

Application du métalangage définitionnel de la BDéf au traitement formel de la polysémie

Lucie Barque (1), Alain Polguère (2)

(1) Lattice - Université Paris 7

UFRL, Case 7003

75251 Paris cedex 5, France

lbarque@linguist.jussieu.fr

(2) OLST - Université de Montréal

C.P.6128, succ. Centre-ville

Montréal (Qc), H3C 3J7, Canada

alain.polguere@umontreal.ca

Mots-clefs : Base de données lexicale, métalangage définitionnel, Lexicologie Explicative et Combinatoire, polysémie

Keywords: lexical database, definitionnal metalanguage, Explanatory Combinatorial Lexicology, polysemy

Résumé Cet article a pour objet le métalangage définitionnel de la base de données lexicale BDéf, plus précisément l'utilisation de ce métalangage dans la modélisation des structures polysémiques du français. La Bdéf encode sous forme de définitions lexicographiques les sens lexicaux d'un sous-ensemble représentatif du lexique du français parmi lequel on compte environ 500 unités polysémiques appartenant aux principales parties du discours. L'article comprend deux sections. La première présente le métalangage de la BDéf et le situe par rapport aux différents types de définitions lexicales, qu'elles soient ou non formelles, qu'elles visent ou non l'informatisation. La seconde section présente une application de la BDéf qui vise à terme à rendre compte de la polysémie régulière du français. On y présente, à partir d'un cas spécifique, la notion de patron de polysémie.

Abstract We present the defining metalanguage of the BDéf lexical database ; more specifically, we focus on how this metalanguage can be used to model relations of polysemy in French. The BDéf contains lexical definitions for a representative subset of the French lexicon : around 500 polysemic words belonging to all major parts of speech. This paper contains two sections. Firstly, the BDéf metalanguage is introduced and positioned relative to different types of existing lexical definitions : formal vs. non formal definitions, definitions that are tailored or not for implementation. Secondly, the paper shows how the BDéf approach is used in a research whose goal is the modeling of the regular polysemy of the French language. The notion of pattern of polysemy is introduced using a specific example.

1 Place de la définition BDéf dans une typologie de la définition en lexicographie et en sémantique formelle

La BDéf, base de données lexicale développée à l'Observatoire de Linguistique Sens-Texte de l'Université de Montréal (Altman et Polguère, 2003), encode les définitions du *Dictionnaire Explicatif et Combinatoire* (Mel'čuk *et al.* 1984, 1988, 1992, 1999) dans un formalisme qui explicite leur structure interne.

Nous classons les différents types de définitions lexicales connus en deux grandes familles, caractérisées non sur la base de leurs fondements théoriques, mais sur celle des formalismes adoptés : d'un côté les approches qui ont recours à la paraphrase pour expliciter le sens d'une unité lexicale, de l'autre celles qui représentent le sens lexical au moyen de structures de traits.

1. **Les formes de la paraphrase** La paraphrase définitionnelle est une décomposition sémantique d'une unité lexicale organisée de façon linéaire. Le caractère linéaire de ce type de représentation lexicale est le point commun des trois formes de paraphrases présentées ci-dessous. Elles divergent par leur degré de formalisation.
 - (a) **Les définitions lexicographiques standard** utilisent un métalangage proche de la langue objet qu'il décrit. Cela leur confère un caractère naturel qui répond bien à l'exigence de substituabilité de la paraphrase.
 - (b) **Les définitions analytiques** suivent les mêmes principes définitoires (Aristote, ed. 2004) que les dictionnaires de langue courants mais sont construites de manière à pouvoir identifier clairement les différents éléments de la décomposition (Wierzbicka, 1987), (Mel'čuk *et al.*, 1995).
 - (c) **Les définitions logiques** comme celles de (Dowty, 1979) sont très différentes des deux autres types de paraphrases du point de vue de la théorie sémantique sous-jacentes. Les formules logiques se présentent cependant elles aussi sous une forme linéaire qui permet une appréhension relativement naturelle du sens lexical.
2. **Les structures de traits** Comme les paraphrases définitionnelles, les structures de traits représentent une décomposition du sens lexical. Ici, la décomposition n'est pas organisée de façon linéaire mais prend la forme d'un ensemble de traits. Les théories de sémantique computationnelle adoptent en général ce mode de représentation de la décomposition lexicale (Pottier 1974), (Rastier 1987), (Pustejovsky, 1995).

Le formalisme d'encodage des définitions lexicales de la BDéf emprunte ses caractéristiques à l'une et l'autre des deux grandes familles qui viennent d'être distinguées. D'une part la BDéf, parce qu'elle se veut un outil servant à réfléchir sur la nature du sens lexical, s'attache à mettre en évidence les différentes composantes de sens qui forment la définition ainsi que la façon dont ces composantes s'organisent. Le formalisme adopté permet de rétablir facilement la forme linéaire de manière à aider le lexicographe lors de la construction des définitions. D'autre part, la BDéf doit constituer une ressource au service de la communauté du TAL et utilise de ce fait un formalisme très proche de celui des réseaux sémantiques. Une définition Bdéf peut ainsi être exploitée dans des calculs, comme nous le verrons dans la seconde section de cet article.

Une présentation détaillée des principes sous-jacents à la base de données BDéf et du formalisme d'encodage des définitions adopté figure dans (Altman et Polguère, 2003). Nous en présentons ici les grandes lignes. La figure 1 ci-dessous reproduit la fiche BDéf modélisant le sens de la lexie MAISONI.1 [*Elle cherche la maison où habite son cousin*]. Cet exemple va nous servir

à illustrer les différentes notions décrivant la structure des définitions : les propositions élémentaires sont organisées en blocs définitionnels élémentaires, eux-mêmes regroupés en blocs de second niveau, la composante centrale et les différences spécifiques.

MAISON I.1
Composante centrale : 1 : habitation de X
différences spécifiques : /*Dimensions*/ 2 : *1 grand.relativement /*Structure*/ 3 : *1 constitué de niveau /*Matériau*/ 4 : *1 fabriqué avec matériau.résistant
Typage des actants : X : individu

FIG. 1 – La définition BDéf de MAISON I.1

La proposition élémentaire C'est la composante sémantique minimale de la définition. Chaque proposition élémentaire est constituée d'un prédicat accompagné des positions actancielles qu'il requiert et est identifiée par un numéro. Une position actancielle peut ainsi être occupée par un pointeur (*1, *2, ...) vers une autre proposition. Par exemple la proposition n°2 de la définition présentée dans la figure 1 est constituée du prédicat *grand* qui a comme argument la proposition n°1.

Le bloc définitionnel élémentaire Les propositions élémentaires sont regroupées en blocs définitionnels qui représentent des composantes « autonomes » de la définition. Chaque bloc est introduit par un en-tête synthétisant l'information qui y est contenue. Dans la définition de MAISON I.1, la proposition n°2 spécifie les *dimensions* d'une maison.

La composante centrale et les différences spécifiques Les blocs définitionnels élémentaires sont organisés au niveau supérieur en deux grands blocs : la composante centrale, qui correspond au sens général de la lexie décrite – ce que la tradition appelle le *genre prochain* – et les différences spécifiques.

2 Modélisation des liens de polysémie

Le développement de la BDéf et de son métalangage a donné lieu à un projet de modélisation des structures polysémiques du français, que nous présentons dans cette section.

2.1 Définitions

Un **vocable polysémique** est constitué d'un ensemble de lexies partageant les mêmes formes et qui ont en commun une composante sémantique non triviale¹. Cette dernière propriété permet de distinguer deux lexies en relation de polysémie de deux lexies en relation d'homonymie (Mel'čuk *et al.*, 1995), (Victorri et Fuchs, 1996), (Kleiber, 1999).

Un **lien de polysémie** se définit comme un lien sémantique entre deux lexies d'un vocable polysémique. Nous nous intéresserons ici à un type particulier de lien de polysémie, celui qui résulte d'un processus de dérivation sémantique. On dira qu'une lexie L2 est dérivée sémantiquement d'une lexie L1 si le sens de L2 est « construit à partir du » sens de L1. On appellera dans ce cas de figure L1 la **lexie source** et L2 la **lexie dérivée**.

Un **patron de polysémie** modélise un lien de polysémie au moyen d'une paire de structures définitionnelles sous spécifiées. Le nom du patron doit rendre compte d'une part du changement (ou non) de l'étiquette sémantique (Polguère, 2003) et d'autre part de la transformation à laquelle on a affaire (par exemple, quel type de métonymie, quel type de métaphore, etc). Un patron de polysémie doit être suffisamment général pour pouvoir s'appliquer à **au moins deux couples** de lexies.

2.2 Un exemple de modélisation : le cas du vocable MAISON

Notre choix s'est porté ici sur deux lexies du vocable MAISON : MAISONI.1 [*Elle cherche la maison où habite son cousin.*] et MAISONI.2 [*Elle a quitté la maison à l'âge de 18 ans.*]². Les définitions des lexies sont présentées ci-dessous sous forme linéaire³ :

MAISONI.1 de X \equiv ' /*habitation*/ habitation de l'individu X, /*Dimensions*/ relativement grande, /*Structure*/ constituée d'un ou plusieurs étages et /*Matériau*/ fabriquée avec des matériaux solides '

MAISONI.2 de X \equiv ' /*lieu occupé*/ lieu occupé par un individu ou un groupe d'individus X /*Limites*/ situé à l'intérieur de l'habitation de X '

Notons que la lexie dérivée est beaucoup plus contrainte dans son fonctionnement dans la phrase que la lexie source puisqu'elle ne s'emploie qu'au défini singulier et n'accepte pas le déterminant possessif : dans la phrase *Elle a quittée sa maison à l'âge de 18 ans*, il s'agit nécessairement de MAISONI.1 et non de MAISONI.2.

La figure 2 ci-dessous représente le patron de polysémie **Synecdoque faible : (partie d'une) construction**→**lieu occupé** s'appliquant au couple MAISONI.1~MAISONI.2. Le lien entre les deux lexies est une **synecdoque** dans la mesure où le sens de la lexie dérivée désigne une partie de ce que dénote la lexie source : comme on le voit dans le patron de la figure 2, le prédicat de synecdoque être_intérieur_de lie le sens général de la lexie dérivée à celui de la lexie source⁴. Le fait que le second argument du prédicat être_intérieur_de, dans la définition

¹Concernant l'identification de la polysémie, nous adoptons les critères descriptifs et méthodologiques de la lexicologie explicative et combinatoire, explicités dans (Mel'čuk *et al.*, 1995).

²Ces deux lexies ont des correspondants en anglais qui n'appartiennent pas au même vocable, respectivement HOUSE et HOME.

³Nous avons introduit (en italique) dans les définitions linéaires les en-têtes de blocs des définitions BDéf correspondantes afin de faciliter par la suite la lecture du patron de polysémie auquel s'applique ce couple de lexies.

⁴le sens général d'une lexie est représenté dans le patron par l'étiquette du bloc de la composante centrale de sa définition, ou une étiquette mère de celle-ci. Par exemple, l'étiquette construction est l'étiquette mère de

de MAISONI.2, soit habitation de X et non MAISONI.1 (*i.e* une inclusion complète) nous amène à qualifier cette synecdoque de **faible**. Cette particularité est marquée linguistiquement par le fait que l'on peut dire *elle a quittée la maison à l'âge de 18 ans*, que cette personne ait habité dans une maison, dans un appartement ou dans une tout autre habitation. À l'inverse, on dira que les lexies ASSIETTEI.1 [*Il a cassé trois assiettes.*] et ASSIETTEI.2 [*Il a mangé trois assiettes de riz.*] sont liées par une **synecdoque forte** car on ne peut pas dire *Il a mangé trois assiettes de riz* si le riz n'est pas contenu dans une assiette (ASSIETTEI.1). Comme il a été signalé plus haut, un patron de polysémie doit rendre compte du type de lien de polysémie (dans notre cas, il s'agit d'une synecdoque faible) mais également de l'éventuel changement du type sémantique entre la lexie source et la lexie cible. Cette information figure dans le patron au niveau de la valeur de l'attribut `Composante centrale`.

LEXIE SOURCE
Composante centrale [1] (partie d'une) construction
Composante exportée [1]
Adresse de la composante exportée /*Construction*/
LEXIE DÉRIVÉE SÉMANTIQUEMENT
Composante centrale [2] lieu occupé
Adresse de la composante importée /*Limites*/
Prédicat de synecdoque être_intérieur_de([2], [1])

FIG. 2 – Patron de polysémie **Synecdoque faible : (partie d'une) construction**→**lieu occupé**

Le patron de polysémie qui vient d'être identifié modélise une alternance récurrente en français, quoique plus marginale que celles de la famille des synecdoques fortes. Le patron **Synecdoque faible : (partie d'une) construction**→**lieu occupé** peut par exemple s'appliquer également au couple de lexies BUREAUI.2 [*Les fenêtres du bureau donnent sur la cour.*]~BUREAUI.3 [*Quand on est au bureau, le temps passe lentement.*] dans la mesure où l'on peut dire que *Jean a passé la journée au bureau* même si Jean n'a pu mettre les pieds dans son bureau (BUREAUI.2), retenu toute la journée en salle de réunion. Le degré de décomposition des définitions et l'organisation explicite de cette décomposition nous permet ainsi distinguer de manière fine des liens de polysémie proches. Cela nous permet par exemple de préciser la distinction entre polysémie systématique et polysémie régulière (Apresjan, 1974) : toute lexie étiquetée `construction` ou `partie de construction` est susceptible de « générer » une lexie étiquetée `lieu occupé` (par exemple ÉCOLE, ÉGLISE, APPARTEMENT, ...) – en cela il s'agit d'une polysémie systématique – mais la définition de la lexie dérivée sera différente selon que les deux lexies seront reliées par un lien de synecdoque forte ou faible.

3 Conclusion

Le travail qui vient d'être présenté doit mener à terme à la constitution d'une typologie des liens de polysémie du français. La classification des différents types de structures polysémiques d'une langue a déjà fait l'objet de plusieurs travaux, notamment sur le français (Martin, 1972a, 1972b) et sur le russe (Apresjan, 1974), mais ces travaux ne reposaient pas sur une exploitation systématique de descriptions lexicales formelles du type de celles qui sont fournies par la BDéf.

habitation. L'opérateur optionnel `partie de` permet de rendre plus général le patron. Pour une présentation détaillée de la notion d'étiquette sémantique dans les lexiques Sens-texte, voir (Polguère, 2003).

Cette recherche présente l'intérêt de contribuer à la rationalisation de la base de données, le but étant d'obtenir des définitions suffisamment homogènes et structurées pour mettre en relation les informations de la BDéf avec celles du DiCo, qui est quant à elle déjà entièrement informatisée (Steinlin *et al.*, 2004). L'exploitation de la BDéf vise par ailleurs à produire un ensemble de patrons de polysémie qui pourront par exemple servir de base à l'élaboration de règles pour le développement de lexiques incrémentaux du type du Lexique Génératif (Copestake et Briscoe, 1995), (Rappaport et Levin, 1998).

Références

- Altman J., Polguère A. (2003), La BDéf : base de définitions dérivée du Dictionnaire explicatif et combinatoire, *Proceedings of the First International Conference on Meaning-Text Theory*, Paris, p. 43-54.
- Apresjan J. (1974), Regular Polysemy, *Linguistics*, 142, p. 5-32.
- Aristote (2004), *Les Topiques*, Librairie Philosophique J.Vrin, Paris.
- Bouillon P., Busa F. (2001), *Generativity in the Lexicon*, Cambridge University Press, Cambridge.
- Copestake A., Briscoe E. (1995), Semi-Productive Polysemy and Sense extension, *Journal of Semantics*, 12(1), p. 15-67 .
- Dowty D. (1979), *Word Meaning and Montague Grammar*, Reidel, Dordrecht.
- Kleiber G. (1999), *Problèmes de sémantique : la polysémie en question*, Presses Universitaires du Septentrion, Villeneuve d'Ascq.
- Martin R. (1972), Esquisse d'une analyse formelle de la polysémie, *Travaux de linguistique et de littérature*, 10, p. 125-136.
- Martin R. (1972), La polysémie verbale, esquisse d'une typologie formelle, *Travaux de linguistique et de littérature*, 17, p. 261-256.
- Mel'čuk I., Clas A., Polguère A. (1995), *Introduction à la lexicologie explicative et combinatoire*, AUPELF-UREF/Duculot, Louvain-la-Neuve.
- Mel'čuk I. *et al.* (1984, 1988, 1992, 1999), *Dictionnaire explicatif et combinatoire du français contemporain*, vol I-IV, Les Presses de l'Université de Montréal, Montréal.
- Polguère A. (2000), Towards a theoretically-motivated general public dictionary of semantic derivations and collocations for French, *Proceedings of EURALEX'2000*, Stuttgart, p. 512-527.
- Polguère A. (2003), Étiquetage sémantique des lexies dans la base de données DiCo, *TAL*, 44, 2, p. ?.
- Pottier B. (1974), *Linguistique générale. Théorie et description*, Klincksieck, Paris.
- Pustejovsky J. (1995), *The Generative Lexicon*, MIT Press, Cambridge, Mass.
- Rappaport Hovav M., Levin B. (1998), Building Verb Meanings, in M. Butt and W. Geuder (eds) *The projection of arguments : Lexical and Compositionnal Factors*, CSLI Publications, Stanford, p. 97-134.
- Rastier F. (1987), *Sémantique interprétative*, Presses Universitaires de France, Paris.
- Steinlin J., Kahane S., Polguère A., El Ghali A. (2004), De l'article lexicographique à la modélisation objet du dictionnaire et des liens lexicaux, *Proceedings of EURALEX'2004*, Lorient, p. 177-186.
- Victorri B., Fuchs C. (1996), *La polysémie, construction dynamique du sens*, Hermès, Paris.
- Wierzbicka A. (1987), *English Speech Act Verbs : A semantic dictionary*, Academic, Sydney.