

De la lexie au vocable : la représentation formelle des liens de polysémie

Lucie Barque
LaTTice
Université Paris 7
lbarque@linguist.jussieu.fr

Résumé - Abstract

Cet article s'intéresse aux définitions formalisées de la base de données BDéf et montre en quoi la structure formelle de ces définitions est à même d'offrir une représentation originale de la polysémie lexicale.

This article deals with formal structure of BDéf's definitions and shows in what way this structure is appropriate to offer an original representation of the lexical polysemy phenomenon.

Mots-clefs – Keywords

LEC, dictionnaire électronique formalisé, structure définitionnelle, TAL
ECL, formalised electronical dictionary, structure of the definition, NLP

1 Introduction

Les dictionnaires de langue s'attachent à rendre compte de la valeur des lexies : on ne cherche pas à définir leur référent mais les différents rapports qu'elles entretiennent entre elles dans le réseau lexical de la langue. C'est en cela que l'on distingue, de manière toute théorique, la définition encyclopédique de la définition linguistique, dans laquelle chaque composante de sens indique un lien, syntagmatique ou paradigmatic, avec une autre lexie (Picoche, 1997). Cet article s'intéresse aux composantes définitionnelles indiquant un lien de polysémie entre deux lexies, et à leur représentation dans les définitions de la BDéf, base de données lexicales en cours de développement à l'université de Montréal (Altman & Polguère, 2003). La BDéf encode les définitions du *Dictionnaire Explicatif et Combinatoire* (Mel'čuk *et al.*, 1984, 1988, 1992, 1999) rédigées suivant les principes de la Lexicologie Explicative et Combinatoire (LEC)¹. La base de données a recours à un formalisme qui explicite la structure des définitions du DEC, donnant ainsi à l'utilisateur, qu'il soit lexicographe ou programme informatique, la possibilité d'accéder aux différentes strates d'une définition. La première section est consacrée à la présentation de

¹La LEC est la composante lexicale de la Théorie Sens-Texte (TST). Pour une introduction à la TST, voir (Mel'čuk, 1997)

la structure formelle de ces définitions. On propose ensuite de les exploiter pour l'étude de la polysémie lexicale : la structuration explicite des définitions BDéf nous invite à nous interroger sur la nature d'une composante liante et sur la place qu'elle occupe au sein de la structure définitionnelle. On envisage enfin, dans une troisième section, d'extraire à partir des vocables de la BDéf des patrons formels de polysémie qui serviront à proposer un traitement approprié de ce phénomène lexical.

2 Présentation des définitions formalisées de la BDéf.

Les définitions de la BDéf se distinguent des définitions traditionnelles en ce qu'elles ont recours à un métalangage contraint, nécessaire à une description logique du lexique et particulièrement bien adapté au traitement automatique des définitions (Altman & Polguère, 2003), (Wanner, 2003). Elles proposent en outre une structuration explicite du sens lexical. La figure 1 présente la définition de MANGER#I.1a encodée dans le formalisme de la BDéf.

MANGER#I.1a	
Forme propositionnelle : X ~ Y	Composante centrale : 1 : et_puis 1.1 : X introduire Y dans Alpha 1.2 : (X mastiquer Y) 1.3 : X avaler Y
Typage des variables : X : être animé Y : matière Alpha : orifice	Différences spécifiques : /*But*/ 2 : *1 dans_le_but *3 3 : X _se_nourrir_#I1 de Y /*Caractéristique.Alpha*/ 5 : Alpha servir à *6 6 : X _se_nourrir_#I1
Rel_Sém : Alpha[X]	

FIG. 1 – Définition de MANGER#I.1a, BDéf

La partie gauche de la figure contient trois blocs relatifs aux actants sémantiques de la lexie, représentés par des variables. Le premier, intitulé **forme propositionnelle**, indique la position des actants par rapport à la lexie vedette (représentée par le signe ~). Le second, intitulé **Typage des variables**, attribue une étiquette sémantique aux différents actants. Le troisième bloc (**Rel_Sém**) indique les éventuelles relations sémantiques entre les actants. Ici, X est le premier actant de Alpha (orifice de l'être animé X). La partie droite de la figure contient la définition proprement dite.

Une définition BDéf présente différents niveaux de structuration : elle est d'abord scindée en deux grands **blocs définitionnels** qui correspondent à la séparation classique genre prochain (appelé ici **composante centrale**) / différences spécifiques. Le genre prochain donne le sens général de la lexie vedette et les différences spécifiques permettent de la distinguer des autres lexies ayant le même genre prochain. Le bloc des différences spécifiques est constitué d'un ensemble de **blocs définitionnels élémentaires** qui sont précédés d'un en-tête indiquant le rôle²

²Le terme *rôle* renvoie à la même notion que les **rôles** de la structure des qualia dans le Lexique Génératif

du bloc par rapport à la composante centrale. Ainsi, les différences spécifiques de la lexie MANGER#I.1a sont constituées d'un premier bloc dont le rôle est d'indiquer le but de l'activité dénotée par la lexie. Le second bloc indique une caractéristique de l'actant Alpha.

Chaque bloc définitionnel est constitué d'un ensemble de **propositions élémentaires** numérotées. Une proposition élémentaire est constituée d'un **prédicat principal** et des **positions actancielles** prévues par ce prédicat³. Par exemple, le prédicat principal de la proposition 1.1 est *introduire*. Ce prédicat requiert trois positions actancielles : une pour celui qui introduit (X), une pour ce qu'il introduit (Y) et une pour l'endroit où il introduit (Alpha). Un prédicat principal peut être modifié au sein d'une proposition élémentaire. On peut imaginer par exemple que la définition de GAVER contienne la même proposition que la proposition 1.1 mais constituée en plus du **modifieur de force** (X introduire Y de force dans Alpha).

Les intérêts pratiques et théoriques d'une telle base de données lexicales ont été présentés dans (Altman & Polguère, 2003) et (Wanner, 2003). La section suivante s'appuie sur les définitions formelles de la BDéf pour entamer une réflexion sur la notion de composante liante.

3 La composante liante dans les définitions de la BDéf.

L'unité polysémique de la BDéf, le vocable, se présente sous la forme d'une liste de lexies correspondant à ses différentes acceptions. Pour figurer au sein d'un vocable, chaque lexie doit être liée de manière directe ou indirecte à la lexie de base (le sens premier de l'unité polysémique)⁴. Si deux lexies homographes n'entretiennent aucun rapport de sens, elles figurent dans des vocables différents, considérés comme homonymes. La LEC préconise donc que soit explicitée la composante de sens liante, appelé **pont sémantique** entre deux lexies. La figure 2 montre les définitions de trois lexies du vocable COUDE, la lexie de base (COUDE#Ia[*Il s'est luxé le coude*]), une acception métonymique (COUDE#II.1[*Les coudes de sa veste sont lustrés*]) et une acception métaphorique (COUDE#II.2[*La route formait un coude à cet endroit*]). Les ponts sémantiques apparaissent en italique dans la figure.

La lexie de base du vocable GENOU, que nous ne pouvons présenter ici faute de place, possède les mêmes différences spécifiques que celle du vocable COUDE à l'exception du modificateur (en formant un angle) de la proposition n°6 (cf figure 2), qui n'y figure pas. Cette divergence s'explique par le fait que le vocable COUDE contient, contrairement au vocable GENOU, une acception métaphorique qui exploite cette composante (voir COUDE#II.2, figure2). Il semble donc que la structure du vocable puisse motiver l'inclusion d'une composante de sens dans la définition de la lexie de base.

(Pustejovsky, 1995). Toutefois :

- Les rôles du Lexique Génératif sont limités au nombre de quatre alors que la liste des différences spécifiques de la BDéf n'est pas a priori close.
- Les rôles du Lexique Génératif servent à la description de toutes les lexies, quel que soit leur type sémantique. On envisage au contraire dans la BDéf de mettre au jour des structures définitionnelles propres aux différents types de lexies.

³Certaines propositions élémentaires peuvent n'être constituées que d'un connecteur. C'est le cas ici de la proposition n°1 : *et_puis* marque une succession temporelle entre les événements dénotés par les propositions 1.1, 1.2 et 1.3 (X introduit Y dans Alpha *et_puis* X mastique Y *et_puis* X avale Y).

⁴Pour les critères permettant d'élire le sens de base et de décider du découpage du polysème en unité de sens discrètes, voir (Mel'čuk *et al.*, 1995).

COUDE#I.a de animé X	COUDE#II.1 de vêtement X	COUDE#II.2 de inanimé X
COMPOSANTE CENTRALE : 1 : partie de n1=bras#I.1a de X DIFFÉRENCES SPÉCIFIQUES : /*Localisation_spatiale*/ 2 : *1 situé au_milieu de *n1 3 : *1 situé sur face#II.2 extérieur de *n1 /*Fonction_physiologique*/ 4 : *1 constituer articulation de *n1 5 : *4 permettre *6 6 : *n1 se_plier à *1 <i>en formant un angle</i>	COMPOSANTE CENTRALE : 1 : partie de X DIFFÉRENCES SPÉCIFIQUES : /*Fonction*/ 2 : *1 couvrir <i>coude#Ia</i> de A quand *3 3 : A porter X	COMPOSANTE CENTRALE : 1 : partie de X DIFFÉRENCES SPÉCIFIQUES : /*Forme*/ 2 : *1 courbe /*Perception*/ 3 : *1 perçu comme *4 4 : *1 <i>former angle</i>

FIG. 2 – Trois acceptions du vocable COUDE, *BDéf*

Cette observation appelle deux commentaires. D'une part, cela revient à dire que, de même qu'il est possible, lorsque l'on emploie une lexie ayant un sens figuré, de faire référence en même temps au sens propre correspondant (certains jeux de mots reposent sur ce principe), de même le sens de la lexie de base peut porter le sens de certaines acceptions figurées. Ce point de vue n'est pas forcément recevable et il faudra se demander si le pont sémantique ne doit pas dans certains cas demeurer implicite. La lexie *SE_PLIER* contient en effet probablement déjà la composante (*former un angle*) dans sa définition. D'autre part, et en faveur cette fois de l'explicitation du pont sémantique, on peut voir cette composante définitionnelle de la lexie de base comme un indice de l'acceptation métaphorique et envisager par conséquent de prévoir, dans une certaine mesure, les transferts métaphoriques (moins réguliers que les transferts métonymiques (Apresjan, 1974)), à partir du sens de base. Un travail de relevé des ponts sémantiques reste à effectuer sur l'ensemble de la base de données, qui permettra de valider ou d'invalider ces prévisions. C'est ce qui est envisagé dans la section suivante.

4 Elaborer des patrons formels de polysémie.

Les définitions de la *BDéf* facilitent par leur structuration explicite l'identification des ponts sémantiques entre les lexies d'un vocable. Le travail consistant à élaborer des ensembles de patrons formels de polysémie à partir des définitions *BDéf* nous semble intéressant pour :

Son originalité. A notre connaissance, aucun travail de ce genre n'a été entrepris à partir de bases de données lexicales couramment utilisées dans des applications de TAL (cf WordNet⁵ (Miller, 1990) et FrameNet⁶). Cette recherche s'effectuera en outre sur des données lexicographiques du français idoines. Les liens de polysémie ne sont en effet que partiellement représentés dans les bases de données lexicales de l'anglais mentionnées ci-dessus : dans WordNet, les différents sens d'un mot sont isolés dans des synsets, unité conceptuelle représentée par un regroupement d'unités lexicales synonymiques. De même dans FrameNet, une unité lexicale est associée à autant de frames (la situation conceptuelle dénotée par la lexie) qu'elle compte d'acceptations. Certains liens de polysémie reposant sur des différences de réalisation syntaxique (par exemple, l'alternance non-causatif/causatif) sont toutefois représentés grâce à la mise en

⁵<http://www.cogsci.princeton.edu/~wn/>

⁶<http://www.icsi.berkeley.edu/~framenet>

correspondance des éléments du frame avec leur différentes réalisations syntaxiques.

La valorisation de la BDéf. Les en-têtes de blocs définitionnels sont à rapprocher des traits de la structure des qualia du Lexique Génératif mais sont en nombre moins restreints (cf note 1). Les mécanismes de générations de sens en contexte proposés par le Lexique Génératif, comme par exemple les règles de coercition de type, reposent en partie sur ces différents traits. Il nous faudra étudier les différents types de blocs abritant le pont sémantique dans une définition et voir *a posteriori* si ils sont pertinents pour éviter les problèmes de surgénération de sens.

L'exploitation des résultats dans le cadre de la lexicologie computationnelle. Les patrons de polysémie pourront servir par exemple à proposer des règles lexicales permettant de générer automatiquement des définitions à partir de la définition de la lexie de base. Par exemple, la règle déduite d'un patron de polysémie extrait du vocable COUDE (figure 3) pourra être appliquée aux lexies JAMBE#1a, DOS#1a, DOIGT#1a, etc.

L1	L2
Etiqu_Sém : partie du corps	Etiqu_Sém : partie de vêtement
	L2 couvrir L1

FIG. 3 – Règle lexicale

Il faudra bien entendu pouvoir contraindre l'application de cette règle au moyen d'autres règles afin d'empêcher la génération d'une lexie étiquetée *partie de vêtement* à partir des lexies de base COU#1a, BRAS#1a, etc. On s'inspirera notamment du type de règles proposées dans (Copestake & Briscoe, 1995). La règle du blocage lexical, par exemple, permet de s'assurer que le sens à générer n'est pas lexicalisé d'une autre manière. L'homogénéité de l'encodage des définitions BDéf devrait permettre d'identifier aisément le prédicat couvrir(L2,L1) dans les définitions de COL#1a et MANCHE#1a.

5 Conclusion

Nous nous sommes intéressés dans cet article à la structure formelle des définitions de la BDéf, base de données lexicales actuellement en cours de développement à l'université de Montréal. Nous avons montré que la BDéf, loin de nier la polysémie, offrait au contraire la possibilité d'en dégager une représentation formelle originale. Nous nous sommes contentés pour l'heure de soulever quelques problèmes liés à la représentation de ce phénomène lexical. L'étape suivante de cette recherche consistera, lorsque le formalisme de la BDéf sera entièrement stabilisé, à élaborer des ensembles de patrons formels de polysémie qui permettront de prendre la mesure exacte du caractère régulier ou non des différents liens de polysémie.

Références

- ALTMAN J., POLGUERE A. (2003), La BDéf : base de définitions dérivée du Dictionnaire explicatif et combinatoire, *Proceedings of the First International Conference on Meaning-Text Theorie*, Paris, p.43-54.
- APRESJAN J. (1974), Regular Polysemy, *Linguistics*, 142, p.5-32.

- COPESTAKE A., BRISCOE T. (1995), Semi-productive polysemy and sense extension *Journal of Semantics* 12(1),15-68.
- MEL'CUK I. *et al.* (1984, 1988, 1992, 1999), *Dictionnaire explicatif et combinatoire du français contemporain*. Recherches lexico-sémantiques I, II, III, IV, Montréal, Presses de l'Université de Montréal.
- MEL'CUK I. (1997), *Vers une Linguistique Sens-Texte*, Leçon inaugurale au Collège de France, Collège de France, Paris.
- MEL'CUK I., CLAS A., POLGUERE A. (1995), *Introduction à la lexicologie explicative et combinatoire*, AUPELF-UREF/Duculot, Louvain-la-Neuve.
- MILLER G. (ed) (1990) *WordNet : An On-Line Lexical Database*, in : *International Journal of Lexicographie*, 3 :4, pp. 233-249.
- PICOCHÉ J. (1997) *Précis de Lexicologie Française*, Nathan,Paris.
- POLGUERE A. (2003), *Lexicologie et sémantique lexicale. Notions fondamentales*, Presses Universitaires de Montréal.
- PUSTEJOVSKY J. (1995), *The Generative Lexicon*, MIT Press.
- VICTORRI B., FUCHS C. (1996) *La polysémie. Construction dynamique du sens*, Hermès, Paris.
- WANNER L. (2003), *Definitions of Lexical Meanings : Some Reflections on Purpose and Structure*, *Proceedings of the First International Conference on Meaning-Text Theory*, Paris, p. 55-65.