

Université Paris 7
UFR Linguistique

DEA de linguistique théorique descriptive et formelle
Option linguistique informatique
Septembre 2003

Opérations sémantiques sur une base de données de
définitions Sens-Texte

Lucie Barque

Sous la direction d'Alexis Nasr

Remerciements

Je remercie tout particulièrement Alexis Nasr de m'avoir aiguillée tout au long de ce travail.

Je remercie aussi Alain Polguère d'avoir apporté des réponses à mes questions et Jacques et mes miens de m'avoir encouragée.

Je remercie enfin Laurent et François de m'avoir aidée à relifter ce mémoire et plus généralement tous les membres du laboratoire Talana d'être si accueillants.

Table des matières

Introduction	4
1 Représentation du sens lexical	7
1.1 Le dictionnaire de la Théorie Sens-Texte	8
1.1.1 La place du dictionnaire dans un modèle Sens-Texte .	8
1.1.2 Présentation du Dictionnaire Explicatif et Combinatoire	15
1.2 Autres approches lexicographiques	18
1.2.1 WordNet	18
1.2.2 FrameNet	20
1.2.3 Le Lexique Génératif	23
1.3 Décomposition du sens et relations lexicales	25
1.3.1 Décomposition du sens lexical	25
1.3.2 Relations lexicales et composantes de sens	27
2 Le formalisme de la BDéf	30
2.1 Présentation de la BDéf	31
2.1.1 Les précurseurs du formalisme de la BDéf : I.Mel’čuk et A.Wierzbicka	31
2.1.2 Une fiche de la BDéf	33
2.1.3 Les deux notions piliers du formalisme de la BDef : la proposition élémentaire et le bloc définitionnel	36
2.2 Contrôle du formalisme de la BDéf	39
2.2.1 Contrôle des définitions	39
2.2.2 Une aide à l’encodage des données	41
2.3 Une grammaire formelle du métalangage de la BDéf	42
2.3.1 Structure de la composante centrale	42
2.3.2 Structure des autres blocs définitionnels	43
2.3.3 Grammaire formelle	48
3 Un modèle calculable	51
3.1 calculs sémantiques sur les définitions	53
3.1.1 Les opérations élémentaires	53
3.1.2 Utilisation des calculs	54

3.2	Une étape en vue d'autres calculs	57
3.2.1	Transparence du lien entre le sens d'une lexie et sa combinatoire lexicale	57
3.2.2	Un formalisme approprié	58
	Conclusion	61
	Références bibliographiques	62
	A fiches BDef	64
A.1	La fiche BRAS#I.1a	64
A.2	La fiche BRAS#I.1b	65

Introduction

Le langage de la définition lexicographique est un métalangage, c'est-à-dire un langage conventionnel utilisé pour décrire un des aspects d'une langue naturelle.

Ce métalangage peut être plus ou moins artificiel c'est-à-dire qu'il peut s'apparenter plus ou moins à la langue naturelle dont il décrit le lexique.

Le métalangage utilisé dans les dictionnaires de langue traditionnels semble peu artificiel comme nous le montre cette définition du Petit Robert.

PASSOIRE n.f. ♦ Récipient percé de trous et utilisé pour égoutter des aliments, pour filtrer sommairement des liquides. *Égoutter des pâtes dans une passoire.*

Petite passoire. ⇒ **chinois, passette ; passe-thé.**

FIG *Sa mémoire est une passoire, ne retient rien.*

Si l'on s'en tient à la définition proprement dite (**PASSOIRE** ♦ Récipient percé de trous et utilisé pour égoutter des aliments, pour filtrer sommairement), il est difficile d'identifier clairement ce qui relève du métalangage définitionnel. Certaines relations lexicales ne sont pas explicitées : le premier mot de la définition (**récipient**) est en relation d'hyperonymie avec la mot vedette (**passoire**) mais cette relation n'est représentée par aucune construction dans la définition. Les unités de la définition qui relèvent du métalangage ne sont en outre pas homogènes : la construction **utilisé pour** indique que le **récipient** a une fonction. Il a une seconde fonction (qui peut être vue comme une généralisation de la première) qui n'est – probablement dans le but d'éviter une répétition – introduite que par la construction **pour**.

Ce type de métalangage a bien sûr l'avantage d'être intuitif : le lecteur comprend immédiatement, justement en raison du caractère conventionnel du langage définitionnel, que **passoire** est un type de **récipient** et que, en raison de son type, **passoire** à une fonction précise.

Il n'est en revanche plus approprié dès lors que l'on souhaite mettre en lumière la structure d'une définition ou en autoriser le traitement automatique.

Ce mémoire va s'intéresser à une théorie particulière de la représentation du sens lexical, la Lexicologie Explicative et Combinatoire (LEC) qui forme

la composante lexicale de la Théorie Sens-Texte (Mel'čuk[95]). Nous allons nous intéresser plus particulièrement à un formalisme associé à cette représentation, celui de la BDéf, base de données lexicales qui encode les définitions lexicographiques Sens-Texte (Altman&Polguère[03]).

Le but principal de ce formalisme est de proposer une structure définitionnelle sur laquelle il soit possible d'effectuer des calculs sémantiques.

Le terme de **calcul sémantique** n'est pas à prendre au sens où on l'entend d'ordinaire dans le domaine de la sémantique formelle : Il ne s'agit pas ici de calculer la valeur de vérité d'une définition en la confrontant à une représentation du monde. Il s'agit plutôt d'opérer des calculs sur la représentation lexicale elle-même (par exemple calculer le degré de synonymie entre deux unités lexicales par comparaison de leurs définitions) qui seront utiles pour la recherche en sémantique lexicale.

L'objectif de ce mémoire est de voir si le formalisme de la BDéf est effectivement approprié pour ce type de calculs sémantiques. Nous décrirons, au moyen d'une grammaire formelle, les définitions BDéf et nous verrons quels sont les calculs pouvant être effectués sur la représentation abstraite issue de cette analyse.

Notre but n'est pas de décrire objectivement le métalangage de la BDéf – qui n'est encore qu'en développement – mais bien d'adopter dans la description un point de vue prescriptif afin que ce métalangage ait toutes les caractéristiques que l'on est en droit d'attendre d'un métalangage (identifiable, non ambigu, etc).

Le premier chapitre va s'intéresser à la question de la représentation du sens lexical.

La représentation lexicale proposée par la LEC fera l'objet de la première section : nous présenterons la structure d'un modèle Sens-Texte et nous verrons la place qu'y occupe le dictionnaire, produit de la LEC.

Cette représentation lexicale sera ensuite mise en regard avec plusieurs autres représentations lexicales, celles du Lexique Génératif (Pustejovsky[95]), de WordNet (Miller[90]) et de FrameNet (Fillmore[76]). Nous tenterons de montrer les alternatives possibles au traitement des aspects du sens lexical auxquels la LEC s'intéresse.

La troisième section considérera le lien que certaines de ces théories établissent entre la décomposition du sens d'une unité lexicale et les relations que celle-ci entretient avec les autres unités du lexique.

Le deuxième chapitre sera consacré entièrement à la description du formalisme de la BDéf au moyen d'une grammaire formelle.

Nous présenterons dans la première section ce formalisme en insistant sur les deux notions principales qui le sous-tendent, les notions de *bloc définitionnel élémentaire* et de *proposition élémentaire*, l'une et l'autre participant du be-

soin d'expliciter la structure définitionnelle.

Nous expliquerons dans une deuxième section l'intérêt de décrire ce formalisme au moyen d'une grammaire formelle.

La troisième section sera consacrée à l'élaboration de cette grammaire.

Enfin le troisième chapitre donnera un aperçu des calculs sémantiques envisageables à partir de la représentation abstraite obtenue au terme de l'analyse d'une définition BDéf.

La première section décrira les calculs élémentaires qui peuvent être effectués sur la structure abstraite des définitions et montrera l'intérêt de ces calculs pour la recherche en sémantique lexicale.

La seconde section évoquera la possibilité d'utiliser ces calculs dans l'effort de mise en évidence des liens entre la définition d'une lexie et les autres aspects de sa description.

Chapitre 1

Représentation du sens lexical

Nous allons nous intéresser, au cours de ce premier chapitre, à la question de la représentation lexicale, plus précisément aux trois questions suivantes :

- Quels sont les différents aspects du sens lexical pris en compte dans la représentation proposée par une théorie sémantique ?
- Quels liens peut-on établir entre ces différents aspects ?
- Sous quelle forme ces différents aspects ainsi que les liens qui les unissent sont-ils représentés ?

La représentation lexicale qui nous intéressera en premier lieu est celle proposée par la Lexicologie Explicative et Combinatoire (LEC), la composante lexicale de la Théorie Sens-Texte.

Créée dans les années soixante à Moscou par A. Žolkovskij et I. Mel'čuk, La Théorie Sens-Texte décrit un type de modèle linguistique modélisant le passage d'un Sens vers l'ensemble des énoncés paraphrastiques (les Textes) qui expriment ce Sens. Le système qui permet d'activer ce modèle, **la grammaire Sens-Texte**, fait appel aux informations lexicales encodées dans le **dictionnaire**.

La première section aura pour but de montrer dans quelle mesure le dictionnaire d'un modèle Sens-Texte entretient dans les calculs permettant ce passage du Sens vers les Textes. Il nous faudra pour cela décrire les trois aspects du sens lexical traités par la LEC : la description du sens d'une unité lexicale au moyen d'une décomposition, la description de sa combinatoire syntaxique et enfin celle de sa combinatoire lexicale restreinte.

La seconde section sera consacrée à la présentation de plusieurs théories de sémantique formelle proposant une représentation du sens lexical. Bien qu'ayant des objectifs linguistiques distincts, nous tenterons de voir en quoi ils se rejoignent concernant la représentation du sens lexical.

La troisième section s'intéressera aux liens que l'on peut établir entre le sens d'une unité lexicale et les relations lexicales qu'elle entretient avec

les autres unités.

1.1 Le dictionnaire de la Théorie Sens-Texte

1.1.1 La place du dictionnaire dans un modèle Sens-Texte

Nous allons présenter un modèle Sens-Texte en insistant sur la place centrale qu'occupe le dictionnaire au sein de ce modèle.

Nous ne présenterons d'un modèle Sens-Texte que ce qui sera nécessaire à la compréhension de la suite de l'exposé. Cette présentation sera elle-même très générale et nous renvoyons donc le lecteur à Nasr[96], Mel'cuk[97] et Polguère[98] pour une introduction complète à la TST.

Présentation d'un modèle Sens-Texte

Un modèle Sens-Texte tente de modéliser l'association que tout locuteur est capable de faire entre un Sens et l'ensemble des Textes (un Texte est une suite de phonèmes) ayant ce Sens.

La figure 1.1, reprise de Polguère[98], présente les principales composantes d'un modèle Sens-Texte.

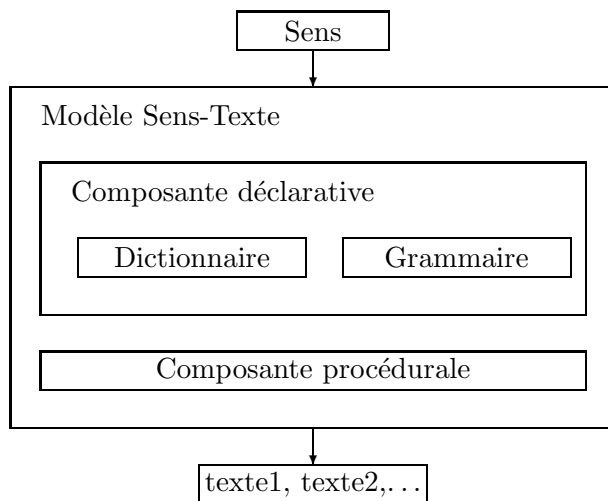


FIG. 1.1 – Modèle Sens-Texte

Un modèle linguistique Sens-Texte prend en entrée un sens et produit en sortie un ensemble de paraphrases, c'est-à-dire un ensemble d'énoncés ayant le même sens linguistique.

La composante procédurale contient les instructions nécessaires à l'activation du modèle Sens-Texte. Nous ne nous intéresserons ici qu'à la com-

posante purement linguistique du modèle, la composante déclarative qui contient la grammaire et le dictionnaire.

La composante déclarative

La **grammaire** regroupe l'ensemble des règles nécessaires aux correspondances entre les sept **niveaux de représentation** adoptés par la TST. Le passage de la représentation du Sens (représentation sémantique) à la représentation du Texte (représentation phonologique) étant trop difficile à établir, La TST préconise, comme la plupart des théories linguistiques formelles, des niveaux de représentation intermédiaires.

La grammaire est ainsi subdivisée en **composantes**, une composante regroupant l'ensemble des règles qui permettent le passage entre deux niveaux de représentation adjacents.

Les règles d'une composante sont plus exactement de deux sortes :

Les règles transformationnelles sont les règles qui agissent au même niveau de représentation.

Les règles dérivationnelles sont les règles qui permettent de passer d'un niveau de représentation à un autre.

Le **dictionnaire**, second module de la composante déclarative, occupe une place centrale dans le modèle Sens-Texte puisque c'est lui qui fournit une partie des informations requises par chaque composante et qui permet donc d'instancier les règles dans le processus de synthèse.

La figure 1.2 représente les sept niveaux de représentation postulés par la théorie, les composantes permettant le passage d'un niveau à un autre ainsi que la contribution du dictionnaire dans ce passage.

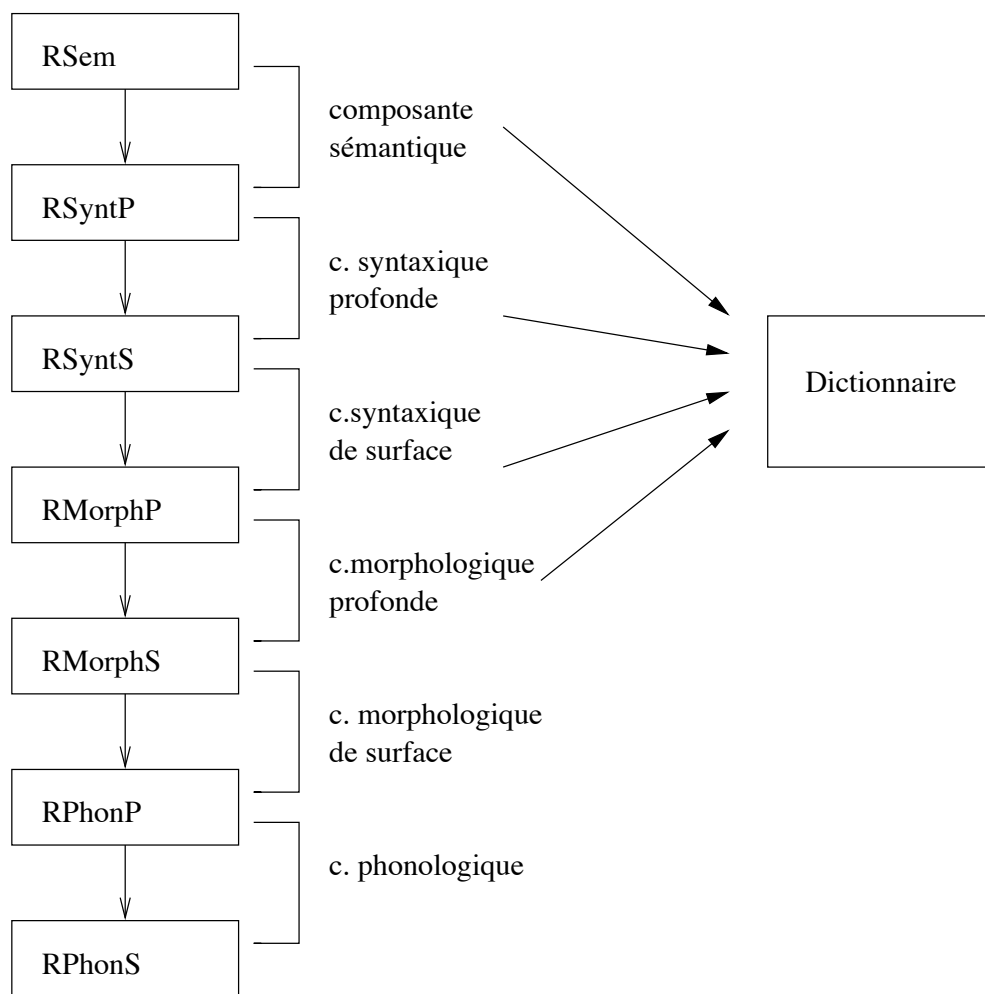


FIG. 1.2 – les niveaux de représentation

Nous choisissons de nous intéresser, pour illustrer le rôle primordial du dictionnaire dans un modèle Sens-Texte, au passage entre le niveau de représentation sémantique et le niveau de représentation syntaxique profond.

Nous décrirons la structure correspondant à chacun de ces niveaux et nous donnerons des exemples de règles permettant de produire des structures de même niveau (des règles transformationnelles).

Nous donnerons également quelques exemples de règles dérivationnelles permettant de produire une structure syntaxique profonde à partir d'une structure sémantique.

Le niveau de représentation sémantique

La structure principale de la représentation sémantique (RSem) est la structure sémantique (SSem). Elle prend la forme d'un graphe connexe dont les nœuds sont étiquetés par des **sémantèmes**. Un sémantème est soit un prédicat (c'est-à-dire un sens qui a besoin d'autres sens, ses actants sémantiques) soit un nom d'objet (un sens qui ne dépend d'aucun autre sens). Les arcs du graphe, qui partent d'un prédicat, sont numérotés selon l'ordre des actants sémantiques du prédicat.

La figure 1.3 représente la SSem de la phrase :

1.1) *La nouvelle de l'assassinat des deux ministres est tombée en Juin.*

et de ses paraphrases :

1.2) *C'est en Juin qu'on a annoncé le meurtre des deux ministres.*

etc. . .

Règles transformationnelles agissant au niveau sémantique :

Une *règle de décomposition sémantique* est une règle transformationnelle qui établit une équivalence entre deux structures sémantiques. Le sémantème de la structure de gauche est décomposé au moyen de sémantèmes plus simples dans la structure de droite.

La figure 1.4 représente (de façon approximative) la règle de décomposition du sémantème *nouvelle* qui pourrait être formulée comme *annonce de X d'un évènement W à l'individu Y, qui sait désormais W*.

Ces règles sont consignées dans la zone sémantique d'un article de dictionnaire où elles sont représentées sous forme linéaire (voir infra 1.1.2).

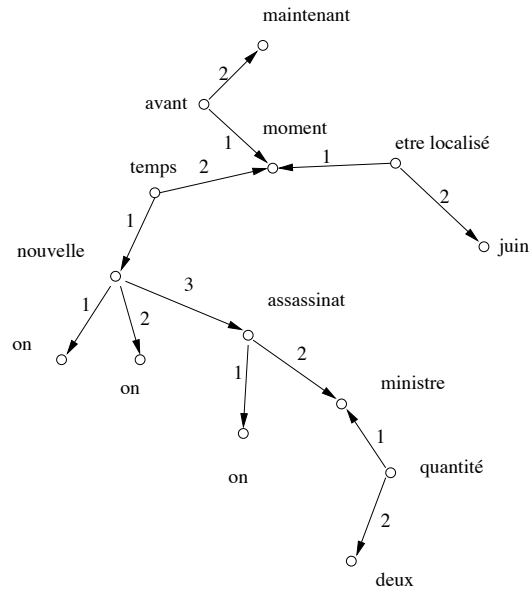


FIG. 1.3 – SSem des phrases 1

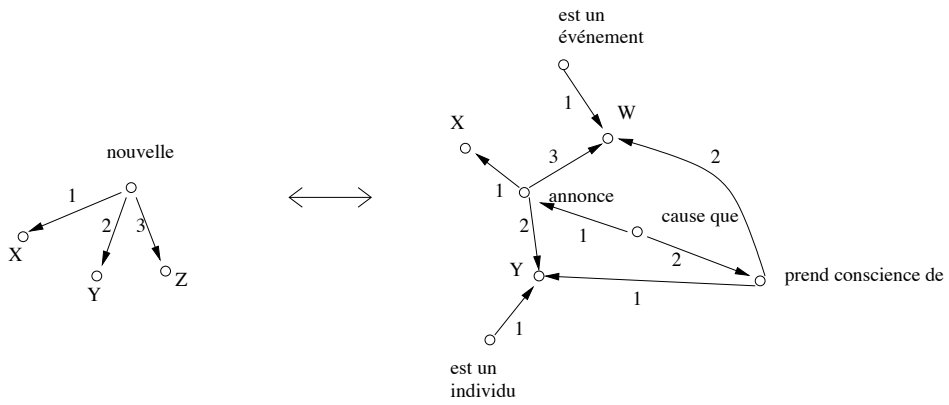


FIG. 1.4 – Une règle de décomposition sémantique

La composante sémantique

Le passage du niveau de représentation sémantique au niveau de représentation syntaxique profond – qui correspond à l’activation des règles de la composante sémantique – consiste en partie à lexicaliser des sémantèmes ou groupes de sémantèmes de la structure sémantique.

La règle de lexicalisation représentée dans la figure 1.5 établit une équivalence entre un sémantème (représenté en lettres minuscules) et la lexie (représentée en lettre capitale) qui lui correspond.

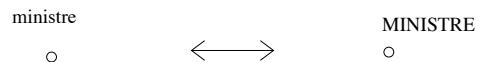


FIG. 1.5 – Une règle dérivationnelle

La règle de lexicalisation représentée dans la figure 1.6 établit une équivalence entre un ensemble de sémantèmes et une lexie¹ qui lui correspond.

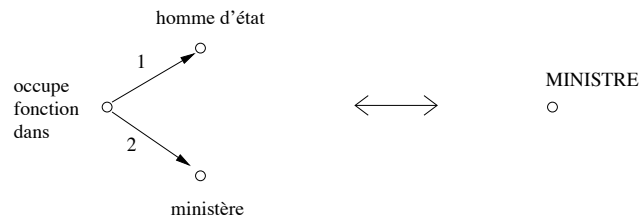


FIG. 1.6 – Une règle dérivationnelle

Le niveau de représentation syntaxique profond

La structure syntaxique profonde (SSyntP) est la structure principale de la représentation syntaxique profonde (RSyntP).

C'est un arbre de dépendance non linéairement ordonné² dont les nœuds sont étiquetés par des **lexies**, c'est-à-dire des unités de la langue prises dans une acception particulière.

¹Le sens de cette lexie est définie par une règle de décomposition. Elle correspond donc à un sémantème mis dans un rapport d'équivalence avec un ensemble de sémantèmes. Il est rare que la partie du réseau sémantique (représentant le sens d'une phrase) que l'on veut lexicaliser corresponde exactement à la décomposition de cette lexie.

Il faut par conséquent réussir à déterminer dans quelle mesure la partie du sens de la phrase correspond à la décomposition du sens de la lexie choisie.

²l'ordre des mots dans la phrase est calculé dans la composante syntaxique de surface

Les arcs de l'arbre syntaxique profond sont étiquetés par les relations syntaxiques profondes qui sont les relations syntaxiques universelles (les relations actancielles (**I,II,...,VI**), la relation attributive (**attr**), la relation appenditive (**append**) et la relation de coordination (**coord**)).

La figure 1.7 présente une structure syntaxique profonde qui peut être générée à partir de la structure sémantique proposée dans la figure 1.3.

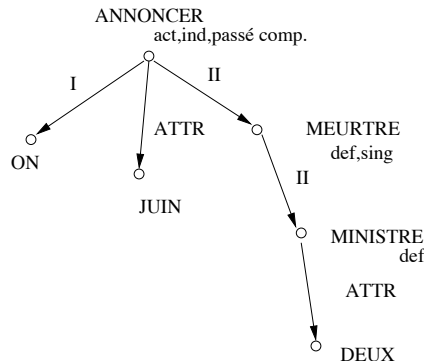


FIG. 1.7 – SSyntP générée à partir de la Ssem de la figure 2.3

Règles transformationnelles agissant au niveau syntaxique profond :

Les **règles de paraphrasage** établissent une équivalence sémantique entre deux structures syntaxiques profondes (Mel'čuk[92], Milićević[03]).

Elles reposent en général sur une ou plusieurs substitutions lexicales représentées en terme de **fonction lexicale** (FL).

Une fonction lexicale s'apparente à une fonction mathématique qui prend en argument un **mot-clé** et renvoie au moins une **valeur**.

$FL(\text{mot-clé}) = \text{valeur}(s)$

Les deux règles lexicales de paraphrasage proposées ci-dessous montrent des exemples de FL. Nous reviendrons de manière plus précise sur cette notion lors de la présentation d'un article du dictionnaire.

règle 1 : $L \Leftrightarrow \text{Syn}(L)$

règle 2 : $L[\text{cat:v}] \Leftrightarrow \text{Anti}(L) \text{ --}\{\text{attr}\}\text{ NE...PAS}$

La règle 1 signifie qu'une lexie pourra être substituée dans l'arbre de droite par un de ses synonymes.

La règle 2 signifie qu'une lexie verbale pourra être substituée par une lexie antonyme qui devra gouverner, par une relation attributive, la lexie 'NE...PAS'.

On peut ainsi, en utilisant la règle 1, produire un deuxième arbre syntaxique profond à partir de celui de la figure 1.7, représenté dans la figure 1.8. Dans l'arbre syntaxique de surface, la fonction lexicale **Syn(meurtre)** sera remplacée par une de ses valeurs, *assassinat* par exemple.

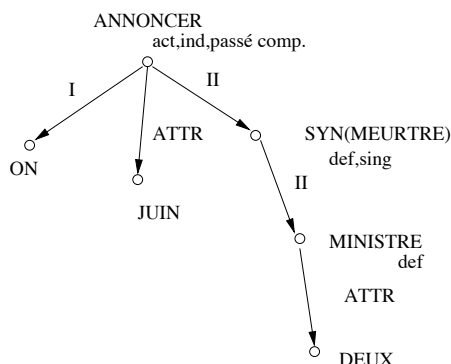


FIG. 1.8 – SSyntP générée à partir de la SSyntP de la figure 1.7

1.1.2 Présentation du Dictionnaire Explicatif et Combinatoire

Le Dictionnaire Explicatif et Combinatoire (DEC) est le dictionnaire théorique décrit par la composante lexicale de la théorie Sens-Texte, la Lexicologie Explicative et Combinatoire (Mel'cuk[95]).

Le dictionnaire est **explicatif** parce qu'il explicite le sens d'une lexie au moyen d'une décomposition utilisant des lexies plus simples. Il est **combinatoire** parce qu'il rend compte d'une part de la combinatoire syntaxique, d'autre part de la combinatoire lexicale d'une lexie.

Avant d'approfondir ces trois aspects de la description lexicale, il nous faut dire deux mots de la macrostructure d'un DEC et définir ainsi les deux notions de **lexie** et de **vocable**.

Une lexie est une unité lexicale monosémique dont la description constitue un article du DEC. Une lexie peut être de deux types :

- Une unité lexicale simple ou **lexème** (aider, ministre, poirier, ...)
- Une expression lexicale figée ou **phrasème** (cordon bleu, tire-au-flanc, voir rouge, ...)

Un vocable regroupe les lexies homographiques ayant entre elles un rapport sémantique c'est-à-dire des lexies qui sont dans une relation de polysémie. Les lexies homographiques qui n'ont pas de rapport sémantique entre elles (les homonymes) sont consignées dans des vocables différents. Un vocable constitue un superarticle regroupant un ou plusieurs articles.

Représentation des informations lexicales d'un article du DEC :

Un article du DEC est subdivisé en quatre zones :

1. La zone sémantique explicite le sens de la lexie vedette sous la forme d'une définition. Rappelons qu'une définition est une version 'Texte' d'une *règle de décomposition sémantique* (voir 1.1.1). Une définition Sens-Texte est, comme dans les dictionnaires de la langue traditionnels, subdivisée en deux parties : le défini et le définissant. Nous donnons ci-dessous la définition d'un article du DEC décrivant une lexie du vocable AIDER.

X aide Y à Z-er par W = l'individu Y étant en train d'effectuer une action Z souhaitable pour Y, l'individu X emploie ses ressources W de sorte que W facilite ou rend possible Z pour Y.

Le défini se présente dans sa forme propositionnelle, c'est-à-dire une formule comprenant la lexie vedette accompagnée, lorsqu'on a affaire à une lexie prédicative, de ses actants sémantiques représentés sous forme de variable (X aide Y à Z-er par W).

Le définissant est une paraphrase analysant le sens du défini au moyen d'unités lexicales plus simples (ce qui se trouve à droite du signe d'égalité sémantique '=').

2. La zone de combinatoire syntaxique se présente sous la forme d'un tableau qui établit la correspondance entre les actants sémantiques de la lexie et ses actants syntaxiques profonds ainsi que la façon dont ils se réalisent en syntaxe de surface.

Le tableau ci-dessous présente les différentes réalisations syntaxiques possibles des actants de la lexie AIDER.

La troisième colonne du tableau indique que l'actant sémantique Z correspond à l'actant syntaxique profond qui entretient une relation de type III avec la lexie vedette.

La seconde ligne indique la nature syntaxique de l'actant syntaxique de surface ainsi que, lorsqu'il y a lieu, la proposition, régie par la lexie vedette, qui introduit cet actant.

3. La zone de combinatoire lexicale se présente sous la forme d'une liste de **fonctions lexicales** (FL).

Le formalisme des fonctions lexicales sert à encoder des informations non prédictibles sur la combinatoire lexicale de la lexie vedette, ce qu'on appelle dans la littérature Sens-Texte la **combinatoire lexicale restreinte** de la

X = I	Y = II	Z = III	W = IV
1. N	1. N	1. à N 2. à Vinf 3. dans N 4. pour N 5. pour Vinf 6. pour que PROPSubj	1. avec N 2. de N 3. par N 4. en Vant

lexie.

Les deux types d'informations non prédictibles encodées dans le langage des fonctions lexicales sont les collocations et les dérivés sémantique d'une lexie :

Une collocation est une expression phraséologique semi-figée, c'est-à-dire une dépendance lexicale entre deux lexies où seule une des deux lexies (la base) conserve son sens propre tandis que l'autre (le collocatif) exprime un sens général appliqué à la base mais aucunement son sens propre³. Le sens du collocatif est représenté par le nom de la fonction lexicale. Par exemple, *noire* a un sens d'intensification (représenté par la FL **Magn**) auprès de *colère*.

Magn(*colère*) = *noire*.

Il arrive toutefois que le sens du collocatif ajouté à celui de la base puisse être exprimé par une seule lexie. C'est ce que l'on appelle dans la littérature Sens-Texte une valeur fusionnée (noté par le signe '//' devant la valeur retournée). Ainsi, dans l'exemple qui suit, *déluge* est une valeur fusionnée puisque le sens de cette lexie unifie le sens du mot-clé et celui de Magn : *déluge* a le même sens que *pluie diluvienne*.
ex : Magn(*Pluie*) = *diluvienne*, //déluge.

Les dérivés sémantiques d'une lexie correspondent à son **paradigme dérivationnel**, c'est-à-dire :

- l'ensemble des lexies qui sont liées à la lexie vedette par une relation sémantique forte (la synonymie, l'antonymie, la conversion).
- les lexies représentant le nom typique des actants de la lexie vedette. Par exemple, le nom typique de X dans 'X aide Y à Z-er par W' peut être dans certains cas *bienfaiteur*.
- Les dérivés syntaxiques de la lexie vedette. Le nom correspondant au verbe *aider* est *aide*

Les dérivés sémantiques sont pris en charge par les fonctions lexicales paradigmatiques.

³Signalons toutefois que la non compositionnalité d'une collocation n'est pas toujours aussi évidente que dans des cas comme *colère noire*, *peur bleue*, etc

4. La zone phraséologique recense tous les phrasèmes dans lesquels apparaît la lexie vedette. Par exemple la définition du nom *aide* contiendra des pointeurs vers les phrasèmes *aide maternelle*, *aide ménagère*, etc.

1.2 Autres approches lexicographiques

Nous avons montré que le mode de représentation du sens lexical choisi par une théorie sémantique dépend du type de calcul que la théorie compte effectuer : la Théorie Sens-Texte a besoin, pour la génération de paraphrases, de décrire le sens lexical selon les trois aspects traités dans un article du DEC :

La décomposition du sens sert au niveau de représentation sémantique (RSem) ainsi qu’au passage entre ce niveau et le niveau de représentation syntaxique profonde (RsyntP).

Les informations concernant la combinatoire syntaxique d’une unité lexicale servent au passage entre la RSem et la RSyntP ainsi qu’au passage entre la RsyntP et la RsyntS.

Les informations sur la combinatoire lexicale sont, quant à elles, utilisées dans les règles de paraphrasage au niveau de la RSyntP.

Nous nous proposons ici de présenter d’autres approches en lexicologie formelle qui offrent des alternatives au traitement de certains aspects du sens lexical proposé par la LEC.

Nous indiquerons d’abord, pour chaque théorie, les objectifs linguistiques envisagés qui permettront d’expliquer le type de représentation proposé par la théorie. Nous décrirons ensuite la (ou les) structure(s) qui forment cette représentation.

1.2.1 WordNet

La base de données lexicales WordNet (Miller[90]) a été élaborée sur la base de principes psychologiques concernant la mémoire lexicale humaine. Les auteurs du projet proposent une organisation du lexique qui rende compte de la manière dont on accède ‘naturellement’ à l’information lexicale.

Dans WordNet, chaque catégorie syntaxique majeure (nom, verbe, adjectif, adverbe) a son organisation lexicale propre⁴. Les noms sont, par exemple, organisés de façon hiérarchique.

Les noeuds de chacun des réseaux lexicaux ne sont pas des unités lexicales mais des **synset** c’est-à-dire des sens représentés par un ensemble d’unités

⁴Ce choix de représentation est justifié par des tests montrant qu’en règle générale les personnes à qui l’on demande de dire à quoi un mot leur fait penser répondent par un autre mot appartenant à la même catégorie syntaxique (Miller[90]).

lexicales synonymes.

La figure 1.9 représente une partie de la hiérarchie des noms de WordNet. Le synset central, représenté par les unités lexicales *buyer*, *purchaser*, *emptor*, *vendee*⁵, est en relation d’hyperonymie avec au moins deux synsets et a deux synset hyponymes.

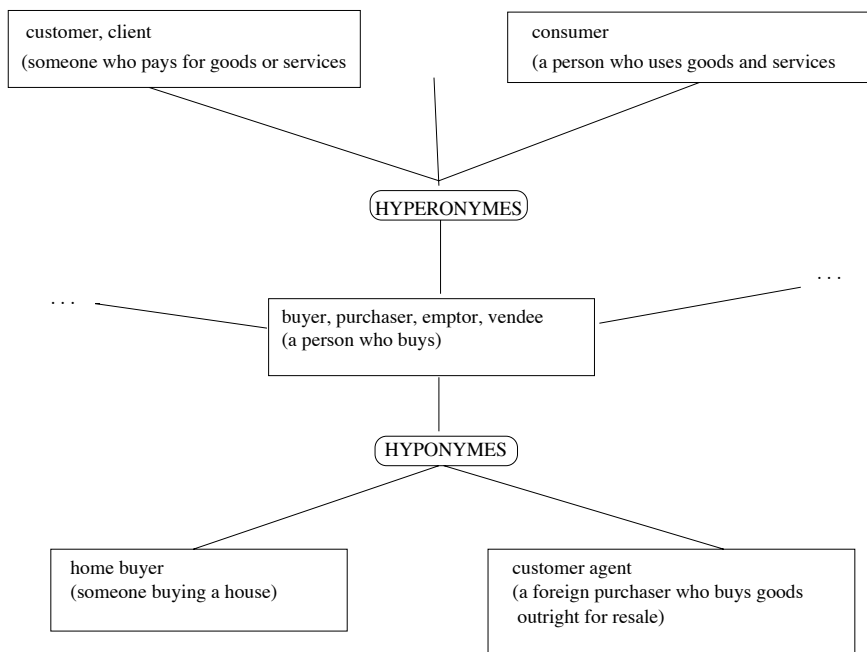


FIG. 1.9 – Une partie du réseau WordNet

La figure 1.9 ne représente pas, en revanche, les traits distinctifs associés aux synsets nominaux. Un synset nominal peut être défini par le (ou les) synset dont il hérite, c’est-à-dire son hyperonyme, et par un ensemble de traits distinctifs qui le distingue des autres synsets qui ont le même hyperonyme. (Miller[90]) donne l’exemple des traits distinctifs associés au synset contenant l’unité lexicale *canary* :

Attributes : small, yellow

Parts : beak, wings

Functions : sing, fly

Un synset hérite des traits distinctifs de son hyperonyme. Dans l’exemple donné ci-dessus, le synset a hérité du synset hyperonyme contenant l’unité

⁵la définition représentée entre parenthèses n’appartient pas, au niveau théorique, à la structure d’un Synset. Elle n’est là que pour ‘éclairer’ l’utilisateur de WordNet

lexicale *bird* une des fonctions, celle de voler, et ses parties, bec et ailes.

Selon la conception de WordNet, le sens d'une unité lexicale émerge de la place qu'elle occupe au sein du lexique. La LEC considère plutôt, pour sa part, que c'est le sens d'une unité lexicale qui détermine la place qu'elle occupe au sein du lexique. Nous reviendrons sur ce point dans la prochaine section consacrée au lien qui unit le sens d'une unité avec les relations sémantiques qu'elle entretient avec les autres unités du lexique.

1.2.2 FrameNet

La base de données lexicales FrameNet ⁶ a été développée pour l'annotation sémantique de corpus.

Fondée sur la sémantique des cadres de Fillmore, la base de données repose sur une hiérarchie de frames (cadres conceptuels) sur laquelle pointent les unités lexicales. Selon Fillmore, on comprend un mot en le reliant à la situation conceptuelle qui lui est sous-jacente. En ce sens, les lexicographes de FrameNet adoptent une démarche onomasiologique puisque leur attention se porte d'abord sur le concept (la description de la situation) pour ensuite établir la liste des unités lexicales qui expriment un des aspects de cette situation.

Un **frame** est composé d'une définition de la situation dans laquelle apparaissent les différents Frame Elements (FE), les participants à la situation. Chaque Frame Element est ensuite caractérisé par une phrase qui réexprime la situation du point de vue du Frame Element décrit. Un frame contient enfin la liste des unités lexicales qui lui sont reliées.

La figure 1.10 reprend la représentation graphique du frame *commerce*.

Une **entrée lexicale** contient une référence à son frame, ainsi qu'aux frames dont celui-ci hérite. Un tableau indique les différentes réalisations syntaxiques des FE pertinents du point de vue de l'unité lexicale, en indiquant leur nature et leur fonction. Un autre tableau indique la co-occurrence de ces FE.

La figure 1.11 reprend la représentation graphique de l'entrée lexicale *buyer.n*, une des unités lexicales reliées au frame *commerce*.

La première ligne du tableau de réalisation syntaxique indique que le FE **Buyer** a été annoté dans 64 phrases. Dans 4 phrases, il est réalisé par un groupe nominal (NP) en position sujet (**Ext**).

La dernière ligne du tableau de valence indique que, dans une phrase du corpus, le FE **Goods** est réalisé comme un groupe prépositionnel dépendant

⁶FrameNet : www.icsi.berkeley.edu/~framenet

COMMERCE	
Definition: These are words describing basic transaction involving a Buyer and a Seller exchanging Money and Goods.	
Frame Elements:	
Buyer	The Buyer has the Money and Wants the Goods
Goods	Goods is anything including labor or time, for exemple, which is exchanged for Money in a transaction.
Money	Money is given in exchange for Goods in e transaction
Seller	The Seller has the Goods and wants the Money
Lexical Units: buy.v, buyer.n, charge.v, cost.n, cost.v, goods.n, lease.v, pay.v, payment.n, price.n, purchase.n, purchase.v, purchaser.n, rent.v, retail.v, retailer.n, sell.v, seller.n, spend.v, vend.v, vendor.n	

FIG. 1.10 – Le frame **commerce**

du FE **Buyer**. Le FE **Seller** est le sujet syntaxique de la phrase.

FrameNet offre donc une représentation des trois aspects d'une unité lexicale traités par la LEC :

La description du **sens** d'une unité lexicale consiste à indiquer la ou les structures conceptuelles auxquelles elle est liée. La structure conceptuelle est décomposée en Frame Elements.

La description de la **combinatoire syntaxique** donne la réalisation des Frame Elements dans les phrases dans laquelle l'unité lexicale apparaît.

La description de la **combinatoire lexicale** indique la manière dont les différents FE peuvent apparaître dans une même phrase.

Comme on peut le remarquer, la différence entre les deux types de représentation repose, dans le cas de FramNet, sur l'utilisation des primitives conceptuelles (les Frame Elements) dans les trois zones de la représentation lexicale. L'objectif de FrameNet étant d'annoter sémantiquement des phrases, c'est-à-dire d'attribuer des rôles sémantiques aux dépendants syntaxiques du prédicat principal de la phrase, il lui faut pouvoir mettre directement en rapport les éléments sémantiques (les Frame Elements) avec les éléments syntaxiques. La LEC, quant à elle, n'attribue pas de rôle aux actants sémantiques mais seulement des numéros qui puissent les identifier au moment où ils se réalisent au niveau syntaxique.

<p>buyer.n</p> <p>Frame: Commerce</p> <p>Definition:</p> <p>COD: a person who buys</p> <p>Frame Elements and Their Syntactic Realizations:</p>			<p>Valence Patterns:</p>			
FE	Number Annotated	Realization(s)	Number annotated	Patterns		
Buyer	61	NP,Ext 4 --,-- 56 NP,Comp 1	24 total	Buyer	Goods	
Goods	26	N.Mod 7 Poss,Gen 1 PP.Comp 15	13 7	--,-- --,--	PP,Comp N,Mod	
Money	1	PP.Comp 1	35 total	Buyer		
Seller	2	NP.Ext 2	32 3	--,-- NP,Ext		
			1 total	Buyer	Goods	Seller
			1	--,--	PP,Comp	NP,Ext

FIG. 1.11 – L'entrée lexicale **buyer.n**

1.2.3 Le Lexique Génératif

Le Lexique Génératif (Pustejovsky[95]) a été développé en opposition à la description lexicale consistant à énumérer les différents sens d'un mot polysémique. Selon J.Pustejovsky, ce mode de description est inapte à représenter la vraie nature de la polysémie et ne peut pas prendre en compte la créativité du sens lexical. Il propose donc de ne décrire qu'une partie des sens lexicaux et de laisser à des mécanismes de composition le soin de générer d'autres sens, dérivés des premiers.

Une structure sémantique du Lexique Génératif est une structure de traits typée composée de trois sous-structures :

1. la structure argumentale recense les différents participants à la situation dénotée par l'objet décrit. Les arguments sont de différents types selon qu'ils sont ou non réalisés au plan syntaxique.
2. la structure événementielle se charge d'offrir une représentation des sous-événements de l'objet décrit, si celui-ci est un événement. Un événement peut être de trois types : un état, un procès ou une action.
3. la structure des qualia indique comment les arguments et les événements participent aux différentes propriétés de l'objet, les propriétés étant classées suivant quatre rôles :
 - (a) le rôle **formel**, i.e grosso modo le type sémantique de l'objet dénoté.
 - (b) le rôle **constitutif**, i.e ce dont est formé l'objet.
 - (c) le rôle **télique**, i.e ce pour quoi l'objet est créé, sa fonction propre.
 - (d) le rôle **agentif**, i.e ce qui crée l'objet.

La figure 1.12 représente l'entrée lexicale **livre**.

LG établit une distinction entre le ou les sens inhérents à un mot et les sens que celui-ci peut prendre en contexte. Les premiers sont représentés par le type (simple ou complexe) associé à la représentation de base (1), les autres sont pris en charge par des mécanismes de composition (2).

- 1) Le type d'un objet, sa dénotation, est le résultat de l'unification des types de la structure des qualia.

Dans certain cas, le mot a un type pointé c'est-à-dire qu'il dénote deux types différents. Le mot est alors dit endocentrique si il dénote les deux types en même temps (l'unité lexicale **livre** est de type **information.objet.physique**), exocentrique quand il ne dénote pas nécessairement les deux types en même temps (l'unité lexicale **repas** est de type **procès.nourriture**). La structure des qualia indique le lien existant entre les types qui forment le type pointé.

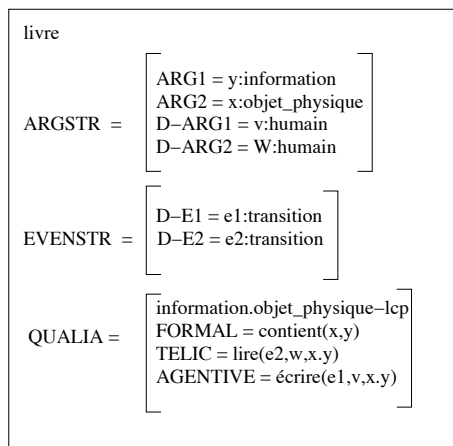


FIG. 1.12 – Structure lexicale de ‘livre’

2) La structure des qualia détermine donc les types potentiels d’un objet. Les mécanismes de composition vont permettre de rendre compte des types effectifs de l’objet en contexte.

La coercition de type permet par exemple à un prédicat qui attend normalement un argument de type x de recevoir un argument de type y si cet argument possède le type x dans sa structure de qualia.

Par exemple, le prédicat COMMENCER, qui attend normalement un argument de type ‘événement’, pourra recevoir **livre** en argument (*commencer un livre*) car **livre** a deux prédicats de type ‘événement’ (l’un comme valeur du rôle **telic**, l’autre comme valeur du rôle **agentive**) dans sa structure des qualia.

Le Lexique Génératif propose donc un traitement de la polysémie en contexte. La LEC, quant à elle, définit a priori les différents sens d’un polysème : une unité polysémique correspond à un vocable du dictionnaire et ses différentes acceptions correspondent à des lexies. Les lexies d’un vocable sont liées entre elle mais sont considérées toutefois comme des sens distincts, autonomes.

L’encodage des collocations sous forme de FL est toutefois à rapprocher du point de vue compositionnel adopté par le Lexique Génératif dans le traitement de la polysémie (Kahane[01]). Le formalisme des FL semble en effet être un outil approprié pour mettre en évidence les régularités du lexique comme propose de le faire le Lexique Génératif avec les mécanismes de composition.

1.3 Décomposition du sens et relations lexicales

Il est d'usage de distinguer, parmi les théories sémantiques lexicales, les théories dites **relationnelles** des théories dites **décompositionnelles** (cf. Bouillon[98]).

Les premières envisagent l'unité lexicale comme une unité atomique, non décomposable, et dont le sens est défini en fonction des relations lexicales qu'elle entretient avec les autres unités du lexique.

Les secondes considèrent que chaque unité lexicale peut être décomposée au moyen d'un ensemble prédéfini de symboles.

Bien que l'on aurait tendance à classer WordNet dans la première catégorie et la LEC dans la seconde, nous avons vu que les deux théories mêlaient ces deux types de description. C'est justement le point qui va nous intéresser au cours de cette section.

Nous souhaitons en effet évoquer à présent la question du lien qui unit la décomposition du sens d'un mot et les relations lexicales que ce mot entretient avec les autres mots du lexique. Sans prétendre bien sûr répondre à cette question, nous voudrions insister sur l'importance d'avoir une représentation lexicale qui réunissent ces deux aspects de la description. Nous pensons en effet qu'il est important qu'une représentation offre les moyens de **vérifier** les données qui y sont encodées :

Nous verrons d'une part que, concernant du moins la LEC, la sélection de certains éléments de la décomposition doit pouvoir être justifiée par certains éléments de représentation des relations lexicales.

Nous verrons d'autre part qu'une relation lexicale ne peut être définie qu'en fonction des éléments de décomposition et que par conséquent l'encodage d'une relation lexicale doit pouvoir être justifiée au moyen de la décomposition des unités mises en relation.

1.3.1 Décomposition du sens lexical

Les quatre théories sémantiques que nous venons de présenter ont en commun de proposer une représentation du sens lexical qui prend, en partie, la forme d'une décomposition, c'est-à-dire qu'elles identifient des composantes de sens contenues dans le sens de l'unité lexicale (UL) décrite.

Avant d'identifier clairement ce que l'on entend par **composante de sens**, il faut s'interroger sur la nature des éléments qui servent à la décomposition.

Nature des éléments d'une composante de sens

Les théories présentées ont recours à deux types d'éléments :

1. Elles peuvent utiliser d'autres unités lexicales, autrement dit des éléments qui sont du même type que l'unité décrite.
2. Elles peuvent définir un ensemble limité de symboles qui serviront à décrire l'ensemble des unités du lexique.

Les symboles utilisés pour le second type de décomposition ressemblent souvent, pour des raisons évidentes de représentativité, aux unités lexicales du lexique et il est parfois difficile, surtout dans les cas où elles mêlent dans la représentation les deux types d'éléments, de savoir à quoi l'on a affaire lorsque la théorie ne l'explique pas.

WordNet et le Lexique Génératif ont l'un et l'autre recours à l'utilisation de **traits sémantiques** pour représenter les différentes composantes de sens d'une UL (ou d'un concept).

Un trait sémantique est une association entre un **attribut** et un ensemble de **valeur(s)** représentée généralement ainsi :

Attribut : valeur(s)

Si les attributs relèvent clairement du métalangage, la nature des valeurs associées à cet ensemble d'attributs est moins évidente.

Dans WordNet, chaque Synset est décrit par un ensemble de traits sémantiques distinctifs. Nous reprenons ici les traits distinctifs du synset contenant l'unité lexicale *canary* :

Attributes : small, yellow

Parts : beak, wings

Functions : sing, fly

Si l'on peut penser que les valeurs sont des unités lexicales, la représentation ne nous permet pas d'en être sûr (l'indication du Synset d'appartenance lèverait par exemple l'ambiguïté).

Il en est de même pour les traits utilisés dans la structure des qualia d'une structure sémantique du lexique génératif : la structure argumentale et la structure événementielle prennent comme valeur des variables typées mais on ne sait pas, concernant les valeurs associées aux rôles (Formel, Constitutif, Télique et Agentif) si elles correspondent directement ou non à des unités du lexique.

La LEC utilise pour décomposer le sens d'une unité lexicale des unités lexicales plus simples. Elle définit le métalangage dans la *règle de standardisation*⁷, mais ces unités définitionnelles n'ont, à notre connaissance, pas été répertoriées et n'apparaissent en outre pas encore clairement au sein des définitions Sens-Texte.

⁷« La définition doit être faite de façon à éviter, dans le définissant, a) les termes AMBIGUS et b) les termes SYNONYMES » (Mel'čuk[95])

Si nous définissons une **composante de sens** comme **une association entre ce qui pourrait s'apparenter à un attribut (qui relève du métalangage) et ce qui pourrait s'apparenter un ensemble de valeurs (qui appartiennent à priori à la langue décrite)**, les composantes ne sont, pour la raison que l'on vient d'invoquer, pas encore clairement identifiable au sein d'une définition linéaire d'un article du DEC.

Le travail de mise en évidence des composantes de sens au sein d'une définition Sens-Texte a été engagé avec le développement de la BDéf. Nous essayerons de poursuivre, au cours de ce travail, l'effort de formalisation du langage définitionnel entrepris par les auteurs de la BDéf. La première étape sera de définir précisément les unités lexicales contraintes qui jouent le même rôle qu'un attribut dans un trait sémantique. La question se posera ensuite de savoir sur quoi repose la sélection des composantes de sens qui doivent entrer dans une définition.

Une partie des éléments qui vont probablement permettre d'apporter un début de réponse à cette question se trouve dans l'encodage des relations lexicales qu'une unité entretient avec les autres unités du lexique.

1.3.2 Relations lexicales et composantes de sens

Des quatre théories lexicales présentées dans la section précédente, seuls WordNet et la LEC proposent une représentation des relations lexicales :

La position tenue par WordNet consiste, on l'a vu, à dire que l'on connaît le sens d'une unité lexicale lorsqu'on connaît la place qu'elle occupe au sein du lexique ce qui explique que WordNet offre une représentation des relations lexicales.

Concernant la LEC, le but de la TST étant de générer des paraphrases à partir d'un sens, il lui faut représenter les relations lexicales qui vont lui permettre d'écrire des règles d'équivalence sémantique.

représentation des relations lexicales dans WordNet :

Il y a dans WordNet des relations à l'intérieur d'une même hiérarchie et des relations entre les hiérarchies (Miller[90]).

Les premières correspondent à des relations paradigmatiques qui peuvent être définies comme des relations sémantiques entre synsets d'une même organisation lexicale, donc entre sens d'unités lexicales qui sont de même nature syntaxique. Ces relations paradigmatiques sont, concernant la hiérarchie des noms, les relations d'hyponymie/hyponymie, la relation d'antonymie et la relation de méronymie.

Les secondes représentent les relations syntagmatiques qui sont définies comme des relations inter-hiérarchiques. Un nom par exemple est relié à la structure des verbes par le trait **Functions** et à la structure des adjectifs

par le trait **Attributes**.

Notons tout de suite ici que ce sont certains des traits sémantiques d'un Synset (en l'occurrence **Functions** et **Attributes** pour un synset nominal) qui sont utilisés pour représenter un certain type de relation lexicale, en l'occurrence les relations syntagmatiques. Le troisième trait associé à un synset nominal (**parts**) renvoie, lui, à la relation paradigmatique de méronymie.

représentation des relations lexicales selon la LEC :

Nous avons vu que la LEC représente les différents types de relations lexicales en terme de fonctions lexicales (FL). La LEC recense une soixantaine de fonctions lexicales (FL) standard⁸qu'elle subdivise également en FL paradigmatique et FL syntagmatique (voir Mel'čuk[95] p. 129-152).

La frontière entre ces deux types de relation est toutefois moins nette que dans le cas de WordNet :

- Les dérivés syntaxiques sont considérés comme des lexies appartenant au même paradigme lexicale que la lexie mot-clé⁹(ex. $S_0(\textit{présenter}) = \textit{présentation}$).
- À l'inverse, des FL syntagmatiques peuvent renvoyer des valeurs qui appartiennent à la même partie du discours que le mot clé. C'est la cas, par exemple, de la FL d'intensification **Magn**. L'intensification s'exprime en général par un adjectif pour un nom ($\textit{Magn}(\textit{pluie}) = \textit{diluvienne}$) et par un adverbe pour un verbe ($\textit{Magn}(\textit{crier}) = \textit{comme un sourd}$) mais il arrive parfois qu'il existe dans le lexique une lexie exprimant l'intensification du mot-clé ($\textit{Magn}(\textit{pluie}) = \textit{déluge}$; $\textit{Magn}(\textit{crier}) = \textit{hurler}$).

La LEC établit donc deux grands types de FL :

1. Les FL qui comparent les éléments de sens du mot-clé avec les éléments de sens de la valeur. Il doit y avoir un nombre important d'éléments de sens en commun qui justifie leur appartenance à un même paradigme. Les éléments qui diffèrent déterminent normalement le type de la FL. Si par exemple le sens de $X \textit{ est ABSENT de } Y$ ne diffère du sens de $X \textit{ est PRESENT à } Y$ que par une négation, on pourra les encoder comme des antonymes.
2. Les FL qui renvoient à un élément du sens du mot-clé. Le sens de la FL se réalise en langue soit sous la forme d'un collocatif (qui a alors le

⁸La LEC définit également des FL non standard. voir (Mel'čuk[95])

⁹Quand deux unités lexicales sont reliées par une FL paradigmatique, elles ont peu de chance d'être en relation de cooccurrence syntaxique, même si elles appartiennent à des parties du discours différentes qui ont le droit de se combiner syntaxiquement. Il y a toutefois parfois des exceptions dues au caractère parfois redondant de la langue : $\textit{monter en haut}$, ($\textit{Adv}_0(\textit{MONTER}) = \textit{EN HAUT}$), $\textit{rentrer à l'intérieur}$ ($\textit{Adv}_0(\textit{RENTREER}) = \textit{A L'INTERIEUR}$)

sens de la FL) soit sous la forme d'une unité autonome qui additionne le sens du mot clé et le sens de la FL. La FL **Magn**, par exemple, ne peut s'appliquer qu'à une lexie qui contient dans son sens une composante qui peut être intensifiée.

On voit donc que, dans le cas de WordNet aussi bien que dans celui de la LEC, la définition des relations lexicales repose sur les composantes de sens de l'unités.

Dans un soucis de justification/vérification des données encodées dans une description lexicale, nous pensons qu'il est utile d' avoir, d'une part, une représentation des relations lexicales, d'autre part d'avoir une représentation des composantes de sens d'une unité (décomposition). Il faut enfin que ces représentations soient suffisamment formalisées pour pouvoir être mises en relation.

La suite de ce travail sera consacrée essentiellement à la question de la formalisation de la représentation des composantes de sens d'une unité lexicale, qui prend, dans le cadre de la LEC, la forme d'une définition.

Chapitre 2

Le formalisme de la BDef

Nous avons présenté dans le chapitre précédent les trois aspects de la description du sens lexical traités par un dictionnaire Sens-Texte, c'est-à-dire la décomposition du sens d'une lexie, sa combinatoire syntaxique et sa combinatoire lexicale restreinte. Nous allons nous intéresser maintenant au premier de ces aspects, pivot de la description lexicologique Sens-Texte : la décomposition du sens d'une lexie prenant la forme d'une définition lexicographique.

Le langage de la définition est, comme tout langage de description linguistique, un métalangage, défini comme un langage contraint utilisé pour décrire certains aspects d'une langue naturelle. Le métalangage peut-être plus ou moins contraint et peu s'apparenter plus ou moins à la langue qu'il décrit. En l'occurrence le métalangage lexicographique traditionnel semble peu artificiel puisqu'on distingue mal, dans la définition d'un dictionnaire de la langue courant, ce qui relève du métalangage de ce qui n'en relève pas. Cela présente un avantage et un inconvénient majeurs : d'un côté la description semble naturelle et donc facilement interprétable par n'importe quel humain, de l'autre, le langage dans lequel la description est faite n'étant pas contraint, elle en rend difficile tout traitement automatique.

L'originalité de la BDef, base de données de définitions Sens-Texte développée à l'OLST par Joël Altman et Alain Polguère, est de parvenir à trouver un compromis entre ces deux types de description. Les auteurs ont entrepris de développer un formalisme de la définition qui d'une part soit aisément interprétable par un humain et qui d'autre part permette un traitement automatique de l'information lexicale.

Notre tâche dans ce travail porte d'avantage sur ce second aspect de la description et consiste à s'assurer que le métalangage de la BDef offre bien une représentation lexicale calculable. Pour cela, comme nous l'avons évoqué à la fin du chapitre précédent, le langage de la définition devra s'apparenter à un

langage contrôlé, c'est-à-dire un langage qui utilise le même type d'éléments qu'une langue non contrôlée mais dont les règles lexicales et sémantiques sont beaucoup plus limitées que celle d'une langue naturelle (voir Nasr[96] pour la définition d'une langue contrôlée dans le cadre de la Théorie Sens-Texte).

La première section de ce chapitre sera consacrée à une présentation de la base de données lexicales BDef. Dans la seconde section, nous expliquerons pourquoi nous entreprenons de décrire ce métalangage définitionnel à l'aide d'une grammaire formelle. La troisième section contiendra une ébauche de cette grammaire formelle.

2.1 Présentation de la BDéf

Nous reprenons dans cette section un nombre important de points évoqués dans l'article qui offre une première présentation de la base de données BDéf (Altman&Polguère[03]), en insistant sur les trois niveaux de structuration d'une définition BDéf : Une **définition** est constituée de **blocs définitionnels**, eux-même constitués de **propositions**.

Nous présenterons dans un premier temps les descriptions lexicales dont la BDef s'inspire pour développer son formalisme (3.1.1). Puis nous nous intéresserons à la forme proprement dite d'une définition de type BDef en décrivant une fiche de la base de données lexicales (3.1.2) et en présentant les deux notions centrales qui servent son formalisme (3.1.3).

2.1.1 Les précurseurs du formalisme de la BDéf : I.Mel'čuk et A.Wierzbicka

Le mode de description de la BDef suit la plupart des principes de la Lexicologie Explicative et Combinatoire (LEC) présentés dans Mel'čuk[95]. Il s'inspire également des travaux d'Anna Wierzbicka (Wierzbicka[92]), notamment en ce qui concerne son effort de mise en évidence de la structure des définitions.

Décomposition du sens lexical en unités plus simples

La lexicologie Sens-Texte partage avec les travaux d'A.Wierzbicka l'idée de définir le sens d'une unité lexicale d'une langue grâce à d'autres unités lexicales de cette même langue.

La LEC préconise que les lexies utilisées pour définir une lexie soient *immédiatement plus simples* : cela signifie d'une part qu'une lexie B utilisée dans la définition d'une lexie A ne devra pas contenir à son tour dans sa définition la lexie A (*plus simple*) et d'autre part que si deux lexies B et C utilisées

dans la définition de la lexie A correspondant au sens d'une lexie D elle aussi plus simple que la lexie A, alors les lexies B et C devront être remplacées par la lexie D dans la définition de A (*immédiatement plus simple*). Ces deux principes correspondent respectivement au *principe de décomposition* et au *principe du bloc maximal* énoncés par la LEC (cf Mel'čuk[95]).

A. Wierzbicka utilise elle aussi les unités de la langue décrite dans les définitions qu'elle propose. En revanche elle ne s'impose pas le principe du bloc maximal de la LEC. Au contraire, A. Wierzbicka a recours dans ses définitions à un ensemble de primitives sémantiques universelles et indécomposables qui devraient constituer le lexique d'une langue sémantique universelle.

La LEC, moins préoccupée par la découverte du langage universel, ne postule pas un ensemble de primitives mais l'envisage au terme de la description du lexique : le principe de décomposition maximale mène en effet de façon progressive aux lexies qui ne peuvent plus être décrites en terme de lexies plus simples.

Vers les notions de proposition et de bloc définitionnel

La LEC identifie différentes composantes de sens au sein de la définition (cf Mel'čuk[95], pp 103-106). Nous reprenons uniquement, parmi ces composantes, celles dont nous reparlerons dans la suite de l'exposé. La définition de 'X aide Y à Z-er par W' nous permettra d'illustrer ces différentes composantes :

X aide Y à Z-er par W : L'individu Y étant en train d'effectuer une action Z souhaitable pour Y, l'individu X emploie ses ressources W de sorte que W facilite ou rend possible Z pour Y.

- la **composante centrale** est ce qui correspond au genre prochain ou à l'hyperonyme de la lexie vedette. Dire que 'X aide Y à Z-er par W' c'est dire que 'X emploie ses ressources W', les différences spécifiques se chargeant de spécifier cette action de X.
- les **différences spécifiques** permettent de spécifier le sens de la composante centrale. La définition de 'X aide Y à Z-er par W' contient une différence spécifique : X emploie ses ressources W (composante centrale) dans le but que Y réussisse ce qu'il a entrepris.
- le **typage des actants sémantiques** consiste à caractériser les actants de la lexie au moyen d'un type sémantique. Dans la définition, X est de type **individu**, Y n'est pas typé et W est de type **ressources**.
- la **composante liante**, appelée également **pont sémantique**, est une composante qui permet de mettre en relation une lexie avec une ou plusieurs autres lexies d'un même vocable. La notion de pont sémantique supposant de considérer plusieurs articles d'un même vocable, nous ne

pouvons pas localiser une composante liante dans cette définition.

La LEC identifie donc différents types de composantes de sens au sein d'une définition. Le problème est que ces différentes composantes de sens n'apparaissent pas explicitement dans les définitions linéaires du type de celle qui est présentée ci-dessus.

Les définitions proposées par A. Wierzbicka exposent, quant à elles, une structuration explicite du sens lexical : les définitions sont découpées en propositions clairement identifiables comme l'illustre la définition de 'X glare at Y' extraite de (Wierzbicka[92]) où l'on voit par exemple que ce qui correspond à la composante centrale (X looked at Y) est isolée au sein d'une proposition.

```
X glare at Y. =  
X looked at Y  
like someone who wants to say:  
    I feel something bad towards you  
    I want you to know this
```

La BDéf va reprendre de la LEC la notion de composante (notion de **bloc définitionnel** dans la BDéf) et de A. Wierzbicka la notion de proposition (notion de **proposition élémentaire** dans la BDéf). Avant de définir ces deux éléments de la structure définitionnelle de la BDéf, il nous faut présenter la structure adoptée dans une fiche de la base de données.

2.1.2 Une fiche de la BDéf

Nous présentons ici succinctement les différents champs d'une fiche de la BDéf. Les notions d'**étiquette sémantique**, de **composante centrale** et de **composantes spécifiques** feront l'objet d'un développement dans la suite de ce chapitre.

Les différents champs pourront être illustrés par la figure 3.1 reproduisant la représentation graphique de la fiche NEZ#I.1a.

Nous signalons que d'autres définitions BDéf, celles des lexies BRAS#I.1a et BRAS#I.1b, sont fournies en annexe.

Lexie : NEZ I.1a

Carac. Gram :

Forme_prop : ~ de X

Exemple : Elle avait les yeux rieurs et un nez mutin couvert de taches de rousseur

DEFINITION

1: partie de visage#1.a de X

*/*Localisation spatiale*/*

2: *1 mediane

*/*Forme*/*

3: *1 saillant

4: *1 allongé

*/*Fonction physiologique*/*

5: *1 etre organe de odorat

6: *1 etre_un_des_organes de respiration

*/*Fonction sémiotique [optionnel]*/*

7: apparence de *1 exprimer Y

Typage

X : individu

Y : propriété psychologique

Rel_SEm

Y[X]

FIG. 2.1 – fiche BDef de NezI.1a

Lexie : Ce champ est constitué du nom du vocable auquel appartient la lexie, suivi d'un numéro indiquant son acception si le vocable est polysémique.

Carac. Gram. : Ce champ, facultatif, regroupe des informations d'ordre divers sur la lexie : registre de langue, genre de la lexie si celle-ci est un nom, etc.

EtiquSém : La première étape de la description d'une lexie consiste à lui attribuer un type, qui prend la forme d'une étiquette sémantique.

Forme_prop : La forme propositionnelle correspond au définiendum de la définition, c'est-à-dire ce que l'on va définir. Dans le cas d'une lexie prédicative, la forme propositionnelle contient la lexie - remplacée par un tilde (\sim) - accompagnée de ses actants sémantiques.

Exemples : Les exemples sont choisis dans un corpus et modifiés si besoin est pour illustrer de façon simple et claire le comportement de la lexie dans un énoncé de la langue.

Définition : La définition est subdivisée en quatre champs :

la composante centrale de la définition occupe le premier cadre. Etroitement liée à l'étiquette sémantique, elle constitue la paraphrase minimale de la lexie, c'est-à-dire une formulation du sens général de la lexie.

les composantes spécifiques occupent le second cadre du champ **Définition**. La fonction précise de ces composantes, appelées blocs définitionnels élémentaires, n'a pas été encore clairement déterminée.

Typage : Le champ Typage isole les différents actants sémantiques de la définition et leur attribue un type quand cela est possible.

Rel_Sem : Ce champ indique les éventuelles relations sémantiques entretenues entre les actants. Dans la fiche donnée en illustration, $Y[X]$ signifie que X est le premier actant sémantique de Y et peut se lire grâce au typage comme *la propriété psychologique Y de l'individu X* ou encore *l'individu X a la propriété psychologique Y* .

Nous allons nous intéresser aux champs qui représentent le corps de la définition : les champs qui encodent la composante centrale et les différences spécifiques de la définition. Les informations contenues dans ces deux champs

sont structurées sous forme de propositions élémentaires regroupées en blocs définitionnels, notions que nous allons définir dès à présent.

2.1.3 Les deux notions piliers du formalisme de la BDef : la proposition élémentaire et le bloc définitionnel

La proposition élémentaire

Qu'est-ce qu'une proposition élémentaire :

Une proposition élémentaire peut être définie comme un énoncé ayant une structure communicative simple, c'est-à-dire un thème (ce dont l'énoncé parle) et un rhème (ce que l'énoncé apporte comme nouvelle information sur le thème) qui ne soient pas à leur tour des énoncés ayant une structure communicative.

Soient les deux énoncés :

- 1) *Les jambes sont l'organe du déplacement de X en position verticale*
- 2) *Les jambes sont l'organe du déplacement de X qui se fait en position verticale*

L'énoncé 2, contrairement à l'énoncé 1, ne constitue pas une proposition élémentaire puisqu'il dit d'une part que les jambes servent à ce que X se déplace et d'autre part que ce déplacement s'effectue en position verticale. Il devra donc, si le lexicologue tient à insister sur ces deux informations, le diviser en deux propositions élémentaires :

- 2a) *Les jambes sont l'organe du déplacement de X*
- 2b) *Le déplacement de X se fait en position verticale*

L'énoncé 1 contient bien les mêmes informations mais l'information qui caractérise le mode de déplacement est 'mise en sourdine' comme une information qui ne serait pas directement pertinente du point de vue du sens de *jambes*.

Forme d'une proposition élémentaire :

Une proposition élémentaire est constituée d'un prédicat principal et des arguments du prédicat principal. Elle est identifiée par un numéro.

num : arg1 prédicat_principal arg2

Remarque 1 : Le numéro associé à la proposition élémentaire joue un rôle d'identification mais peut dans certains cas avoir une signification. C'est le

cas lorsqu'il est composé et qu'il représente alors une disjonction inclusive entre deux propositions.

5.1: *1 laisser_passer n1={rayon}

5.2: *1 percevoir *n1

doit se lire comme :

5: *1 laisser_passer n1={rayon} ET/OU *1 percevoir *n1

Remarque 2 : Une proposition élémentaire est débarrassée de toutes les flexions morphologiques dans le but d'en faciliter l'analyse automatique. Les flexions morphologiques qui ont un sens pertinent pour la définition sont rétablis sous une autre forme. Une lexie au pluriel, par exemple, sera représentées entre accolades.

5.1: *1 laisser_passer n1={rayon}

Nature du prédicat principal de la proposition :

– une lexie

*1 percevoir *n1

– une expression sémantiquement compositionnelle, représentée par des traits de soulignement intérieurs.

*1 laisser_passer *n1

Nature d'un argument du prédicat principal de la proposition :

– une proposition. Dans ce cas, la proposition sera représentée par son numéro précédé du signe '*'.
*1 permettre#II à X de *9

– une lexie. On crée une variable pour les lexies qui sont utilisées dans une autre proposition de la définition (la variable est alors précédée du signe '*').

*1 laisser_passer n1={rayon}

*1 percevoir *n1

– un groupe nominal (si celui-ci n'est pas marqué communicativement)

*1 être organe de déplacement de X en position verticale

– un actant de la forme propositionnelle

FRONT#I.a (Y) de X

apparence de *1 exprime Y

Le bloc définitionnel

Un **bloc définitionnel** se définit formellement par un ensemble de propositions qui, soit forment entre elles une proposition complexe,

/*Fonction physiologique*/

3: *1 permettre#II à X de *4

4: X absorber nourriture#I

soit sont indépendantes syntaxiquement les unes des autres

```
/*Forme*/  
3: *1 saillant  
4: *1 allongé
```

Nous avons signalé, au cours de la présentation d'une fiche BDéf, que la fonction définitionnelle d'un bloc n'est dans certains cas pas encore identifiée. Intuitivement, les lexicographes de la BDéf encodent dans un bloc un groupe de propositions qui n'apporte des informations que sur un seul aspect de la lexie. Cette information peut être, si besoin est, décomposée à nouveau et les propositions concernant cette décomposition formeront alors un **sous-bloc définitionnel**.

```
OEILI.1a  
/*Parties*/  
2: *1 constitué de n1=globe & n2=2{pli_de_peau}  
  
/*PartiesI.Localisation*/  
7: *n1 placé dans cavité  
  
/PartiesII.Forme*/  
10: *n2 horizontal
```

Le rôle informationnel d'un bloc ou d'un sous-bloc est, comme on le voit dans les exemples, explicité dans un **titre de bloc définitionnel** (représenté entre les signes /* et */).

La présence de ce titre est doublement justifiée :

1. L'information est parfois explicitée dans une proposition du bloc définitionnel, parfois non¹. Rien n'indique par exemple, dans les propositions 3 et 4 du bloc définitionnel /*Forme*/ que les prédicats **saillant** et **allongé** renvoient à la forme de ce qui est dénoté par la lexie vedette.

```
/*Forme*/  
3: *1 saillant  
4: *1 allongé
```

Par contre, le prédicat (être) **placé** de la proposition 7 indique que ce qui suit concerne la /*Localisation*/ d'un élément de ce qui est dénoté par la lexie.

¹Rappelons que les définitions de la BDef sont construites à partir des définitions du DEC papier.

`/*PartiesI.Localisation*/`

`7: *n1 placé dans cavité`

2. L'utilisation de titre va permettre l'élaboration d'un patron lexical associé à un groupe de lexies. Ce patron représentera l'essentiel des contraintes sémantiques propres au type de la lexie.

Les lexicographes de la BDef ont choisi pour le moment d'indiquer systématiquement cette information sous forme de titre. Nous souhaitons pour notre part encoder directement cette information au sein d'une proposition, au risque de réduire encore le côté 'naturel' d'une proposition BDef. Il faut pouvoir identifier clairement le prédicat qui se chargera de reprendre l'information encodée dans le titre d'un bloc définitionnel. Nous expliquerons ce choix dans la prochaine section où nous justifions l'entreprise de description du métalangage définitionnel.

2.2 Contrôle du formalisme de la BDéf

Le formalisme de la BDéf a été conçu en partie pour permettre un traitement automatique des définitions.

Nous cherchons à nous assurer de cette propriété du formalisme. Nous allons pour cela décrire – parfois agir sur – le **métalangage définitionnel** de la BDéf au moyen du formalisme d'une **grammaire formelle**. Cette grammaire va établir les règles sémantiques et syntaxiques du métalangage de la BDéf et refléter son degré de formalisation. L'idée est de voir si la **représentation abstraite** obtenue au terme de l'analyse du métalangage va effectivement être adaptée au calcul sémantique.

2.2.1 Contrôle des définitions

Nous souhaitons dans un premier temps dégager les règles syntaxiques du métalangage définitionnel, c'est-à-dire déterminer la façon dont :

- les éléments s'agencent dans une proposition,
- les propositions s'agencent dans un bloc définitionnel,
- et les blocs définitionnels s'agencent dans la définition.

Ces règles, qui doivent être appliquées quel que soit le type de la lexie décrite, vont permettre de vérifier que les définitions sont bien formées syntaxiquement.

Le contrôle sémantique des définitions va reposer principalement sur la définition de **types**.

Nous distinguons les types associés au métalangage des types des unités lexicales. Ces derniers prennent la forme, dans la BDéf, d'une **étiquette**

sémantique. (Polguère[03]) donne les principales caractéristiques d'une étiquette sémantique, que nous rappelons succinctement.

Rappel des principales caractéristiques des étiquettes sémantiques

La hiérarchie d'étiquettes sémantique se construit au fur et à mesure de la description du lexique. Cela signifie que la hiérarchie d'étiquettes n'est pas une ontologie, une organisation de concepts extralinguistiques, mais une représentation du lexique d'une langue donnée.

Il y a trois types d'étiquetages sémantiques :

étiquette sémantique simple : Une lexie de la langue (ex :|**attitude**|, |**sentiment**|) ou une expression sémantiquement compositionnelle c'est-à-dire une expression libre de lexies (ex :|**partie du corps**|).

étiquette sémantique composée : Elles s'obtiennent de deux façons, soit avec deux étiquettes simples dont l'une est l'argument de l'autre (ex : |**ensemble d'individus**|), soit par composition d'une étiquette sémantique avec un opérateur, qui lui n'appartient pas à la hiérarchie proprement dite.

disjonction d'étiquette : Certaines lexies dénotent plusieurs types.
(ex :|**attitude/sentiment**|)

L'étiquette sémantique ayant une fonction classificatrice, un prédicat trop particulier ne pourra faire office d'étiquette sémantique. Si l'étiquette choisie pour une lexie ne peut s'appliquer à au moins deux autres lexies, elle ne sera pas adoptée et le lexicographe devra trouver l'hyperonyme le plus proche.

Les types du métalangage

Il faut définir le lexique du métalangage. Chaque unité lexicale du métalangage doit, plus précisément, être définie par sa fonction définitionnelle (ce que pour l'instant on ne sait pas toujours déterminer) et par la caractérisation, au moyen du typage, de ses arguments.

Les unités lexicales contraintes sont des prédicats à deux arguments. Afin de les identifier clairement, nous les représenterons dorénavant en lettres capitales dans les propositions.

Prenons la définition (partielle) de *bras de individu* :

1 : **partie du corps de individu**

2 : *1 **ETRE_DE_FORME** articulé

L'unité lexicale définitionnelle X **ÊTRE_DE_FORME** Y sera définie ainsi dans le lexique définitionnel :

X ÊTRE_DE_FORME Y : x=`|partie du corps|`, y=`|forme|`

Les types attribués aux arguments de l'unité lexicale définitionnelle sont, rapellons-le, des étiquettes sémantiques puisque ces arguments sont réalisés par des unités lexicales de la langue décrite.

Ainsi la proposition 2' provoquera une erreur dans l'analyse car l'unité lexicale *bleu* n'est pas du type requis par le prédicat X ÊTRE_DE_FORME Y (c'est-à-dire n'est pas étiquetée `|forme|`).

1' : `partie du corps de individu`

2' : `*1 ETRE_DE_FORME bleu`

Si l'on fait le parallèle avec les représentations lexicales qui ont recours à des traits sémantiques (des couples *attribut : valeur(s)*), les unités lexicales du métalangage correspondent aux attributs et les unités lexicales de la langue décrite correspondent aux valeurs associées à ces attribut.

Le type du prédicat principal, selon qu'il représente l'attribut ou la valeur, va déterminer le type de la proposition (cf 3.2.2).

2.2.2 Une aide à l'encodage des données

Le fait d'introduire certaines contraintes dans le langage définitionnel risque, comme nous l'avons dit, de nuire au caractère naturel de la description BDéf. Les unités lexicales contrôlées du métalangage rendrons parfois les propositions 'lourdes de sens' puisque tout y sera explicité.

La BDéf ne s'adresse toutefois pas, comme le signalent eux-même ses auteurs, à un large public mais à un public restreint de lexicographes Sens-Texte que le langage contrôlé ne devrait pas décontenancer.

La LEC pose en principe qu'à un même sens ne doit correspondre qu'une seule expression, que cette expression ne doit pas être autant que possible ambiguë et qu'elle ne doit pas non plus figurer parmi les idiomatismes de la langue. Par exemple, le lexicographe préférera, pour exprimer la fonction d'un artefact, l'expression 'destiné à' plutôt que la préposition 'pour', jugée trop polysémique (cf Mel'cuk[95]). La définition d'un lexique définitionnel va donc dans le sens des principes de la LEC.

Nous pensons en outre que l'élaboration de la grammaire peut apporter une aide lors de l'encodage des données :

Une équipe lexicographique, étant donné l'ampleur de la tâche que représente la construction d'un dictionnaire, est nécessairement amenée à se renouveler. Il arrive souvent également que plusieurs équipes travaillent à un même projet en des lieux différents. Pour peu que les décisions d'encodage émanent de différentes personnes à différents endroits, il devient nécessaire

de mettre en place un moyen de normaliser les différentes productions, sans quoi on ne peut prétendre à une description formelle et homogène.

L'analyse d'une définition au moyen de la grammaire formelle va donc être un bon moyen de guider le lexicographe dans ses choix de rédaction.

Le lexicographe, en rédigeant une entrée, n'a, en effet, pas toujours connaissance des expressions choisies pour exprimer des sens récurrents qui apparaissent dans les structures lexicales. L'analyse des définitions va donc permettre de lui signaler qu'une des expressions qu'il a choisies n'est pas prévue par la grammaire et de lui soumettre les formes choisies d'ordinaire pour exprimer un sens donné. Libre à lui ensuite de trouver une justification à son expression de départ ou au contraire de choisir d'utiliser une des expressions prévues par la grammaire.

2.3 Une grammaire formelle du métalangage de la BDéf

Nous souhaitons dégager de l'observation d'un ensemble de lexies |**partie du corps**| les règles générales de la syntaxe d'une définition BDéf ainsi que le lexique définitionnel propre à ce type de lexies.

La tâche de description est, comme on le verra, parfois délicate étant donné le caractère encore non stable du formalisme de la BDéf.

2.3.1 Structure de la composante centrale

La composante centrale d'une définition, encodée dans le premier sous-champ du champ Définition d'une fiche BDéf, est un bloc définitionnel et consiste donc en un ensemble de propositions.

Nous allons, dans un premier temps, décrire la structure de la proposition principale (la première proposition) de la composante centrale des lexies |**partie du corps**|, en sachant que, pour ce type de lexie, la composante centrale se limite en général à cette seule proposition.

Nous nous intéresserons ensuite aux lexies dont la composante centrale contient plus d'une proposition en s'interrogeant sur la notion de paraphrase minimale.

Composition de la proposition principale.

La proposition principale des lexies |**partie du corps**| peut prendre deux formes :

une relation prédicat relationnel-argument . Dans le cas de nos lexies |**partie du corps**|, le prédicat relationnel exprime la relation de méronymie (PARTIE DE).

FRONT#I.c

1: PARTIE DE de tête#I.1b de X

L'argument du prédicat est soit l'holonyme (CORPS) soit une lexie |**partie du corps**|. On peut avoir également une relation de méronymie entre la lexie vedette et une partie du corps non lexicalisée :

VENTRE#I.1a

1: PARTIE DE partie_inférieure de corps#II.1d de X

une lexie unique étiquetée |**partie du corps**|.

POUCE#I.a

1: doigt#I.a de X

Nous proposons, dans le second cas, de rétablir le prédicat relationnel TYPE DE indiquant que l'argument de ce prédicat relationnel est l'hyperonyme de la lexie vedette.

La paraphrase minimale.

Les auteurs de la BDef définissent la composante centrale comme la **paraphrase minimale** de la lexie c'est-à-dire une première formulation suffisante de son sens.

La question de la délimitation de la composante centrale recoupe celle, ancienne en lexicologie, du choix du genre prochain. Elle repose en fait sur l'**intuition** du lexicographe qui doit percevoir le moment où les propositions de la composante centrale forment la paraphrase minimale.

La composante centrale de la lexie POUCEI.a est :

1: doigt#Ia de X

2: *1 le_plus large

Elle pourrait peut-être aussi bien être limitée à la proposition 1' :

1': TYPE DE doigt#Ia

La part d'intuition qui entre dans le choix de la délimitation d'une composante centrale en rend difficile la description formelle.

2.3.2 Structure des autres blocs définitionnels

La composante centrale est un bloc définitionnel dont au moins une des fonctions est identifiée : celle d'encoder la paraphrase minimale du sens de la lexie décrite.

La fonction des autres blocs définitionnels n'est en revanche pas clairement identifiée. On peut envisager trois fonctions pour un bloc définitionnel au sein d'une définition :

Fonction 1 : servir à rendre compte de la cooccurrence lexicale libre de la lexie.

Fonction 2 : servir à rendre compte de la cooccurrence lexicale restreinte de la lexie.

Fonction 3 : servir à indiquer un pont sémantique entre la lexie vedette et une autre lexie du vocable.

Nous nous proposons d'étudier la structure des blocs définitionnels des lexies **|partie du corps|** afin de voir si il est possible d'identifier leur fonction au sein de la définition.

Nous avons dit que les unités lexicales du métalangage n'étaient dans l'état actuel de la BDéf pas toujours explicitées au sein des propositions. Nous allons donc porter dans un premier temps notre attention sur les titres des blocs définitionnels pour pouvoir opérer des regroupements. Nous caractériserons ensuite la structure des différents blocs identifiés en explicitant dans les propositions les informations lexicales manquantes.

Les titres de blocs définitionnels des lexies **|partie du corps|**.

La première colonne du tableau 2.2 contient les différents titres de blocs définitionnels des lexies **|partie du corps|**². La seconde colonne reprend le titre dans une formule qui explicite sa structure.

Les titres simples de la ligne **1** représente le patron lexical associé à ce type de lexies, c'est-à-dire l'ensemble des caractéristiques de ce qui est dénoté par une lexie **|partie du corps|** susceptibles d'entrer dans la définition d'une lexie **|partie du corps|**.

Les titres de blocs des lignes **2** et **3** ont une structure complexe de type **(partieGauche.partieDroite)** où la partie droite est l'argument sémantique de la partie gauche.

Dans le cas des lexies **|partie du corps|**, la partie gauche est le titre simple **Parties**, la partie droite est également un titre simple.

²Cette liste constitue l'ensemble des attributs possibles d'une lexie étiquetée **|partie du corps|**. Cela signifie que tous ces attributs ne sont pas nécessairement définis dans chacune des structures des lexies.

	TITRES DE BLOC	GLOSE
1	/*Fonction physiologique*/ /*Localisation spatiale*/ /*Fonction*/ /*Parties*/ /*Forme*/	Fonction physiologique/ Parties/etc de la lexie vedette
2	/*Parties.Localisation*/ /*Parties.Forme*/ /*Parties.Autres caractéristiques*/	Localisation/Forme/etc d'une Parties de la lexie vedette
3	/*PartiesI.Localisation*/ /*PartiesII.Forme*/	Localisation/Forme d'une PARTIES nommée de la lexie vedette
4	/*Lien métaphorique*/	Il y a un lien métaphorique entre la lexie vedette et une autre lexie du vocable
5	/*Lien métaphorique.Fonction*/ /*Lien métaphorique.Forme*/	Il y a un lien métaphorique entre la lexie vedette et une autre lexie du vocable et ce lien repose sur le bloc Forme/Fonction de la lexie vedette

FIG. 2.2 – Titres de blocs définitionnels des lexies |**partie du corps**|

L'exemple ci-dessous reprend une partie de la définition de BRAS#I.1a proposée en annexe.

BRAS#I.1a

Ex : elle m'a assomé avec ses bras

/*Parties*/

2: *1 constitué de n1=2{partie}

/*Parties.Localisation_spatiale*/

3: *n1 latéral

4: *n1 supérieur

5: *n1 symétrique

/*Parties.Forme*/

6: *n1 allongé

7: *n1 articulé

/*Parties.Autres caractéristiques*/

8: *n1 mobile

9: *n1 distinct de corps(1)#II1d de X

Remarque 1 : la partie droite du titre complexe */*Parties.Autres Caractéristiques*/* montre que la BDef est encore en développement et que la liste des titres simples, représentant les *caractéristiques* d'une lexie **|partie du corps|**, n'est pas encore close.

Remarque 2 : Etant donné que ces blocs semblent dépendre d'un bloc définitionnel élémentaire de la définition, nous les appellerons **sous-bloc définitionnel**.

Les titres de blocs des lignes **4** et **5** servent à rendre compte explicitement³ d'un **pont sémantique** entre la lexie vedette et une autre lexie du vocable. Nous définissons pour l'heure un bloc indicateur de pont sémantique comme *un bloc constitué d'une proposition élémentaire contenant une lexie plus simple du vocable*.

La structure complexe (**partieGauche.partieDroite**) des blocs indicateurs de pont sémantique (ligne **5** du tableau) ne doit pas être interprétée de la même manière que les titres complexes de sous-blocs :

La partie gauche indique le type du pont sémantique. Dans le cas de nos lexies **|partie du corps|**, ce type est toujours **Lien métaphorique**. La partie droite indique le bloc de la lexie plus simple sur laquelle porte le pont sémantique.

Dans l'exemple donné ci-dessous, les deux blocs définitionnels indiquent qu'il y a un **Lien métaphorique** entre BRAS#I1b et BRAS#I1a et que ce lien repose sur la **Forme** de ce qui est dénoté par la lexie plus simple. Comme on peut le voir dans les fiches BRAS#I1b et BRAS#I1a données en annexe, la partie droite du titre */*Lien_métaphorique.Forme*/* du bloc de la définition BRAS#I1b ne correspond pas exactement au bloc de la définition de BRAS#I1a auquel il réfère */*Parties.Forme*/*.

BRAS#I1b

Ex: La blonde héroïne tremblait entre les bras de l'énorme gorille

*/*Lien_métaphorique.Forme*/*

6: forme de *1 ressembler à forme de {bras#I1a} de individu

*/*Lien _métaphorique.Fonction*/*

³Il y a d'autres réalisations possibles d'un pont sémantique comme on le verra au prochain chapitre

7: fonction de *1 ressembler à fonction de {bras#I1a} de individu

Les deux types de propositions d'un bloc définitionnel

On distingue deux grands types de propositions, selon le type de leur prédicat principal :

Le **prédicat indicateur de bloc** est un prédicat qui reprend au sein d'une proposition l'information apportée par le titre du bloc définitionnel.

4: *1 destiné à *5

Le prédicat de cette proposition (DESTINÉ À) est le prédicat qui représente la fonction (/Fonction*/) de la lexie définie.

Ce type de prédicat est contraint et constitue, comme nous l'avons dit, un prédicat à deux arguments du métalangage définitionnel.

Dans l'état actuel de la BDéf, ce prédicat est celui de la première proposition du bloc définitionnel, il est unique et associé au titre du bloc. On peut relever, dans l'état actuel de la BDéf, trois exceptions à cette règle :

1. certains blocs définitionnels ne possèdent pas d'indicateur de bloc (c'est, on l'a vu, une des justifications de la présence d'un titre pour chaque bloc définitionnel).

NEZI1.a

/*Forme*/

3: *1 saillant

4: *1 allongé

2. On peut avoir plusieurs prédicats indicateurs (ÊTRE ORGANE DE et ÊTRE UN_DES_ORGANES_DE pour un seul bloc définitionnel).

NEZI1.a

/*Fonction physiologique*/

5: *1 être organe de odorat

6: *1 être un_des_organes de respiration

3. un indicateur de composante peut être discontinu (APPARENCE+EXPRIMER).

NEZI1.a

/*Fonction sémiotique[optionnel]*/

7: apparence de *1 exprimer Y

Dans le cas de l'exception 1, le prédicat indicateur de bloc devra être rétabli :

3: *1 ÊTRE_DE_FORME saillant

Dans le cas de l'exception 2, il faudra veiller à s'assurer que les prédicats indicateur de blocs expriment des sens définitionnels différents, autrement

dit ils ne devront pas être synonymes.

Dans le cas de l'exception 3, le prédicat indicateur de bloc, pour une question d'homogénéité dans l'analyse, prendra la forme d'un prédicat non discontinu :

7: *1 EXPRIME_PAR_SON_APPARENCE Y

Une proposition dont le prédicat principal est un PRédictat Indicateur de Bloc (PRIB) est une Proposition Indicatrice de Bloc (PIB).

Le second type de prédicat est, contrairement au prédicat indicateur de bloc, un prédicat appartenant à la langue décrite. Il occupe, au sein d'un bloc définitionnel, la position du second argument d'un prédicat indicateur de bloc.

Le second argument d'un prédicat indicateurs de blocs peut être :

1. une proposition (*11). Le prédicat de cette proposition ne pourra pas être un indicateur de bloc.
10: *1 permettre#II à X de *11
11: X faire {action_physique}
2. un prédicat (au milieu de).
2: *1 situé au_milieu de *n1
3. une lexie appartenant au même vocable que la lexie vedette (pouce#Ia).
POUCEI.d de X:
3: *1 rappeler pouce#Ia
4. un actant sémantique de la lexie (Z).
LEVREI.1a (Z) de X [énonçant Y]:
7: apparence de *1 exprimer Z

Le cas 1 renvoie au second type de proposition identifié :

Une proposition dont le prédicat principal est un PRédictat Valeur (PRV) est une Proposition Valeur (PV).

2.3.3 Grammaire formelle

Nous reprenons ici, sous la forme de règles de réécriture, la grammaire formelle ébauchée au cours de cette section.

Une définition (DEF) est constituée d'une composante centrale (CC) et d'un ensemble de blocs définitionnels (BD).

$$\begin{aligned} \text{DEF} &\rightarrow \text{CC BD} \\ \text{BD} &\rightarrow \text{BD} \mid \epsilon \end{aligned}$$

Une composante centrale (CC) est constituée d'une proposition indicatrice

de composante centrale (PICC).

$$CC \rightarrow \text{PICC}$$

Une proposition indicatrice de composante centrale (PICC) est constituée d'un prédicat relationnel (PrR) et du second argument (Arg2) du PrR.

$$\begin{aligned} CC &\rightarrow \text{PrR Arg2} \\ \text{PR} &\rightarrow \text{TYPE DE} \mid \text{PARTIE DE} \end{aligned}$$

Un bloc définitionnel (BD) est constitué d'une proposition indicatrice de bloc PIB.

$$BD \rightarrow \text{PIB}$$

Une proposition indicatrice de bloc (BIP) est soit une proposition indicatrice de pont sémantique (PIPS), soit une proposition indicatrice de bloc non identifié (PIBNI).

$$BIP \rightarrow \text{PIPS} \mid \text{PIBNI}$$

Une proposition indicatrice de pont sémantique (PIPS) est constituée d'une proposition indicatrice de composante centrale (PICC), d'un prédicat indicateur de pont sémantique (PrIPS) et d'une lexie du vocable (LV).

$$\begin{aligned} \text{PIPS} &\rightarrow \text{PICC PrIPS LV} \\ \text{PrIPS} &\rightarrow \text{RAPPELLE} \mid \text{RAPPELLE_PAR_LA_FORME} \mid \\ &\quad \text{RAPPELLE_PAR_LA_FONCTION} \mid \dots \\ \text{LV} &\rightarrow \text{lexie du vocable} \end{aligned}$$

Une proposition indicatrice de bloc non identifié (PIBNI) est constituée d'un prédicat indicateur de bloc (PrIB) et des deux arguments (Arg1,Arg2) du PrIB.

$$\begin{aligned} \text{PIBNI} &\rightarrow \text{Arg1 PrIB Arg2} \\ \text{PrIB} &\rightarrow \text{ÊTRE_DE_FORME} \mid \text{DESTINÉ_À} \mid \text{ÊTRE_PLACÉ} \mid \dots \end{aligned}$$

Le premier argument (Arg1) est soit un prédicat valeur (PrV), soit une proposition valeur (PV), soit une variable d'actant (AVar) soit une proposition indicatrice de bloc (PIB).

$$\text{Arg1} \rightarrow \text{PrV} \mid \text{PV} \mid \text{AVar} \mid \text{PIB}$$

Le second argument (Arg2) soit un prédicat valeur (PrV), soit une pro-

position valeur (PV), soit une variable d'actant (AVar).

$$\text{Arg2} \rightarrow \text{PrV} \mid \text{PV} \mid \text{AVar}$$

Une proposition valeur (PV) est constituée d'un prédicat valeur (PrV).

$$\text{PV} \rightarrow \text{PrV}$$

Un prédicat valeur (PrV) est une lexie de la langue définie.

$$\text{PrV} \rightarrow \text{lexie}$$

Chapitre 3

Un modèle calculable

Nous avons proposé, dans le chapitre précédent, une ébauche de grammaire formelle permettant d'analyser une définition BDef. La figure 3.1 propose une des représentations abstraites possibles obtenue à l'issue de l'analyse.

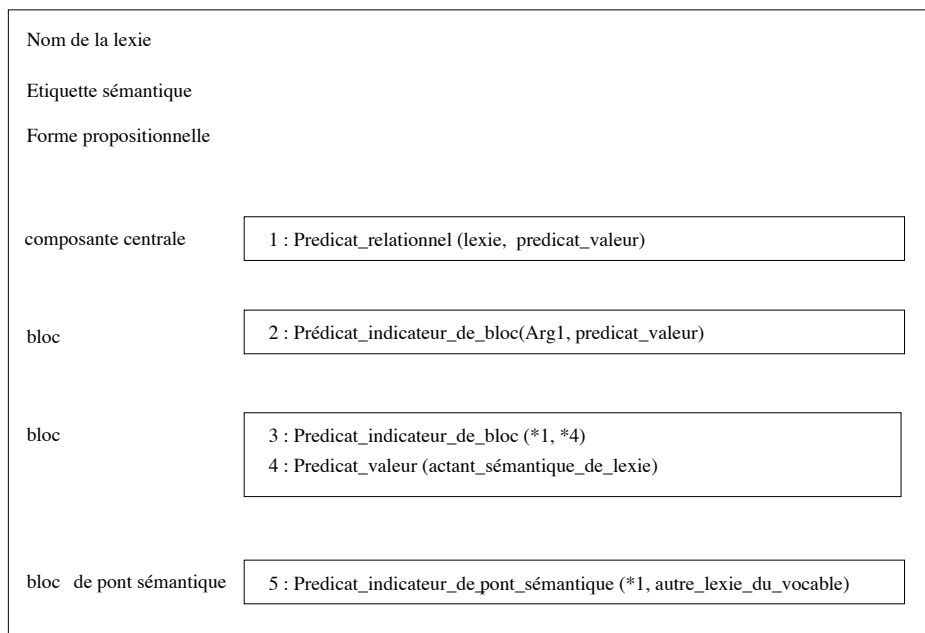


FIG. 3.1 – Représentation abstraite d'une définition BDef

Nous allons à présent nous intéresser, au cours de ce chapitre, aux différents calculs sémantiques qui vont pouvoir être effectués à partir de cette représentation abstraite.

Il nous faut avant cela définir clairement ce que l'on entend par *calcul sémantique*.

Nous définissons un *calcul sémantique* comme une opération effectuée sur un ensemble de structures sémantiques et donnant comme résultat :

- une structure sémantique (un calcul qui permette par exemple de composer plusieurs sens)
- un booléen (un calcul qui dise par exemple si oui ou non deux sens sont identiques).

Le type d'un calcul sémantique dépend des structures qui apparaissent en opérandes :

1. Les calculs dépendent de l'unité sémantique représentée par la structure : le mot, la phrase, le discours.
2. Les calculs dépendent des aspects de l'unité sémantique représentés par la structure.
3. Les calculs dépendent du degré de formalisation de la structure.

Choix de l'unité sémantique traitée. Selon les objectifs linguistiques qu'elle s'est fixés, une théorie de sémantique formelle propose une structure sémantique qui prenne ou non en compte les différentes unités sémantiques. La Théorie de la Représentation du Discours par exemple prévoit une structure sémantique représentant le sens d'une phrase (non le sens des mots qui la composent) et calcule, à partir de phrases organisées en discours, la structure sémantique représentant le sens du discours. La Théorie Sens-Texte (TST), à l'inverse, doit décrire les sens lexicaux qui vont lui permettre de réaliser son objectif qui est de générer des ensembles de paraphrases (cf. chapitre 1). Cet objectif explique que la TST ne prenne pas en compte la représentation du discours.

Traitement des différents aspects de l'unité sémantique. Si, par exemple, la représentation lexicale ne prend pas en compte la description de la combinatoire lexicale, il sera impossible de calculer le sens d'une expression semi-figée à partir de ces parties : si on ne représente pas le fait que *noire* prend un sens spécial auprès de *colère*, on ne pourra pas calculer le sens de l'expression *colère noire*¹.

Degré de formalisation de la structure. Lorsqu'une représentation lexicale prend la forme d'une définition lexicographique, le langage utilisé dans cette définition doit être contraint pour pouvoir être aisément analysé (cf. chapitre 2). Si le langage définitionnel utilise, comme la fait la langue naturelle, différentes expressions d'un même sens (*pour, a pour fonction de, destiné à, faire en sorte que*, etc. peuvent toutes exprimer la fonction de ce

¹le type de calcul évoqué par cet exemple, qui renvoie à la notion de compositionnalité du sens, ne sera pas pris en compte dans la suite de ce chapitre.

qui est dénoté par l'unité lexicale décrite) il sera plus difficile de décrire la démarche à suivre pour isoler ce sens que si le langage ne prévoit qu'une manière de l'exprimer (par exemple, utiliser toujours DESTINÉ À pour exprimer la fonction de ce qui est dénoté par l'unité lexicale).

Ce dernier critère concernant la structure sémantique que l'on veut soumettre à un calcul est celui qui a retenu jusqu'à présent notre attention. La première section présentera les opérations élémentaires que l'on pense pouvoir effectuer sur la représentation abstraite d'une définition BDéf. On évoquera également l'intérêt de tels calculs pour la recherche en sémantique lexicale.

Dans la seconde section, nous mentionnerons un autre type de calcul qui pourra être effectué en partie sur cette représentation abstraite : le calcul de liens entre blocs définitionnels et éléments de la représentation de la combinatoire lexicale d'une unité.

3.1 calculs sémantiques sur les définitions

3.1.1 Les opérations élémentaires

Nous donnons ici les opérations élémentaires qui peuvent être effectuées sur la représentation abstraite des définitions BDéf. Ces opérations élémentaires concernent l'identification d'un élément au sein d'une définition et la comparaison de définitions.

Recherche d'un élément dans une définition

Fonction qui prend en argument une définition et renvoie la composante centrale de cette définition.

`DonneCC(Def)`

Fonction qui prend en argument une Valeur (ce qui correspond à un Arg2 dans la grammaire). Renvoie le Prédicat Indicateur de Bloc qui lui est associé.

`DonnePrIB(Valeur)`

Fonction qui prend en argument un PRédicat Indicateur de Bloc. Renvoie la valeur qui lui est associée.

`DonneValeur(PrIB)`

Fonction qui prend en argument une Lexie du vocable. Renvoie le PRédicat Indicateur de Pont Sémantique qui lui est associé.

`DonnePrIPS(L)`

Comparaison de définitions

Fonction qui compare deux blocs définitionnels élémentaires. Les blocs sont identiques si ils ont le même PRÉdicat Indicateur de Bloc et la même valeur.
`blocsIdentiques(bloc1,bloc2)`

3.1.2 Utilisation des calculs

Recherche sur la polysémie

La LEC propose un traitement de la polysémie fondée sur l'énumération des sens : une unité polysémique correspond à un vocable et chaque sens de l'unité polysémique correspond à une lexie.

Comme dans les dictionnaires classiques, les sens sont classés en fonction d'un sens premier que les lexicographes de la LEC élisent sur la base de leur intuition et sur celle du *principe de décomposition maximale* : La première lexie du vocable, appelée **lexie de base**, doit être plus simple que les autres lexies du vocable.

La notion de **pont sémantique** est déterminante dans le traitement de la polysémie proposée par la LEC. Un pont sémantique est défini dans Mel'čuk[95] comme *une composante caractéristique commune à deux lexies d'un même vocable*.

La composante commune est tantôt explicite, tantôt implicite. Pour le montrer, nous reprenons l'exemple donné dans Mel'čuk[95] :

BUREAUI 'table destinée à faire des travaux écrits dessus'

bureau de style empire

BUREAUII 'pièce destinée à faire des travaux écrits dedans'

je suis entré dans la bureau de l'écrivain

BUREAUIII.1 'organisme destiné à faire des travaux écrits'

le 3e bureau, le bureau de planification de la mairie

BUREAUIII.2 'personne travaillant ensemble dans un BUREAUIII.1'

Son bureau lui a offert des fleurs

BUREAUIII.2 'groupe de personnes travaillant ensemble pour assurer la direction d'une association'

Bureau du parti radical

Un pont sémantique peut être réalisé de trois manières différentes :
Soit les trois lexies L1, L2, L3.

1. L3 utilise une lexie plus simple (L1 ou L2) dans sa définition. Dans l'exemple donné ci-dessus, la lexie BUREAU**III.2** utilise la lexie BUREAU**III.1** dans sa définition².
2. L3 et L2 ont une partie de leur sens en commun, c'est à dire une ou plusieurs composante de sens en commun. Les lexies BUREAU**I**, BUREAU**II** et BUREAU**III.1** partagent la composante *destiné à faire des travaux écrits*.
3. L3 est relié à L1 par un pont sémantique indirect autrement dit il y a un pont entre L3 et L2 et entre L2 et L1. Les lexies BUREAU**III.2** est reliée à la lexie BUREAU**I** par l'intermédiaire de BUREAU**III.2** et de BUREAU**III.1**

Dans les cas **1** et **2**, le pont sémantique est explicite, dans le cas **3**, il est implicite.

Concernant le cas **1**, les calculs proposés nous permettent de connaître directement le bloc définitionnel dans lequel apparait la lexie qui constitue le pont sémantique et de savoir quel est le type de ce pont sémantique. (DonnePRIPS(L)).

Concernant les cas **2** et **3**, l'opération élémentaire de comparaison permet de connaître les blocs identiques au sein des définitions d'un même vocable et d'identifier ainsi les différents ponts sémantiques du vocable.

Il sera alors envisageable de voir en quoi la nature d'un pont sémantique est significative de la **distance sémantique** entre deux lexies.

Dans les dictionnaires Sens-Texte ayant recours aux définitions, les lexies d'un vocable sont en effet dotées d'un numéro qui, outre sa fonction distinctive, donne une indication sur les distances sémantiques entre les lexies. Grosso modo, les chiffres romains marquent une distance importante, les chiffres arabes une distance moindre et les lettres une courte distance.

Dans l'exemple donné ci-dessus, les définitions des trois premières lexies, qui ont chacune un chiffre romain différent, contiennent chacune deux composantes : la composante centrale et la composante *destiné à faire des travaux écrits*. Est-ce à dire qu'une composante centrale différente établit automatiquement une distance importante entre deux lexies ? Pourtant, le calcul permettant de comparer des blocs dira que les composantes centrale des lexies BUREAU**III.1** et BUREAU**III.2**, ne sont pas identiques bien qu'il y est un rapport sémantique fort entre *organisme* et *groupe*.

On voit que les calculs ne vont pas toujours apporter une réponse directe au besoin de caractériser de manière précise la description lexicale. En l'occurrence, le problème évoqué montre l'importance d'avoir une hiérarchie

²Rappelons que cela constitue la définition partielle que nous avons donnée d'un pont sémantique lors de l'identification des blocs définitionnels (cf. 2)

d'étiquettes sémantiques sur laquelle on puisse opérer conjointement des calculs.

Recherche sur la dérivation sémantique

La notion de partage de composantes non triviales utilisée pour rendre compte du fait de polysémie est également pertinente concernant les relations inter-vocable et plus précisément les relations de dérivation sémantique. Le formalisme adopté va, par exemple, permettre d'entreprendre une étude systématique sur le degré de synonymie existant entre deux lexies : la question de savoir à partir de quand un hyperonyme d'une lexie L n'est plus considéré comme un quasi-synonyme de cette lexie, autrement dit qu'elle est la taille ou l'importance de la composante de sens nécessaire pour établir un lien de synonymie pourra probablement trouver une réponse grâce aux calculs envisagés.

Ces calculs trouveront leur place dans des applications variées telles que la génération automatique de texte (entre autre au moment de la lexicalisation) et la traduction automatique³ (voir Altman&Polguère[03] et Wanner[03])

La possibilité de comparer des définitions sur la base de la représentation abstraite obtenue au terme de l'analyse des définitions BDef va permettre plus généralement de vérifier les liens de dérivation encodés en terme de FL.

On trouve dans la littérature Sens-Texte une typologie des antonymes fondée sur le type de composantes apparaissant dans les lexies qui sont en relation d'antonymie et sur la place de cette composante au sein de la définition (cf. Milićević[03]).

On y distingue deux grands types de relation d'antonymie entre deux lexies selon que :

1. Une des deux lexies ne diffère de l'autre que par une négation.

$$(\text{Anti}_{neg}(L1) = L2)$$

X est présent à Y = X se trouve à un endroit Y

X est absent de Y = X NE se trouve PAS à un endroit Y

2. Les deux lexies ne diffèrent que par 'plus'/'moins'.

$$(\text{Anti}_{<}(L1) = L2)$$

X augmente de Y = la valeur v du paramètre X

devient PLUS grande qu'avant de Y

X diminue de Y = la valeur v du paramètre X

devient MOINS grande qu'avant de Y

³le formalisme de la BDef a été développé pour la description d'une partie du lexique du français. L'utilisation d'une telle base de données dans des applications de traduction automatique implique bien sûr qu'elle ait des correspondants dans d'autres langues.

Le type 1 est subdivisé en deux sous-types selon que :

1. La négation porte sur la composante centrale de la lexie.

(*Anti_{neg}non*(L))

X est présent à Y = X se trouve à un endroit Y

X est absent de Y = X NE se trouve PAS à un endroit Y

2. La négation porte sur un bloc autre que la composante centrale.

(*Anti_{neg}-*(L))

X apparaît dans Y = X commence à se trouver dans Y

X disparît de Y = X commence à NE PAS se trouver dans Y

La FL *Anti_{neg}non* met en relation deux lexies dont le sens ne diffère que par une négation dans une des composantes centrales. Si le résultat de la comparaison des composantes centrales ne correspond pas à la définition de la FL, c'est qu'il y a une erreur d'encodage.

Nous sortons alors du seul domaine de la définition et considérons à présent la question du lien entre les éléments de la définition et ceux de la zone de combinatoire lexicale.

3.2 Une étape en vue d'autres calculs

La représentation abstraite des définitions va faciliter la mise en correspondance des liens entre la zone sémantique et la zone de combinatoire lexicale restreinte.

3.2.1 Transparence du lien entre le sens d'une lexie et sa combinatoire lexicale

Le PRINCIPE DE COHÉRENCE INTERNE énoncé par la LEC veut que *dans un DEC, les descriptions sémantique, syntaxique et cooccurentielle de la lexie vedette L montrent un accord complet*⁴.

Comme nous l'avons déjà signalé, le DEC papier, du fait d'une définition lexicographique dont la structure n'est pas explicitée, n'est pas à même de mettre en évidence cette correspondance entre les informations consignées dans les trois zones d'un article. Cette opacité conduit à différentes erreurs d'encodages, symptomatiques d'une lacune dans la formalisation du langage des fonctions lexicales aussi bien que de celui de la définition lexicographique, comme il est montré dans Kahane[01].

Du point de vue de l'encodage des données, la description du sens d'une

⁴« Par accord, nous entendons ici une correspondance réciproque -explicitement indiquée- entre les composantes sémantiques dans la définitions de L, ses dépendants syntaxiques actanciels et ses cooccurents lexicaux » (Mel'čuk[95]) p.36

lexie se fait simultanément avec celle de sa combinatoire lexicale :

Les informations sémantiques et grammaticales de la lexie définissent un sous-ensemble de relations sémantiques et lexicales possibles.

C'est l'idée, adoptée également dans WordNet, selon laquelle certains liens sémantiques sont propres à la partie du discours de l'unité lexicale : toutes les FL qui prennent par exemple un nom comme mot-clé seront évidemment exclues de la description d'une unité lexicale verbale. De même, si l'unité décrite est un objet sémantique, c'est-à-dire si elle ne dénote pas une situation faisant intervenir des participants (ce à quoi correspond un prédicat), on évitera naturellement de l'appliquer à des FL de type S_i , qui donnent comme valeur les noms typiques des participants. En délimitant bien le sous-ensemble des FL pertinentes du point de vue de la lexie, on diminue par là même le risque de trouver encodées dans l'article des informations surfaîtes.

La méthodologie d'encodage des fonctions lexicales nécessite donc une définition, au moins partielle, du sens de la lexie ainsi que de ces caractéristiques grammaticales.

A l'inverse puisque la définition lexicographique encode le sens linguistique de l'unité lexicale, les composantes de la définition trouvent leur justification dans l'utilisation de la langue, ce qu'encodent en partie les fonctions lexicales.

Lorsque le lexicographe encode les liens collocationnels d'une lexie et que cette lexie a un Réal (qui réfère à l'utilisation typique de l'objet dénoté par la lexie) il saura que la définition devra contenir un bloc de sens relatif à la fonction de l'objet.

Une bonne caractérisation d'une unité lexicale naît donc du va et vient entre la zone sémantique et la zone de combinatoire lexicale. La représentation abstraite des définitions BDéf sera appropriée pour expliciter 'l'historique' de la description lexicale.

3.2.2 Un formalisme approprié

Le développement de la base de données BDéf, consécutif à celui de la base de données DiCo (Polguère[03]), a permis de remettre à l'ordre du jour la question de la mise en relation effective des informations de la zone sémantique avec celles de la zone de combinatoire lexicale.

La base de données lexicales DiCo encode les informations lexicales qui ne sont pas traitées par la BDéf : les informations sur la combinatoire syn-

taxique et celles sur la combinatoire lexicales restreinte de la lexie⁵.

Nous avons indiqué dans la chapitre précédent la structure d'un **bloc définitionnel**. Nous appellerons par analogie **bloc collocationnel** un couple $FL(\text{Lexie vedette}) = \text{valeur}(s)$.

La figure 3.2 montre une relation entre un bloc définitionnel du sens de la lexie BOISSON et un bloc collocationnel de la zone de combinatoire restreinte de cette lexie.

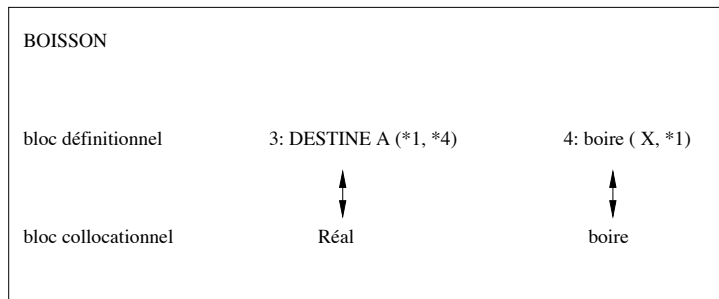


FIG. 3.2 – Mise en relation entre un bloc définitionnel et un bloc collocationnel

Le lien, dans l'exemple donné, peut être établi relativement aisément : concernant la lexie BOISSON, la FL Réal, dont le sens est *réaliser les "objectifs" inhérents de la chose désignée par le mot-clé* (cf Mel'čuk[95]), est clairement en lien direct avec le PrIB DESTINÉ À(X,Y). La FL ne renvoie en outre ici qu'une valeur, boire, que l'on peut ainsi mettre directement en rapport avec le PrV boire(x,y).

L'établissement du lien entre les deux types d'éléments, qui ne va pas de soi contrairement à ce que pourrait laisser entendre l'exemple donné ci-dessus, pose un certain nombre de questions :

Bi-univocité? Il faudra se poser la question de savoir si il y a bi-univocité entre bloc définitionnel et bloc collocationnel ou si au contraire un bloc définitionnel peut être mis en relation avec plusieurs blocs collocationnels et à l'inverse un bloc collocationnel avec plusieurs blocs définitionnels.

Bonne délimitation des blocs? La question précédente ne pourra être résolue que par tâtonnements étant donné que l'on n'a pas encore les

⁵Une fiche DiCo propose également une caractérisation sémantique de la lexie sous forme d'étiquette sémantique (cf Polguère[03])

moyens de s'assurer du caractère 'élémentaire' d'un bloc.

Dans le cas des blocs définitionnels, nous avons déjà évoqué le problème de la délimitation de la composante centrale. Dans le cas des blocs collocationnels, certaines FL renvoient un nombre important de valeurs montrant des nuances de sens qui pourraient être représentées au moyen de FL sémantiquement plus complexes.

Considérer tous les blocs collocationnels ? On peut supposer qu'il faudra effectuer un tri dans les blocs collocationnels. Les FL sémantiquement vides, comme par exemple les dérivés syntaxiques, ne seront pas, a priori, mises en relation avec un bloc de la définition. Il n'est toutefois pas toujours évident de déterminer précisément quelle est l'information sémantique véhiculée par un bloc définitionnel et dans quel type d'élément du bloc collocationnel elle trouve son origine.

Conclusion

Ce mémoire s'est intéressé à une représentation du sens lexical, celle de la Lexicologie Explicative et Combinatoire, et à un formalisme associé à une partie de cette représentation, celui de la BDéf, une base de données lexicales encodant les définitions Sens-Texte.

Le but principal de ce formalisme est de mettre en lumière la structure du sens lexical et d'autoriser un traitement automatique des définitions.

Nous nous sommes proposés de voir si ce but était en voie d'être atteint en décrivant – parfois de manière normative – , le formalisme de la BDéf au moyen d'une grammaire formelle. Nous avons ensuite défini des calculs élémentaires opérables sur la représentation abstraite des définitions BDéf obtenue à l'issue de l'analyse.

Notre description du formalisme de la BDéf s'est toutefois fondée sur l'observation d'un nombre restreint de lexies, qui plus est appartenant toutes au même type sémantique. Il faudra par la suite élargir cette observation à d'autres ensembles de lexies pour, d'une part, voir si les règles de la grammaire formelle que nous avons proposée sont bien applicables à tous les types de lexies et pour, d'autre part, poursuivre la description du lexique du métalangage définitionnel de la LEC.

Une des principales questions que nous avons laissées en suspens concerne l'identification précise de la fonction des blocs de sens formant la structure d'une définition. Nous avons déterminé de manière formelle la (ou une des) fonction de certains blocs définitionnels, comme par exemple l'identification de blocs servant à rendre compte d'un pont sémantique entre les lexies d'un vocable, mais nous n'avons pu mettre en lumière la fonction précise des autres types de blocs définitionnels. Si l'on admet, avec les auteurs de la BDéf, que la définition n'encode que le sens linguistique d'un mot, c'est-à-dire un certain nombre d'aspects – ceux qui sont encodés par la langue – du concept auquel réfère une lexie, la réponse est probablement à chercher du côté des autres zones de la représentation lexicale de la LEC, notamment de la zone de combinatoire lexicale.

Références bibliographiques

- ALTMAN J., POLGUERE A. (2003) *La BDef : base de données dérivée du Dictionnaire explicatif et combinatoire*, Proceedings of the first International Conference on Meaning-Text Theorie (MTT-03), Paris, p.43-54.
- BOUILLON P. et al. (1998) *Traitement Automatique des Langues Naturelles*, AUPELF-UREF/Duculot, Louvain-la-Neuve.
- FILLMORE C. (1976) *Frame semantics and the nature of language*, in Annals of the New York Academy of Sciences : Conference on the Origin and development of Language and Speech, Volume 280 (pp.20-32).
- KAHANE S. (2001) *Une Blessure profonde dans le DEC. Sur le lien entre la définition lexicographique et les fonctions lexicales*, à paraître in F.Grossmann & A.Tutin, Modélisation des collocations. Journal de linguistique appliquée.
- 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. (1992), *Paraphrase et lexique : la théorie Sens-Texte et le Dictionnaire Explicatif et Combinatoire*. In MEL'CUK I. et al (1992).
- MEL'CUK I., CLAS A., POLGUERE A., (1995), *Introduction à la lexicologie explicative et combinatoire*, AUPELF-UREF/Duculot, Louvain-la-Neuve.
- MEL'CUK I. (1997), *Vers une linguistique Sens-Texte*, Collège de France, Paris.
- MILICEVIC J.(2003), *Etude 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.
- MILLER G. (ed) (1990) *WordNet : An On-Line Lexical Database*, in :

International Journal of Lexicographie, 3 :4, pp. 233-249.

NASR A. (1996) *Un modèle de reformulation automatique fondé sur la Théorie Sens-Texte : application aux langues contrôlées*, thèse de doctorat, Université de Paris 7.

PICOCHÉ J. (1997) *Précis de Lexicologie Française*, Nathan.

POLGUÈRE A. (1992), *Remarques sur les réseaux sémantiques Sens-Texte*. In A. Clas(Ed.) *Le mot, les mots...les bons mots*, Montréal : Presse de l'Université de Montréal, 109-148.

POLGUÈRE A. (1998), *La théorie Sens-Texte*, Dialogue, Vol.8-9, Université du Québec à Chicoutimi, pp.9-30.

POLGUÈRE A. (2002), *Notions de base en lexicologie*

POLGUÈRE A. (2003), *Étiquetage sémantique des lexies dans la base de données DiCo*, TAL, Volume 44- n2.

PUSTEJOVSKY J. (1995), *The Generative Lexicon*, The MIT Press.

WIERZBICKA A.(1992), *Back to definitions : Cognition, semantics and lexicography*, Lexicographica, 8 : 146-174.

Annexe A

fiches BDef

A.1 La fiche BRAS#I.1a

BRAS I.1a

Carac. Gram. :

EtiquSém : partie du corps

Forme_prop : \sim [pl] de X

Exemples : Elle m'a assomé avec ses bras.

Définition :

composante centrale :

1: partie_de_corps(1)#III1a de X

différences spécifiques :

/*Parties*/

2: *1 constitué de n1=2{partie}

/*Parties.Localisation_spatiale*/

3: *n1 latéral

4: *n1 supérieur

5: *n1 symétrique

/*Parties.Forme*/

6: *n1 allongé

7: *n1 articulé

```
/*Parties.Autres caractéristiques*/
8: *n1 mobile
9: *n1 distinct de corps(1)#II1d de X
```

```
/*Fonction physiologique*/
10: *1 permettre#II à X de *11
11: X faire {action_physique}
```

Typage :

X: individu

Rel_Sém :

A.2 La fiche BRAS#I.1b

BRAS I.1b

Carac. Gram. :

EtiquSém : partie du corps

Forme_prop : ~[pl] de X

Exemples : La blonde héroïne tremblait entre les bras de l'énorme gorille.

Définition :

```
composante centrale :
1: partie_de_corps(1)#II1b de X
```

différences spécifiques :

```
/*Parties*/
2: *1 constitué de n1=2{partie}
```

```
/*Parties.Localisation_spatiale*/
3: *n1 latéral
4: *n1 supérieur
5: *n1 symétrique
```

```
/*Lien_métaphorique.Forme*/
6: forme de *1 ressembler à forme de {bras#I1a} de individu
```

/*Lien _métaphorique.Fonction*/

7: fonction de *1 ressembler à fonction de {bras#I1a} de individu

Typage :

X: singe

Rel_Sém :