

Extension de l'encodage formel des fonctions lexicales dans le cadre de la Lexicologie Explicative et Combinatoire

Anne-Laure JOUSSE
OLST – Université de Montréal
Lattice – Université Paris 7
anne.laure.jousse@umontreal.ca

Résumé. Dans les ressources dictionnairiques développées à partir du cadre théorique de la Lexicologie Explicative et Combinatoire telles que le DiCo, les relations sémantico-lexicales sont modélisées au moyen de fonctions lexicales. Cependant, seulement la majorité d'entre elles (dites standard) répondent véritablement à un encodage formel. Les autres (dites non standard), représentant des relations plus spécifiques à certaines unités lexicales, sont écrites sous la forme d'un encodage hétérogène et très peu formalisé. Par conséquent, certaines relations ne peuvent entrer en ligne de compte dans les traitements automatiques. Nous proposons dans cet article une méthodologie pour la normalisation des fonctions lexicales non standard afin de les rendre exploitables dans des applications telles que l'analyse et la génération de texte. Pour ce faire, nous discutons certains principes théoriques associés à ce formalisme de description et esquissons des propositions pour un traitement global et homogène de l'ensemble des relations décrites dans le DiCo.

Abstract. In the lexicographical products developed within the framework of the Explicative and Combinatorial Lexicology such as the DiCo, the lexico-semantic links are modeled by means of lexical functions. However, only a part of them (called standard) happen to appear as a real formal encoding. The others (called non-standard), which represent links more specific to some lexical units, are written in a heterogeneous and barely formalized way. Therefore, some relations can't be taken into account in automatic processings. We propose, in this paper, a methodology for the normalization of non standard lexical functions in order to make them machine readable in applications such as text-analysis and generation. To complete this work, we discuss some theoretical assumptions drawn upon this formalism and sketch some propositions for a global and homogeneous processing of all the lexical links described in the DiCo.

Mots-clés : Fonctions lexicales (non standard), modélisation des relations sémantico-lexicales, DiCo.

Keywords: (non standard) Lexical Function, modelling of lexico-semantic links, DiCo

1 Introduction

Dans diverses applications pour le traitement automatique des langues, il devient nécessaire de posséder un système formel pour décrire les relations sémantico-lexicales idiomatiques entre les unités lexicales d'une langue. Par relations sémantico-lexicales, nous entendons plus précisément, des relations syntagmatiques ou collocations (ex: pour *abeille* : *piquer*, *butiner*,

polliniser, ...) ainsi que les relations d'ordre paradigmatique, identifiées dans la Lexicologie Explicative et Combinatoire sous le nom de dérivés sémantiques (ex : pour *abeille* : *ruche*, *essaim*, *apiculteur*, ...). Un des objectifs de la modélisation des relations lexicales en analyse est de permettre des inférences à partir d'une unité lexicale : par exemple, l'encodage de la relation entre un animal et son logis permettra d'inférer à partir de la lexie *ruche*, la lexie *abeille*, à partir d'*écurie*, la lexie *cheval*, etc. À l'inverse, en génération, une bonne description des liens lexicaux dans une base de données lexicale fournie permet de doter les textes générés d'un caractère idiomatique. Aussi, pour employer la relation unissant une unité lexicale dénotant un objet et la marque laissée par cet objet, on pourra employer divers types de substantifs selon les unités lexicales concernées : pour un navire, on parlera de *sillage*, pour un pneu ou un pas de *trace* ou d'*empreinte*, pour une plaie, de *cicatrice* ou de *balafre*, etc.

Les bases de données lexicales ne formalisent les relations lexicales souvent que de façon partielle : Wordnet (Fellbaum 1998) ne prend pas en compte les collocations, MindNet (Richardson 1998) se limite aux relations les plus fréquentes. FrameNet (Fillmore *et al.* 2003) propose pour chaque « cadre » un ensemble de relations lexicales, cependant, celles-ci ne sont pas identifiées en elles-mêmes mais par le biais des Frame Elements qui sont de nature conceptuelle et dont le nom peut varier en fonction du cadre dans lesquels ils s'inscrivent. La Théorie Sens-Texte et plus particulièrement la lexicologie explicative et combinatoire (Mel'čuk *et al.* 1995) propose, quant à elle, une modélisation d'un très large éventail de relations sémantico-lexicales au moyen des fonctions lexicales. Celles-ci permettent d'encoder le sens et les caractéristiques syntaxiques d'une relation sous la forme d'une formule succincte, en prenant un mot-clé et en retournant une ou plusieurs valeurs (selon ce schéma noté ici : *mot-clé* FL *valeur(s)*), par exemple :

nom d'endroit typique :	<i>abeille</i>	S_{loc}	<i>ruche</i>
verbe de réalisation typique :	<i>ciseaux</i>	Fact₀	<i>couper</i>
qualificatif positif :	<i>aspiration</i> ¹	Bon	<i>noble</i>

Les fonctions lexicales (désormais FL) ont déjà été largement utilisées dans les applications en traitement automatique (Apresjan *et al.* 2000). Que ce soit pour des programmes de paraphrasage et génération de texte (Nasr 1996, Lareau 2002), de résumés automatiques (Kittredge & Bélanger 2005), de traduction automatique (Apresjan 2003), ou encore pour développer des outils d'apprentissage des langues (Diachenko, 2006, Boguslavsky *et al.* 2006). Cependant, les FL utilisées dans ces projets ne représentent qu'une partie des collocations et dérivés sémantiques propres aux lexies. En effet, l'encodage des FL n'est pas homogène mais varie en fonction du degré de fréquence des relations lexicales : plus une relation sémantique est applicable à un large ensemble d'unités lexicales et plus elle est sera considérée comme standard. Ainsi, alors que la relation modélisée par **S_{loc}** s'applique à beaucoup d'unités lexicales, la relation entre *criminel* et *se repentir* que l'on peut décrire par 'éprouver du regret pour un méfait', ne concerne qu'un très petit nombre d'unités lexicales (*criminel*, *bandit*, *meurtrier*, etc.). Ces relations spécifiques à certaines unités lexicales vont être décrites au moyen de FL dites non standard (ou FLNS). Elles sont encodées selon un métalangage naturel qui met l'accent sur la lisibilité et la clarté de la formulation. Ceci a pour conséquence l'existence de variantes synonymiques comme l'illustrent les trois exemples suivants :

¹ Afin d'alléger le texte nous n'avons pas reproduit le formalisme de la numérotation des lexies du DiCo.

<i>médicament</i>	Discipline qui étudie les M.	<i>pharmacologie</i>
<i>poisson</i>	Science qui étudie les P.	<i>ichtyologie</i>
<i>signe</i>	Étude des S.	<i>astrologie</i>

Tel quel, cet encodage est très hétérogène et inexploitable pour des applications en TAL : il n'est ni repérable ni analysable de façon automatique. Il est donc nécessaire de le normaliser, c'est là l'objectif de notre travail. Nous illustrons, dans cet article, notre démarche en prenant en compte les relations nominales et adjectivales extraites du DiCo (Mel'čuk et Polguère 2006). Nous consacrons la première partie à une discussion théorique du statut et de la légitimité d'une distinction entre relations sémantico-lexicales. La deuxième partie rend compte du travail de normalisation réalisé sur les FL nominales et adjectivales.

2 Enjeu théorique d'une normalisation des fonctions lexicales non standard (FLNS)

2.1 Définition et problèmes

L'objectif des FL est de décrire une relation sémantico-lexicale en un encodage synthétique rendant compte du sens et des caractéristiques syntaxiques d'une relation. Par exemple, la fonction **Real**₁ s'applique à des substantifs pour retourner des verbes ayant un sens de réalisation et prenant pour objet le mot-clé de la relation :

voiture **Real**₁ *conduire* [*une voiture*].

Cependant, comme nous l'avons brièvement évoqué, les FL ne sont pas homogènes, elles s'organisent autour de trois statuts : on passe ainsi d'une description très synthétique et générale (les FL standard) à une description très précise (les FL non standard). Notons à ce propos que ces statuts ne sont pas étanches mais se situent plutôt sur un continuum.

a) Les **FL standard** sont construites à partir d'un noyau de 60 FL simples (Mel'čuk *et al.* 1995 : 125) ; elles doivent répondre au principe d'universalité, c'est-à-dire, exister dans toutes les langues. Elles peuvent être composées d'un seul élément (cf. 1) ou combinées entre elles (cf. 2).

- | | | | |
|-------------------|----------------------|---------------|----------------------------------------|
| (1) <i>joyeux</i> | Anti | <i>triste</i> | : 'antonymie' |
| (2) <i>abcès</i> | IncepPredPlus | <i>mûrir</i> | : 'commencer à devenir plus important' |

b) Les **FL semi-standard** sont constituées d'une FL standard et d'un élément en français venant ajouter une composante de sens non prise en charge par la FL standard. Par exemple, la description de la relation (3) ayant le sens 'causer le silence de qqn', se différencie de la relation (4) par la présence de l'élément **en échange de qqch.** qui vient apporter une composante de sens supplémentaire à la FL standard.

- | | | | |
|--------------------|--------------------------------------------------|----------------|-------------------------------|
| (3) <i>silence</i> | CausOper ₁ | <i>réduire</i> | [<i>qqn. au silence</i>] |
| (4) <i>silence</i> | en échange de qqch. CausOper ₁ | <i>acheter</i> | [<i>le silence de qqn.</i>] |

c) Les **FL non standard** (FLNS), quant à elles, sont écrites intégralement dans la langue de description, ici le français.

- | | | |
|---------------------|--------------------------------------------------------|----------------------|
| (5) <i>faillite</i> | Qui est liée à des irrégularités commises par X | <i>frauduleuse</i> |
| (6) <i>bière</i> | Produite à l'étranger | <i>d'importation</i> |

Polguère (2003 : 4) propose d'assigner le statut standard aux FL satisfaisant les deux conditions suivantes :

- **condition de cardinalité** : La FL doit s'appliquer à un nombre important d'unités lexicales et non être limitée aux unités lexicales d'un seul champ sémantique. La relation

suivante, par exemple : *canard* **Qui ne vit pas à l'état sauvage** *d'élevage, domestique*, ne peut s'appliquer qu'à certains noms d'animaux et ne pourra donc pas prétendre au statut standard.

- **condition de diversité** : Les valeurs retournées par les FL doivent être diversifiées. Ainsi, les valeurs de la FL **Bon** 'qualificatif positif' sont diversifiées selon les mots-clé auxquels elle s'applique, par exemple :

aspiration **Bon noble** ; *bijou* **Bon somptueux** ; *déjeuner* **Bon succulent**, etc.

On peut, par conséquent la considérer comme une FL standard.

Ces conditions offrent des critères opératoires pour la description des relations mais ne se révèlent pas toujours suffisants. Sans vouloir tout à fait remettre en cause cette distinction, nous voudrions réfléchir à sa légitimité et sa pertinence par rapport aux objectifs inhérents à l'existence des FL. Le problème fondamental que nous soulevons ici est le suivant : il semble que la façon d'encoder les relations non standard contribue à reléguer un ensemble non négligeable de relations en marge des descriptions formelles réalisées dans le DiCo et d'autres bases de données du même type. Les FLNS comptent pour 16% des relations du DiCo (1931 non standard sur 11912 relations), or, si l'on veut modéliser le lexique au moyen d'un encodage formel, pourquoi se réduire à laisser de côté une partie relativement importante des relations lexico-sémantiques ? Nous souhaitons remettre en cause, d'une part, le postulat d'un ensemble prétendument fermé de relations standard, d'autre part, la façon dont sont encodées les FLNS. Alors que le nombre de FL standard simple est fixé à une soixantaine, les FLNS constituent un ensemble non fermé de relations sémantiques. Délimiter un nombre de FL est nécessaire pour en conserver un ensemble cohérent et une maintenance raisonnable, cependant, une limitation trop péremptoire pourrait constituer une entrave au bon développement du système des FL. Il nous semble que, si la création de nouvelles FL standard peut amener à supprimer un grand nombre de non standard, il est peut-être légitime d'en ajouter quelques-unes à la liste.

Par ailleurs, la distinction sévère entre FL standard et non standard a entraîné une façon radicalement différente de traiter les relations sémantico-lexicales. Nous pensons que limiter le nombre des FL standard ne devrait pas empêcher une formalisation stricte des autres relations. Dans l'état actuel du DiCo, aucune normalisation des FLNS n'a encore été proposée malgré des régularités flagrantes. Pour palier ce problème, Polguère (2007 : 4) a proposé une autre catégorie de FL à statut hybride entre le standard et le non standard pour les FLNS dont la nature universelle n'est pas démontrée mais qui peuvent être appelées standard pour une langue naturelle donnée. Il s'agit des **fonctions localement standard**. Celles-ci sont écrites à l'aide de formules en français plutôt qu'en latin. Ainsi, la fonction non standard **De_nouveau** illustrée ci-dessous,

hostilité **De_nouveauncepFunc₀** *reprendre* : 'commencer à nouveau'
goût **De_nouveauncepOper₁₂** *reprendre* [*goût* à *qqch.*] : 'avoir de nouveau'

est très fréquente en français mais n'est pas standardisée, car son universalité reste à prouver. Toutefois, elle est considérée comme une fonction lexicale standard locale du français. Cette initiative nous semble être une bonne solution pour réduire le nombre massif de FLNS. Malheureusement, ce principe n'a pas été véritablement étendu ni appliqué à d'autres FL que celles décrites dans l'article cité.

2.2 Vers un niveau de description formel et équilibré : quelle granularité adopter ?

La question centrale lorsque l'on cherche à décrire le sens d'une relation est d'opter pour un degré de précision idéal. Nous voudrions introduire ici le concept de granularité dans la description d'une relation entre unités lexicales. Comme Polguère l'a énoncé dans les conditions citées plus haut, la granularité de la relation ne doit pas être trop restreinte, c'est-à-dire que la description de la relation doit être suffisamment large pour concerner un grand ensemble de lexies. Cependant, aucun critère n'est posé pour le cas contraire, à savoir, pour éviter la description trop générale d'une relation. Le défi est donc de trouver le bon équilibre entre un degré trop fin ou trop large de granularité. Si l'on opte pour des descriptions très précises, on verra le nombre de formules augmenter, en revanche, si l'on vise une généralisation maximale, la clarté de la relation sera mise en péril. Considérons les exemples de la figure 1 ci-dessous. Ils représentent tous une relation entre une entité et un produit qui en est dérivé. On pourrait choisir de regrouper les noms d'animaux et les plats préparés avec leur chair. Dans ce cas, la FL choisie reflèterait une granularité fine qui exclurait les relations entre *éléphant* et *ivoire*, ou entre *pain* et *tartine*, etc. Au contraire, choisir une description plus large comme **Produit dérivé**, par exemple, permet de les regrouper toutes sous une seule et même formule.

Produit dérivé

canard	Foie de C. produit par gavage	foie gras
pain	Tranche de P. enduite de N	beurrée, tartine
sucre	Friandise contenant beaucoup de S.	sucrerie
agneau	Plat à base d'A.	blanquette, rôti, méchoui
cerise	Boisson alcoolisée faite avec des C.	cherry,kirsch, guignolet
coq	Plat à base de viande de C.	au vin
éléphant	Matériau que l'on obtient à partir des défenses d'É.	ivoire
lait	Aliment préparé dérivé du L.	fromage, laitage, beurre

Figure 1 : Regroupements de FLNS nominales sous la formule **Produit dérivé**

Dans le DiCo, chaque FL est accompagnée d'une formule de vulgarisation qui explicite en détail le sens de la relation, par exemple la relation suivante,

gare **IncepReal1** *arriver, entrer* [en ~],

modélisée par **IncepReal1** est vulgarisée par la formule [X] arriver dans une G.. Nous optons donc, en ce qui concerne les FLNS, pour une généralisation maximale de la relation tout en assignant à chacune d'entre elle une vulgarisation sur le modèle des FL standard. Ainsi, il est possible d'obtenir à la fois un encodage formel et une formule explicite et détaillée destinée aux utilisateurs.

3 Propositions pour la normalisation des FLNS

Plusieurs recherches ont été menées pour proposer de nouvelles fonctions lexicales et tenter de normaliser les FLNS. Erastov (1969, cité dans Polguère 2007), Grimes (1990) et Fontenelle (1997) ont proposé quelques nouvelles fonctions lexicales d'après des régularités parmi des relations lexicales observées dans le lexique. Frawley (1998), L'Homme (2002), Jousse (2002), Jousse & Bouveret (2003) et Bouveret (2006) ont postulé quelques adaptations des fonctions lexicales aux relations terminologiques. Grizolle (2003) et Jousse (2003) ont réfléchi à des moyens d'homogénéiser l'encodage des FLNS, Popovic (2004) a travaillé sur l'homogénéisation des gloses de vulgarisation des FL standard. Cependant, la question de

l'encodage des FLNS dans sa globalité n'a pas encore été abordée. Notre objectif dans ce travail est de proposer une normalisation de l'encodage des FLNS. Il est important de distinguer la standardisation de la normalisation. La première consiste à créer de nouvelles FL standard, ce qui nécessite de confronter les candidates aux critères énoncés plus haut ; la normalisation consiste, quant à elle, à observer des régularités parmi les relations lexicales et à proposer une homogénéisation et une formalisation plus stricte de leur encodage sans pour autant les faire passer au statut standard. Elle constitue, en un sens, la première étape vers la standardisation, si cette dernière se révèle possible. Pour ce faire, nous distinguons les relations selon les parties du discours concernées : les relations adjectivales, nominales, verbales et adverbiales. Notons que, dans cet article, nous ne rendrons compte que des deux premiers types de relations. Notre corpus se mesure en terme de FL. Nous avons extrait les 820 FLNS nominales ainsi que les 727 FLNS adjectivales du DiCo et procédé à un traitement différent dans les deux cas.

3.1 Normalisation des FLNS nominales

Les relations encodées par les FLNS nominales sont majoritairement des dérivés sémantiques. Ce sont des unités lexicales partageant une composante de sens avec le mot-clé (ainsi, *apiculteur* est un dérivé sémantique de *abeille* car il est défini comme un 'individu élevant des abeilles'). Nous détaillons ci-dessous six relations lexico-sémantiques nominales extraites du corpus.

a) Les relations **Matériau et Ingrédient**

Les relations 'matériau typique' et 'ingrédient typique' sont parfois considérées comme de la méronymie. On retrouve dans WordNet (Fellbaum 1998), par exemple, trois types de méronymie : « member meronymy » (*association* → *associate*), « substance meronymy » (*steel* → *iron*) et « part meronymy » (*table* → *leg*). Dans le DiCo, la méronymie classique (relation partie-tout) est représentée au moyen de la FL **Mero**. Nous proposons de normaliser les relations ci-dessous par les formules **Matériau** et **Ingrédient**.

Matériau

assiette	Matériau typique dont sont faites les A.	porcelaine
bijou	Matériau dont on fait des B.	pierre, argent, or
corde	Matériau pour C.	lin, coton, nylon
pneu	Matériau dont sont faits les P.	caoutchouc
vaisselle	Matériau précieux dont peut être fait la V.	argent, porcelaine

Ingrédient

bière	Ingrédient utilisé pour faire de la B.	blé, houblon, malt
pain	Substance alimentaire avec laquelle on fait le P.	farine
yaourt	Ingrédient avec lequel on fait le Y.	lait fermenté

Figure 2 : Les fonctions **Matériau** et **Ingrédient**

b) Les relations **Masc, Fem et Infant**

Nous avons repéré trois relations 'mâle ou équivalent masculin de X', 'femelle ou équivalent féminin de X' et 'petit de X ou jeune X' dont l'encodage semble inutilement hétérogène (cf. figure 3). Nous proposons de les regrouper sous trois formules uniques : **Masc** pour les équivalents masculins ou mâles, **Fem** pour les équivalents femelles ou féminins et **Infant** pour les noms de petits d'animaux ou de jeunes individus.

Masc					
abeille	Mâle	faux bourdon			
mouton	M. mâle	bélier			
poule	Mâle de la P.	coq		Fem	
				avocat	De sexe féminin
				canard	Femelle du C.
				chien	Femelle
				mouton	M. femelle
					avocate
Infant					cane
chat	Petit du C.	chaton			chiennne
coq	Jeune C.	coquelet			brebis
grenouille	Larve de la G.	têtard			

Figure 3 : Les fonctions **Masc**, **Fem** et **Infant**

Notons que ces fonctions peuvent se combiner entre elles pour représenter des relations du type cheval / poulain ou cheval / pouliche :

cheval **MascInfant** poulain
cheval **FemInfant** pouliche

c) La relation entre un fait ou une entité et la discipline scientifique qui l'étudie : **Schol**

Des études sur le lien entre dérivés morphologiques et FL standard (Jousse 2002, Jousse et Bouveret 2003) ont montré que la plupart des dérivés morphologiques typiques du français se trouvent déjà modélisés par des FL. Il nous semble pertinent, pour le traitement de certaines FLNS, de se référer aux lexies relevant des compositions gréco-latines les plus courantes pour normaliser certaines relations. Dans bien des cas, la relation entre un fait ou une entité et son étude scientifique est morphologiquement marquée par le suffixe *-logie*, ce qui témoigne d'une récurrence notoire de la relation (cf. figure 4). Nous proposons donc d'encoder cette relation par la formule **Schol**.

Schol		
crime	Science qui étudie ce qui à rapport aux C.	criminologie
poisson	Science qui étudie les P.	ichtyologie
langue	Discipline qui étudie les L.	linguistique
astre	Étude des A.	astronomie

Figure 4 : La relation **Schol**

Faute de place, nous ne pouvons faire un inventaire détaillé de toutes les FLNS que nous avons normalisées. Dans l'état actuel de nos travaux, nous avons réussi à traiter 720 FLNS nominales (sur les 820 de départ) que nous avons réparties en 30 FLNS normalisées. Nous devons convenir que les relations restantes sont difficilement généralisables. Il en est ainsi, par exemple, du lien entre *chat* et *chatière*, représenté par la FLNS **Petite ouverture pratiquée en bas d'une porte qui permet à un C. d'entrer et sortir**. Toutefois, nous tenons à signaler que le DiCo n'est encore que peu développé (1000 acceptions de lexies), ce qui ne permet pas toujours de faire émerger des régularités. Ce travail est donc corrélé à l'avancement du DiCo, il ne peut être fait *a priori* et il est évident que les données évolueront au fur et à mesure de la description de nouvelles relations lexicales.

3.2 Normalisation des FLNS adjectivales

Pour traiter les FLNS adjectivales nous adoptons une démarche très différente dans la mesure où les relations en jeu ne sont pas du même ordre. Les relations encodées par les FLNS adjectivales sont en très grande majorité de type syntagmatique (des collocations) prenant pour mots-clé des noms et retournant comme valeurs des adjectifs modificateurs des mots-clé.

Par exemple : victoire **Dont la probabilité est faible** douteuse

Afin de mener à bien un traitement cohérent et global des FLNS adjectivales, nous faisons l'hypothèse que tout type d'entité ou de fait possède des attributs, par exemple : 'fonction', 'taille', 'forme', 'appréciation', etc., susceptibles d'être exprimés au moyen de collocatifs adjectivaux. Ces attributs peuvent être comparés aux éléments de cadre (= Frame Elements) décrits dans FrameNet (Fillmore *et al.* 2003) qui spécifient les éléments entrant en jeu dans un cadre conceptuel. Par exemple, sous le cadre *Artifact*, on retrouve les éléments : *creator, material, name, time of creation, type, use*. Ces éléments d'ordre conceptuel sont ensuite susceptibles d'être lexicalisés dans les phrases mettant en scène le concept d'artefact.

D'après l'analyse des différentes FLNS, nous avons dégagé un certain nombre d'attributs (environ une vingtaine). Nous cherchons à assigner un attribut à chaque FLNS afin d'en proposer une première formalisation homogène. Ce nouveau type d'encodage permet de recenser les différentes relations, de les organiser et d'y accéder plus facilement dans une base de données. La figure 5 présente (sous la forme **ancienne FL** → **FL normalisée**) quelques-uns de ces attributs et illustre la façon dont ils s'intègrent pour former ce nouvel encodage des FLNS adjectivales.

COULEUR			
	barbe	De couleur grise → Couleur:gris	grisonnante, poivre et sel
FORME			
	barbe	Qui a une forme évasée → Forme: évasée	en éventail
TAILLE			
	drap	Utilisable pour deux personnes → Taille : +	double
	drap	Utilisable pour une personne → Taille : -	simple
MATÉRIAU			
	chapeau	Qui est fait de feutre → Matériau: feutre	de feutre
	chapeau	Qui est fait de paille → Matériau: paille	de paille
FONCTION			
	wagon	Équipé pour que les passagers y dorment → Fonction: dormir	-lit
	wagon	Équipé pour que les passagers y mangent → Fonction: manger	-restaurant
FONCTIONNEMENT			
	horloge	Qui possède un carillon → Fonctionnement: à carillon	à carillon
	horloge	Qui possède un balancier → Fonctionnement: à balancier	à balancier
CAUSE			
	célibat	Qui a lieu malgré la volonté de X → Cause : non voulu	forcé, obligé
PARTICIPANT			
	restaurant	Qui est plutôt fréquenté par des familles → Participants: familles	familial
STATUT			
	wagon	Dans lequel il est interdit de fumer → Statut:interdit de fumer	non-fumeurs
CONSÉQUENCE			
	coup de feu	Qui tue l'être Y → Conséquence:mort de Y	mortel

Figure 5 : Exemple de FLNS adjectivales normalisées

Certains de ces attributs sont polarisables, par exemple : **Taille: +** ou **Taille: -** et constituent ainsi une modélisation directement analysable ; d'autres, en revanche, appellent de fines descriptions pour être compréhensibles, par exemple : **Statut:interdit de fumer**.

La liste présentée ci-dessus n'est pas exhaustive, elle vise simplement à illustrer notre démarche. Nous avons réussi à classer l'ensemble des FLNS adjectivales en une vingtaine d'attributs en essayant au maximum d'en restreindre le nombre. Toutefois, nous avons conscience que certains regroupements ont été un peu forcés. À l'instar des FLNS nominales, nous rappelons que ce travail ne peut se prétendre abouti puisqu'il fera l'objet d'une constante évolution au fur et à mesure du développement du DiCo.

4 Bilan et exemple d'application

D'après l'analyse de régularités parmi les relations sémantico-lexicales décrites dans le DiCo, nous avons cherché à normaliser les fonctions lexicales non standard. Nous avons remis en cause certains principes liés aux fonctions lexicales qui nous semblent figés dans le but d'en envisager de nouvelles possibilités d'encodage. Normaliser de la sorte permet d'obtenir une représentation formelle homogène de l'ensemble des relations lexicales dans les programmes de traitement automatique. Prenons l'exemple d'une application pour la génération d'expressions référentielles (cf. Reiter et Dale 1992). Grâce à ces formules, il est possible de proposer des choix de modificateurs saillants et concis plutôt que des périphrases pour l'identification des objets. Par exemple, pour générer des phrases décrivant les différents objets de la figure 6, on dispose de propriétés discriminatoires telles que la forme, la position, la fonction, la qualité ou le statut.

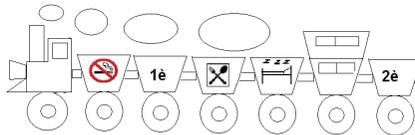


Figure 6 : Illustration d'un train

Ces propriétés pourront être exprimées au moyen des expressions collocatives idiomatiques associées à la lexie *wagon* et encodées par des FLNS normalisées suivantes :

Forme : avec parois et toit	couvert	Fonction : dormir	-lit, -couchette
Forme : avec deux niveaux pour les passagers	à impériale	Fonction : boire ou manger	-bar
Forme : sans paroi ni toit	à plateforme, découvert, plat	Qualité : meilleur confort aux passagers	de première (classe)
Position : fin	de queue	Qualité : plus économique pour les passagers	de seconde (classe)
Position : début	de tête	Statut : interdit de fumer	non-fumeurs
Position : suit ou précède un autre	voisin-adj	Statut : permis de fumer	fumeurs
Fonction : manger	-restaurant		

Notons que ce travail constitue une première étape dans le sens où nous n'avons pour l'instant pris en compte que les relations nominales et adjectivales. On doit également préciser qu'il s'agit d'une tâche évolutive parallèle au développement du DiCo et que les résultats seront amenés à des changements certains. Il nous semble donc primordial de poser un cadre méthodologique précis pour le traitement des relations lexico-sémantiques non standard.

Remerciements

Je remercie chaleureusement Sylvain Kahane, Alain Polguère et Frédéric Landragin pour la lecture de cet article.

Références

- APRESJAN J., BOGUSLAVSKY I. , IOMDIN L., TSINMAN L.(2000). Lexical Functions in NLP: Possible Uses. *Computational Linguistics for the New Millenium: Divergence or Synergy*. Heidelberg, 2000, p. 1-11.
- APRESJAN J. D. *et al.* (2003). Lexical Functions as a Tool of ETAP-3, *Actes de MTT 2003*, Paris, 16-18 juin 2003.
- BOGUSLAVSKY I., BARRIOS RODRÍGUEZ M., DIACHENKO P. (2006). CALLEX-ESP: a software system for learning Spanish lexicon and collocations, *Current Developments in Technology-Assisted Education*.
- BOUVERET M. (2006). Fonctions lexicales pour le typage de relations syntagmatiques et

- paradigmatiques en bioindustries, une approche lexicographique du terme, *The Processing of Terms in Dictionaries : New models and techniques, special Issue of Terminology* 12(2), John Benjamins Publishing Company, Amsterdam/Philadelphia. (à paraître)
- DIACHENKO P. (2006). Lexical functions in learning the lexicon, *Current Developments in Technology-Assisted Education*.
- ÉRASTOV K. O. (1969). Primery slovarnyx opisaniy [Some sample lexicographic descriptions]. *Mašinnyj perevod i prikladnaja lingvistika* [Automatic Translation and Applied Linguistics], 11, 36-59.
- FELLBAUM C. (Ed.). (1998). *Wordnet: An Electronic Lexical Database*. MIT Press.
- FILLMORE C., JOHNSON C. et PETRUCK M. (2003). Background to FrameNet, *International Journal of Lexicography*, Vol. 16, n°3 : 235-249.
- FONTENELLE T. (1997). *Turning a bilingual dictionary into a lexical-semantic database*, Tübingen : Niemeyer.
- FRAWLEY W. (1998). New forms of Specialized Dictionaries, *International Journal of Lexicography*, vol. 1, n°3, 1988, p.89-213.
- GRIMES, J. (1990). Inverse Lexical Functions, in: J. Steele (ed.) (1990). *Meaning-Text Theory: Linguistics, Lexicography and Implications*, Ottawa: Ottawa University Press, pp. 350-364.
- GRIZOLLE B. (2003). *Classification des fonctions lexicales non-standard du DiCo*, rapport de stage de maîtrise, OLST, Université de Montréal.
- JOUSSE A.-L. (2002). *Dérivation morphologique de termes, analyse en corpus spécialisé et modélisation au moyen des fonctions lexicales*, Mémoire de maîtrise, Université du Maine.
- JOUSSE A.-L. (2003). *Normalisation des fonctions lexicales*, Mémoire de DEA, Paris 7.
- JOUSSE A.L., BOUVERET M. (2003). Lexical functions to represent derivational relations in specialized dictionaries, *Terminology*, 9(1), 71-98.
- KITTREDGE R., BÉLANGER P. (2005). Paraphrasing with Space Constraints: Linguistic Operations in Journal Abstracting, *Actes de MTT 2005*, Moscou, 23-25 juin 2005.
- LAREAU F. (2002). *La synthèse automatique de paraphrases comme outil de vérification des dictionnaires et grammaires de type Sens-Texte*. Mémoire de maîtrise, Université de Montréal.
- L'HOMME M.-C. (2002). Fonctions lexicales pour représenter les relations sémantiques entre termes, *Traitement automatique de la langue*, 43(1), pp. 19-41.
- MELČUK I., CLAS A. et POLGUÈRE A. (1995). *Introduction à la lexicologie explicative et combinatoire*, Louvain-la-Neuve (Belgique), Duculot.
- MELČUK I. et POLGUÈRE A. (2006). Dérivations sémantiques et collocations dans le DiCo/LAF. *Langue française*, numéro spécial sur la collocation « Collocations, corpus, dictionnaires », sous la direction de P. Blumenthal et F. J. Hausmann, 150, juin 2006, 66-83.
- NASR A. (1996). *Un modèle de reformulation automatique fondé sur la Théorie Sens Texte: Application aux langues contrôlées*, Thèse de doctorat en informatique, Université Paris 7
- POLGUÈRE A. (2000). Towards a theoretically-motivated general public dictionary of semantic derivations and collocations for French, *Proceedings of EURALEX, ...*
- POLGUÈRE A. (2007). Lexical function standardness, in Wanner L. (ed.). *Selected Topics in Meaning Text Theory, In Honour of Igor Mel'čuk*. (à paraître).
- POPOVIC S. (2003). Métalangage de vulgarisation des liens de fonctions lexicales, *Actes de la première conférence internationale sur la Théorie Sens-Texte*, Paris, juin 2003.
- REITER E. et DALE R. (1992). A fast algorithm for the generation of referring expressions. *Proc. of the 14th Int. Conference on Computational Linguistics*, 232-238, Nantes, France.
- RICHARDSON S., DOLAN W., et VANDERWENDE L. (1998). MindNet: Acquiring and Structuring Semantic Information from Text, *COLING-ACL 1998*: 1098-1102.