Computing semantic relations on structured lexical definitions

Lucie Barque - Alexis Nasr
LATTICE-CNRS (UMR 8094)
Université Paris 7
lbarque@linguist.jussieu.fr, alexis.nasr@linguist.jussieu.fr

1 Introduction

The BDéf was introduced in (Altman and Polguère, 2003) as a formal database derived from the Explanatory Combinatorial Dictionnary of Contemporary French (ECD). One of the major features of the BDéf entries is the formal description of the internal structuring of lexical meanings. The need to precisely describe such a structuring led to the definition of a formal metalanguage for lexicographic definitions. The work presented in this paper is based on the BDéf approach and shows how the BDéf definitions can be used in order to compute semantic relations between lexical units.

The structure of the paper is the following : in section 2 we describe the grammar of the BDéf metalanguage, in section 3 we show how semantic relations between lexical units can be modelled using the decomposition proposed by the BDéf. We illustrate this, in section 4, on the antonymy relation. Section 5 concludes the paper.

2 Formalizing the BDéf definitional metalanguage

In this section we quickly review the structure of a BDéf definition, as introduced in (Altman and Polguère, 2003); we then propose to describe the BDéf definitional metalanguage by means of a formal grammar and to represent the entries as typed feature structures (Carpenter, 1992), (Copestake and Briscoe, 1995). The grammar allows us to produce feature structures as the output of a parsing process that takes as input the definitions. The feature structures produced are then used in a calculus.

We will illustrate the BDéf metalanguage on an example, the French verb APPLAUDIR**I** (*to applaud*), whose definition, translated into English, is represented below, in a linear form, as it might appear in the ECD (Mel'čuk *et al.* 1984, 1988, 1992, 1999,). Its BDéf definition appears on the left in figure 1.

X applaud Y for Z ≡ ' the person X produces a meaningful sound by means of clapping hands that express that X congratulates the person Y for his performance Z'

The BDéf definition of the verb APPLAUDIR**I** decomposes the ECD definition into smaller parts (propositions and groups of propositions) and describes explicitly the relations that hold between these parts.

APPLAUDIR **I**

Propositional Form

X ∼ Y pour Z

Semantic label

son expressif

DEFINITION

Central Component

/*son expressif*/
1: X produire son_expressif

Definition body

/*Contenu*/
2: *1 exprimer *3
3: X complimenter Y pour Z

/*Manière*/
4: *1 se_faire_au_moyen_de *5
5: X taper_dans_les_mains

Actant typing

X: personne
Y: personne
Z: performance

LU	APPLAUDIR.I			
P-FORM	[1] [NAME X, TYPE personne], [2] [NAME Y, TYPE personne], [3] [NAME Z, TYPE performance]			
SEM-LAB	son_expressif			
DEF	CENTRAL-COMP	LABEL	son_expressif	
		FIRST-PROP	[4] [PRED son_expressif, PRED-ARG [1]]	
	BODY	LABEL	contenu	
		FIRST-PROP	[5] [PRED exprimer, PRED-ARG [4], [6]]	
		CONTENT	[6] [PRED complimenter, PRED-ARG [1], [2], [3]]	,
		LABEL	manière	
		FIRST-PROP	...	

Figure 1: The BDéf description of the French verb APPLAUDIR**I** (left) and its representation as a feature structure (right)

More precisely, a BDéf entry describing the meaning of a lexical unit L consists of four parts, a propositional form, which introduces the actants of L, a semantic label, which encodes the general meaning of the unit, the definition proper, which constitutes the main part of the semantic description of the unit and the actant typing section which attributes a type to each actant introduced in the propositional form[1]. The feature structures defining the types **bdef entry** and **lu argument** are presented below.

bdef entry	
LU	**lexical unit**
PROP-FORM	list_of **lu argument**
SEM-LAB	**semantic label**
DEF	**definition**

lu argument	
NAME	**var**
TYPE	**semantic label**

The definition itself comprises two parts : the central component and the definition body, which respectively correspond to Aristotle notions of *genus* and *differentiae* (Aristote, ed. 2004). Briefly, the central component represents the general meaning of the defined lexical unit (also represented, but more concisely, by the semantic label). The definition body represents components (*differentiae*) which specify further the central component and distinguish the lexical unit from other units that share the same general meaning. Both the central component and the definition body are made of blocks (one for the central components and several for the definition body), as shown below in the feature structure that defines the type **definition**.

definition	
CENTRAL-COMP	**block**
BODY	list_of **block**

block	
LABEL	**block label**
FIRST-PROP	**first proposition**
CONTENT	list_of **proposition**

[1]A BDéf entry contains another section describing the semantic relations (if any) that hold between the actants. For instance, the actant Y of APPLAUDIR is the first actant of the predicate Z that denotes a performance. Due to space limitations we will ignore this part of the description here.

A block represents an "autonomous" component of the definition, which means that one can remove a block from a definition and keep the definition well-formed whereas removing a part of it does not. Each block is tagged with a block label (between `/* */` signs in the definition, see figure 1). Such a label accounts for the informational purpose of the block. In our example, the second block of the definition body specifies *how* an applause is performed.

A block itself contains a list of propositions. Each of them is numbered so that it can be refered to anywhere in the definition. A proposition is made of a predicate and its arguments which can be a semanteme, a reference to another proposition, or an actant introduced in the propositional form. For instance, in figure 1 the proposition number 2 contains the predicate `exprimer` and its two arguments : a pointer to (the predicate of) proposition 1 and a reference to (the predicate of) proposition 3. Propositions may also include modifiers of the predicate (for example a negation adverb) and support verbs, when the predicate is not a verb (for example, in proposition number 2, `produire` is a support verb of the main predicate `son_expressif`.). The first proposition of a block plays a special role : its predicate is tightly linked to the block label.

first proposition	
PRED	**block indicator predicate**
PRED-ARG	list_of **pred argument**

proposition	
PRED	**semanteme**
MOD	**semanteme**
ARG-STRUCT	list_of **pred argument**

The "lexical units" of the metalanguage are described in another database[2] where the semantemes used in the propositions are associated with a lexical unit. For example the configuration of semantemes (also called BDéf word) `se_faire_au_moyen_de` will be associated with the lexical unit MOYEN in order to access its meaning. The description also accounts for the function of the unit in the definition. For instance, `se_faire_au_moyen_de` is a block indicator predicate of the block `manière`. The labels (the semantic label and the block labels), which belong to the label hierachy, are described in another part of the BDéf lexicon (Polguère, 2003a).

The conditions of structural well-formedness of a definition are represented as a context free grammar, more specifically as an XML Document Type Definition (DTD). This DTD defines a set of XML tags as well as the rules that describe the structure of the definitions. It is therefore possible to check the structural well-formedness of a given definition using standard XML parsers. A definition is represented as an XML document[3] and it is also possible, using standards tools, to produce the original representation for the entry, as it appears on the left side of figure 1, from its XML representation.

To be semantically well-formed, a definition must verify some "semantic" constraints. Let's mention two of them : first, the labels of the blocks that can appear in the definition of L are controlled by its semantic label. In our example, the semantic label `son expressif` accounts for the occurrence of the two blocks `Contenu` (what does X mean when he produces this sound) and `Manière` (how does X produce this sound). Second, the label of a block controls the predicate of its first clause : the predicate must express the informational purpose of the block.

The semantic tags, block labels, first proposition predicates and the relations that hold between them are represented in another XML document whose structure is described in another DTD.

As noted, the lexical entries are represented as typed feature structures in order to carry out computations on them. The feature structure representation of the definition[4] of APPLAUDIR**I** is

[2]The development of this database is in progress at OLST, Université de Montréal.

[3]Space limitations preclude us from presenting here an example of a definition in the XML format.

[4]The representation of a definition as a typed feature structure can be produced automatically from the XML representation of a definition.

represented on the right in figure 1. This means of representation allows us to use the operation of unification for our calculus, as in (Pustejovsky, 1995) and (Copestake and Briscoe, 1995). The following sections present one of them : to check whether a given semantic relation exists between two given lexical units.

3 Semantic relations

3.1 Definition

Semantic relations are defined in (Polguère, 2003b) in terms of set-theoretic operations on lexical definitions. A lexical unit L is seen as a structure based on a set of simpler lexical meanings (the semantemes that appear in the definition of L). A semantic relation holds between two lexical units L_1 and L_2 if there is either identity, inclusion or a non null intersection between the definitions of L_1 and L_2. Such relations are seen as basic semantic relations in terms of which more complex relations, as hyperonymy/hyponymy, synonymy, antonymy, . . . are built.

As we have seen in section 2, the BDéf definition of a lexical unit describes precisely its lexical meaning in terms of other lexical units, uncovering the "organization" that is mentionned in the definition above. This explicit decomposition allows for a precise definition of lexical relations : the decomposition of the definitions in terms of blocks, propositions and predicates, and their representation as feature structure enable the description of a lexical relation as a pair of underspecified feature structures. Such feature structures describe the properties that two lexical units must have for a relation to hold between them. The fine grained level of decomposition of the definitions allows for a perspicuous definition of relations. One can, for example, define a relation between two lexical units by referring to specific blocks, propositions or predicates of their definitions.

3.2 Defining lexical relations as feature structure pairs

We define a lexical relation R as a couple of feature structures S_1 and S_2 (we note $R = \langle S_1, S_2 \rangle$)[5]. The relation R holds between two lexical units L_1 and L_2 (we note $R\langle L_1, L_2 \rangle$) if $*L_1$ unifies with S_1 and $*L_2$ with S_2, or the opposite, where $*L$ denotes the feature structure that represents L's definition. In other words, the operator '$*$' is used to access the semantic decomposition of a lexical unit.

$$R\langle L_1, L_2 \rangle \Leftrightarrow (S_1 \sqcup *L_1 \ \wedge \ S_2 \sqcup *L_2) \ \vee \ (S_1 \sqcup *L_2 \ \wedge \ S_2 \sqcup *L_1)^6$$

The type of relation defined above is called direct relation since it is directly defined on two lexical units. Indirect relations between lexical units L_1 and L_2 imply that another relation R (direct or indirect) exists between components of L_1 and L_2 (we note $IR = \langle S_1, S_2 \rangle \wedge R\langle S_1.x, S_2.y \rangle$ where $S_1.x$ and $S_2.y$ are paths in the feature structures S_1 and S_2 leading to semantemes and R is a semantic relation). Checking for the definition of an indirect relation IR between L_1 and L_2 is defined below :

$$IR\langle L_1, L_2 \rangle \Leftrightarrow ((S_1 \sqcup *L_1 \ \wedge \ S_2 \sqcup *L_2) \ \vee \ (S_1 \sqcup *L_2 \ \wedge \ S_2 \sqcup *L_1)) \ \wedge R\langle L_1.x, L_2.y \rangle$$

[5] R is actually defined as a feature structure having S_1 and S_2 as substructures. We have chosen to represent it here as a couple for readability reasons.

[6] Unification is seen here as a logical predicate. The expressions $A \sqcup B$ is true if A and B unify.

4 Computing antonymy

Two lexical items L_1 and L_2 are in an antonymy relation if they share the same part of speech and if some components of the definition of L_1 are negated or, more generally "put in opposition" in the definition of L_2 or the opposite (Polguère, 2003b).

When formally defining (as a pair of feature structures) the antonymy relation, many subcases can be distinguished according to the position that the negation occupies in the two feature structures. We describe below three types of antonymy which correspond to three different positions of the negation in the patterns[7]. Note that we only consider the cases where negation is located in the central component of the definiton[8].

Subtype 1 : Inclusive antonymy In this case, a lexical unit appears, modified by a negation, in the definition of another one[9], as in the example below, where the definition of RÉUSSIR[10] (*to succeed*) is included in the definition of ÉCHOUER (*to fail*).

$$\begin{bmatrix} \text{LU} & \text{ÉCHOUER} \\ \text{DEF} & \begin{bmatrix} \text{C-COMP} & \begin{bmatrix} \text{FIRST-PROP} & \begin{bmatrix} \text{PRED} & \text{réussir} \\ \text{MOD} & \text{ne\_pas} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

The definition of this type of antonymy is represented below :

$$Anti_1 = \left\langle \begin{bmatrix} \text{LU} \ \boxed{1} \end{bmatrix}, \begin{bmatrix} \text{DEF} & \begin{bmatrix} \text{C-COMP} & \begin{bmatrix} \text{FIRST-PROP} & \begin{bmatrix} \text{PRED} & \boxed{1} \\ \text{MOD} & \text{ne\_pas} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle$$

Subtype 2 : Intersective antonymy In this case, two lexical units share a predicate, that is negated in one of them. An example of this subtype is presented below with the pair PROVOQUER (*to cause*) $\sim$ EMPÊCHER (*to prevent*)[11].

Lexical unit X provoquer Y		**Lexical unit** X empêcher Y	
Central component		**Central component**	
1: X causer *2	(*1: X cause *2)*	1: X causer *2	(*1: X cause *2)*
2: Y avoir_lieu	(*2: Y happen*)	2: Y ne_pas avoir_lieu	(*2: Y not happen*)

The definition of intersective antonymy is represented below :

[7]This typology is mainly inspired by the typology of antonymy presented in (Apresjan, 1992) and (Milićević, 2003). The reader should note that antonymy has been chosen here as a case in point. For the detailed literature on this this type of semantic relation, see among others, (Apresjan, 1992), (Cruse, 1986), (Lehrer and Lehrer, 1982), (Lyons, 1995).

[8]For instance, pairs of antonyms like APPLAUDIR$\sim$SIFFLER (*applaud*/*boo*) are not taken into account here. Indeed, they share the same central component (`X produire son expressif`) while the opposition takes place in the body of the definition, more precisely in the `Contenu` block (X expresses that he is happy *vs* X expresses that he is not happy)

[9]More precisely, the first lexical unit must appear as the predicate of the first proposition of the central component of the other lexical unit.

[10]We have only represented the relevant part of the definition.

[11]Due to lack of space, we represent the central component of these two lexical units in the BDéf formalism and not as feature structures.

$$Anti_2 = \left\langle \begin{bmatrix} \text{DEF} & \begin{bmatrix} \text{C-COMP} & \begin{bmatrix} \text{FIRST-PROP} & \begin{bmatrix} \text{PRED} \ \boxed{4} \\ \text{PRED-ARG} \ \boxed{1} \end{bmatrix} \\ \text{PROP} \ \boxed{1} & \begin{bmatrix} \text{PRED} \ \boxed{2} \\ \text{MOD} \ \emptyset \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}, \begin{bmatrix} \text{DEF} & \begin{bmatrix} \text{C-COMP} & \begin{bmatrix} \text{FIRST-PROP} & \begin{bmatrix} \text{PRED} \ \boxed{4} \\ \text{PRED-ARG} \ \boxed{3} \end{bmatrix} \\ \text{PROP} \ \boxed{3} & \begin{bmatrix} \text{PRED} \ \boxed{2} \\ \text{MOD} \ \text{NE\_PAS} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle$$

Subtype 3 : Indirect antonymy In this case, the central component of one lexical unit includes a predicate which is in an antonymy relation with a predicate of the other lexical unit[12]. These two predicates must be located at the same place in the definition. If they appear in a subordinate proposition, they should share the same governor. The pair EXHIBER (*to exhibit*) $\sim$ ESCAMOTER (*to make sth vanish*) illustrates this third subtype.

Lexical unit X exhibe Y	**Lexical unit** X escamoter Y
Central component	**Central component**
1: X causer *2 (*1: X cause *2)*	1: X causer *2 (*1: X cause *2)*
2: Y apparaître (*2: Y appear*)	2: Y disparaître (*2: Y disappear*)

The definition of indirect antonymy is represented below :

$$Anti_3 = \left\langle \begin{bmatrix} \text{DEF} & \begin{bmatrix} \text{C-COMP} & \begin{bmatrix} \text{FIRST-PROP} & \begin{bmatrix} \text{PRED} \ \boxed{5} \\ \text{PRED-ARG} \ \boxed{1} \end{bmatrix} \\ \text{PROP} \ \boxed{1} & \begin{bmatrix} \text{PRED} \ \boxed{2} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}, \begin{bmatrix} \text{DEF} & \begin{bmatrix} \text{C-COMP} & \begin{bmatrix} \text{FIRST-PROP} & \begin{bmatrix} \text{PRED} \ \boxed{5} \\ \text{PRED-ARG} \ \boxed{3} \end{bmatrix} \\ \text{PROP} \ \boxed{3} & \begin{bmatrix} \text{PRED} \ \boxed{4} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle$$

$$\wedge \ Anti_2 \left\langle * \boxed{2}, * \boxed{4} \right\rangle$$

5 Conclusions and perspectives

We have presented in this paper a methology that consists in modelling semantic relations by means of BDéf definitions. The formal patterns that result of this modelling are closely related to lexical functions (LF), another formalism developped by the *Explanatory Combinatorial Lexicology* to model semantic relations in the lexicon. Formalization based on BDéf descriptions may help to tackle issues like the control of LF's values, the organization of the set of an LF's values, etc. In this paper we have used the relation patterns to check for the occurrence of a relation R between two lexical units L_1 and L_2. Such patterns can also be used in another way : having the definition of L_1 and the relation R, we can automatically produce an (underpecified) definition of L_2. This technique could be used by the lexcographer when producing new entries.

Acknowledgements

We are grateful to Sylvain Kahane and Alain Polguère for helpful comments on the preliminary draft of this paper.

[12]Note that this third subtype has to be taken into account with respect to the computations on definition but may be not relevant linguistically.

References

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. (1992), *Lexical Semantics : User's Guide to Contemporary Russian Vocabulary*, Ann Arbor.

Aristote (2004), *Les Topiques*, Librairie Philosophique J.Vrin, Paris.

Carpenter B. (1992), *The logic of typed feature structure*, Cambridge University Press, Cambridge, England.

Copestake A., Briscoe E. (1995) Semi-Productive Polysemy and Sense extension *Journal of Semantics*, 12(1), p.15-67.

Cruse D.A. (1986) *Lexical semantics*, Cambridge University Press, New York.

Lehrer A., Lehrer K. (1982), Antonymy, *Linguistics and phylosophy*, 5, p.483-501.

Lyons J. (1995), *Linguistics semantics*, Cambridge University press, London.

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.

Milićević, Jasmina. (2003), *Étude des aspects sémantiques et syntaxiques de la paraphrase : application à la génération automatique des phrases françaises.*, Thèse de doctorat, Département de linguistique et de traduction, Université de Montréal.

Polguère A. (2003a), Étiquetage sémantique des lexies dans la base de données DiCo, *TAL*, 44, 2, p.39-68.

Polguère A. (2003b), *Lexicologie et sémantique lexicale*, Presses de l'université de Montréal, Montréal.

Pustejovsky J. (1995), *The Generative Lexicon*, MIT Press, Cambridge, Mass.