
Un modèle formel de descriptions lexicales

formalisme BDéf et structures de traits typées

Lucie Barque*, Alexis Nasr**

* *Laboratoire Lattice
UFRL
Université Paris 7 - Case 7003
2, place Jussieu,
75251 PARIS cedex 05
lucie.barque@gmail.com*

** *Laboratoire d'Informatique Fondamentale
CNRS - Aix Marseille Université
Parc Scientifique et Technologique de Luminy
163, avenue de Luminy - Case 901
F-13288 Marseille Cedex 9
alexis.nasr@lif.univ-mrs.fr*

RÉSUMÉ. L'article s'intéresse à la formalisation, dans une optique de calcul, des définitions lexicographiques BDéf (Altman et Polguère, 2003). La richesse informationnelle de ces définitions laisse envisager de nombreux calculs utiles aussi bien à la pratique lexicographique qu'à une réflexion plus générale sur la modélisation du sens lexical. De tels calculs ne peuvent toutefois être mis en œuvre sans une formalisation poussée des définitions, qui se traduira ici par le recours aux structures de traits typées (Carpenter, 1992). En permettant à chaque instant de s'assurer de manière automatique de la cohérence générale du lexique, la formalisation nous permet de mettre en place une nouvelle méthodologie de description lexicale par « allers-retours » successifs entre les définitions lexicographiques et une métadescription de ces dernières.

ABSTRACT. We propose in this paper a formalization of the BDéf lexicographic definitions proposed by (Altman et Polguère, 2003), in order to carry calculus on them. The information richness of such definitions suggest many calculus that are useful for lexicographic practice as well as for a more general reflection on modelling lexical semantics. Such calculus cannot be realized without a thorough formalization of such definitions that we propose to represent as typed feature structures (Carpenter, 1992). The formalization proposed allow to automatically check for the lexicon consistency, which suggest a new methodology for lexical description based on successive enrichment of the lexical database and the meta data that describe it.

MOTS-CLÉS : définitions lexicographiques formalisées, règles lexicales, structures de traits typées, calculs liés aux descriptions lexicales.

KEYWORDS: formalized lexicographic definitions, lexical rules, typed featured structures, computing lexical data.

1. Introduction

Le langage de la définition lexicographique est un métalangage, c'est-à-dire un langage conventionnel utilisé pour décrire un aspect de la langue, en l'occurrence le sens des unités lexicales. Parce qu'il s'agit d'un langage conventionnel très proche de la langue naturelle dont il décrit le lexique, les dictionnaires de langue traditionnels tels que *Le Petit Robert* ne mettent pas en évidence sa structure sous-jacente. Cela ne nuit pas à la consultation de ces dictionnaires et aurait même tendance à la simplifier. Mais cela ne convient plus dès lors que l'on souhaite expliciter l'organisation des définitions pour les soumettre à un traitement automatique.

L'étude présentée ici s'intéresse à la formalisation poussée du métalangage définitionnel de la BDéf, en vue d'une automatisation des descriptions utilisant ce métalangage. La BDéf (pour base de définitions) est un projet lexicographique dans lequel les définitions du *Dictionnaire explicatif et combinatoire*, ou DEC (Mel'čuk *et al.*, 1984, 1988, 1992, 1999)¹ sont encodées dans un format qui rend explicite la structuration des composantes définitionnelles, offrant ainsi une vision claire de l'organisation du sens lexical (Altman et Polguère, 2003). Toutefois, les descriptions en format BDéf nécessitent, pour être exploitées dans des calculs, une formalisation plus poussée que la formalisation actuelle, conçue pour les lexicographes. Parmi ces calculs, on peut citer le contrôle de la bonne formation syntaxique et sémantique des entrées, la possibilité de savoir s'il existe une relation sémantique particulière entre deux entrées ou encore le calcul de la différence de deux entrées, pour être à même de savoir ce qui distingue précisément deux entrées apparentées.

La caractérisation formelle du métalangage BDéf présente en outre un intérêt lexicographique important à l'heure du développement de lexiques sémantiques pour le TAL². L'explicitation des règles de bonne formation syntaxiques et sémantiques d'une définition ainsi que la possibilité d'en contrôler la satisfaction permettent en effet la mise en place d'une méthodologie de description lexicale par allers-retours successifs entre les définitions et les outils descriptifs servant à la construction de ces définitions. Les données et les métadonnées sont de fait amenées à évoluer au fur et à mesure qu'augmente la couverture du lexique et la diversité des cas rencontrés. Le problème de la cohérence devient alors crucial, comment vérifier que la partie du lexique élaborée jusque-là reste cohérente ? Comme nous le montrerons dans l'article, une telle vérification n'est possible que si l'on adopte une formalisation poussée des descriptions permettant à chaque instant de s'assurer de manière automatique de la cohérence générale du lexique.

1. Rappelons que le DEC ne décrit qu'un sous-ensemble du lexique français (environ 1 500 lexies regroupées sous 500 vocables). Les 1 500 lexies définies sont des lexies du français courant appartenant aux différentes parties du discours et correspondant aussi bien à des unités simples qu'à des unités complexes (locutions).

2. cf. le projet lexicographique fondé sur l'exploitation du *Trésor de la Langue Française informatisé* (Barque et Polguère, 2009)

Il existe déjà un certain nombre de projets de formalisation de lexiques sémantiques destinés au TAL. Citons parmi ceux-ci le projet LMF (*Lexical Markup Framework*), qui propose une norme pour l’encodage des informations sémantiques (entre autres) dans ces lexiques (LMF, 2008). Cette norme peut tout à fait être utilisée pour mettre en évidence la structure interne des définitions BDéf, comme le montre (Francopoulo, 2005), et devra l’être lors du développement d’un lexique de type BDéf à large couverture. Notre étude se penche pour sa part davantage sur la mise en œuvre d’un certain nombre de calculs sur ces descriptions et sur les intérêts, en premier lieu métalexicographiques, de pouvoir mettre en œuvre de tels calculs.

La suite de l’article s’organise de la façon suivante : la section 2 présente les descriptions lexicales sur lesquelles se fonde notre étude, en l’occurrence les définitions BDéf et les règles lexicales qui en sont dérivées. La section 3 insiste sur l’intérêt de pousser la formalisation de ces descriptions pour les soumettre à un certain nombre de calculs. La section 4 décrit le formalisme des structures de traits typées dans lequel seront encodées les descriptions BDéf et montre l’adéquation de ce formalisme aux calculs envisagés.

2. Le formalisme BDéf

La lexicologie explicative et combinatoire (Mel’čuk *et al.*, 1995), cadre théorique dans lequel s’est développé le formalisme BDéf, adopte une approche décompositionnelle du sens puisqu’elle propose de décrire le sens des lexies au moyen d’une définition lexicographique. En voici un exemple avec la définition de la lexie GAVER#1³ :

X GAVER#1 Y avec Z : X_{individu} nourrit#1 Y_{animal} avec Z_{aliment}, de force et beaucoup, dans le but d’engraisser#2 le foie#1 de Y.

On remarque que cette définition répond à un certain nombre d’exigences formelles : la structure actancielle de l’unité est définie (on définit « X GAVER#1 Y avec Z » et non simplement « GAVER#1 »), les unités utilisées dans la définition sont désambiguïsées (le verbe NOURRIR est employé dans son sens 1, le verbe ENGRAISSER dans son sens 2...), etc.

Toujours d’un point de vue formel, on observe *a contrario* que les relations entre les différentes composantes de sens de la définition ne sont pas totalement explicites. Au niveau de la structure informationnelle, il n’est pas indiqué explicitement que « X nourrir#1 Y avec Z » constitue le genre prochain de la définition, ni que l’engraissement du foie de l’animal est l’objectif du gavage, etc. Les relations prédicats-

3. La lexicologie explicative et combinatoire adopte par ailleurs une approche discrète du sens : les vocables polysémiques sont des regroupements de lexies distinctes les unes des autres car ayant leur fonctionnement linguistique propre mais liées entre elles par des liens sémantiques plus ou moins évidents. La lexie GAVER#1 est par exemple identifiée comme étant le premier sens du vocable polysémique GAVER. Elle est liée aux autres lexies du vocable, telles que GAVER#2[*La presse est là pour nous gaver de précisions sur le coût des grèves.*].

actants ne sont pas non plus explicitées dans la définition et si un sujet humain est capable de comprendre automatiquement quel est le premier actant d'engraisser, il n'en sera pas de même d'un ordinateur. Par ailleurs, cette forme de définition, très naturelle lorsqu'on est habitué à consulter des dictionnaires classiques, n'est pas forcément la plus appropriée pour ceux qui, voulant modéliser le sens lexical, sont amenés à observer et à comparer des données. Le formalisme BDéf est justement né de la volonté de répondre à cette double attente de calculabilité des données lexicographiques et de réflexion sur l'organisation du sens lexical (Altman et Polguère, 2003). Nous le présentons maintenant.

2.1. Les définitions BDéf

La principale caractéristique du formalisme BDéf est de rendre explicite l'organisation des différentes composantes de sens d'une définition. La définition linéaire du verbe GAVER#1 présentée plus haut en format linéaire sera par exemple encodée sous la forme d'une fiche, telle que dans la figure 1 ci-dessous.

Comme on le voit dans cette figure, une fiche BDéf décrivant le sens d'une lexie donnée se décompose en six champs. Le premier champ indique le nom de la lexie définie, ici GAVER#1. S'ensuit une phrase d'exemple, extraite de corpus et illustrant l'utilisation de la lexie définie. Le champ *forme propositionnelle* introduit les actants de la lexie en indiquant leur organisation syntaxique. Dans notre exemple (figure 1), la lexie GAVER#1 a un premier actant X qui se réalise en tant que sujet, un second actant Y qui se réalise en tant qu'objet direct et un actant Z qui se réalise en tant que complément de moyen. C'est donc la proposition *X gaver#1 Y avec Z* qui sera définie dans cette fiche. L'*étiquette sémantique*, quant à elle, exprime de manière concise le sens général de la lexie définie (Polguère, 2003b), ce que fait l'étiquette *nourrir* dans l'entrée de GAVER#1. Le champ *définition* offre une décomposition du sens de la lexie suivant une forme qui sera expliquée ci-après. Enfin, le *typage des actants* attribue un type à chacun des actants introduits dans la forme propositionnelle. Les trois actants de GAVER#1 sont ainsi typés respectivement *individu*, *animal* et *aliment*.

Considérons maintenant la définition proprement dite. Celle-ci comprend deux grandes parties, la *composante centrale* et les *composantes périphériques*, qui correspondent aux notions classiques de *genus* et *differentiae* héritées d'Aristote. La composante centrale encode le sens général de la lexie (également encodé – de façon souvent plus concise – par l'étiquette sémantique). Les composantes périphériques, quant à elles, ont pour fonction de spécifier l'information contenue dans la composante centrale, distinguant ainsi la lexie définie des autres lexies ayant le même sens général. Par exemple, la composante centrale de la définition de GAVER#1 indique que le sens général de cette lexie est *nourrir* et, pour la distinguer des autres lexies dénotant aussi le fait de nourrir (tel que ALLAITER#1), indique dans le champ des composantes périphériques ce que cette action de nourrir a de particulier (elle est faite *de force*, elle implique que l'on nourrisse *beaucoup* l'animal, etc.).

GAVER#1
<i>Avec cette méthode, il est recommandé de gaver les oies quatre fois par jour.</i>
Forme propositionnelle $X \sim Y$ avec Z
Étiquette sémantique nourrir
DÉFINITION Composante centrale 1 : X nourrir#1 Y avec Z Composantes périphériques /*quantité*/ 2 : *1 beaucoup /*but*/ 3 : *1 _dans le but de_ *4 4 : *1 engraisser#2 foie#1 de Y /*manière*/ 5 : *1 _de force_
Typage des actants X : personne Y : animal Z : aliment

Figure 1. La fiche BDéf de la lexie GAVER#1

La partie « composantes périphériques » est constituée d'un ensemble de *blocs définitionnels* formant des composantes autonomes de la définition. Ces composantes sont autonomes dans la mesure où elles peuvent être retirées du corps de la définition (ou ajoutées) sans remettre en cause la bonne formation de cette dernière. Par exemple, si l'on décide d'enlever le bloc intitulé /*but*/ de la définition de GAVER#1, celle-ci restera bien formée, quoique incomplète. En revanche, il n'est pas possible d'enlever une partie de ce bloc sans compromettre la bonne formation de la définition. Les blocs définitionnels sont introduits par l'indication du *rôle* joué par le bloc vis-à-vis de la composante centrale. Dans la définition de GAVER#1, la composante dans le but d'engraisser le foie de Y exprime par exemple le but de l'action de nourrir de force une volaille.

Outre l'indication d'un rôle, les blocs sont constitués d'une liste de *propositions*, toutes identifiées par un numéro afin de pouvoir y faire référence de n'importe quel endroit de la définition. Le bloc /*but*/ regroupe par exemple les propositions n° 3 et n° 4. Les propositions s'articulent autour d'un prédicat principal⁴ et de ses arguments. Par exemple, dans la proposition n°4, le prédicat engraisser#2 a deux actants (*1 et foie de Y). Un prédicat peut par ailleurs être modifié au sein d'une proposition, typiquement par une négation, ou encore par l'expression d'une temporalité.

Les éléments lexicaux des définitions BDéf sont de deux types : les variables, qui représentent soit des actants (X, Y ...) soit des propositions (*1, *2 ...), et les *mots BDéf*, unités de base significatives de la définition. Par *unités significatives*, nous entendons des unités qui ont un sens pouvant être à son tour décomposé. Les mots BDéf sont des lexies de la langue décrite (nourrir#1, foie#1, etc.). Les locutions sont indiquées par des soulignés extérieurs, comme dans le but de et de force dans la définition de GAVER#1. Les lexies utilisées dans le corps des définitions sont déclarées comme appartenant au *métalexique BDéf*, dont la nomenclature ne sera connue qu'à l'issue de la description du lexique.

2.2. Les éléments structurants

Parmi les éléments d'une définition BDéf qui viennent d'être présentés, deux sont importants pour l'organisation générale du lexique : les étiquettes sémantiques et les rôles joués par les composantes périphériques. Les étiquettes sémantiques proviennent, comme nous l'avons dit, d'une hiérarchie développée en même temps que le lexique (Polguère, 2003b). À chaque étiquette est associée la liste des rôles susceptibles d'entrer dans la composition du sens des lexies qui portent l'étiquette en question. Par exemple, à l'étiquette *animal* est associée la liste des rôles qui entrent dans la définition des lexies étiquetées *animal*, comme illustré ci-dessous :

étiquette : *animal*

rôles associés : {caractéristiques physiques, comportement, sexe}

Cette description indique qu'une lexie étiquetée *animal* peut avoir, dans sa définition, un bloc qui rendra compte des *caractéristiques physiques* de l'*animal* (comme par exemple pour GORILLE#1, « singe d'une grande force physique ») et/ou de son *comportement* (comme TAUPE#1, « animal qui se déplace sous terre ») et/ou du *sexe* de l'*animal* (comme pour BŒUF#1, qui désigne un « mâle castré »). Insistons bien sur le fait qu'il n'est question ici que de proposer au lexicographe un choix contrôlé de types de composantes périphériques en fonction d'une composante centrale donnée, non de répondre à l'épineuse question de savoir quelle composante périphérique choisir lors de l'écriture d'une définition. Le lexicographe sait ainsi qu'il peut utiliser un bloc de type *caractéristiques physiques* dans la définition d'une

4. On parle de prédicat principal parce qu'une proposition est susceptible de contenir un autre prédicat. C'est presque le cas dans la proposition n° 4 avec *foie de X*, qui est un semi-prédicat.

lexie étiquetée `animal` mais rien ne lui indique, pour un animal donné, s'il doit le faire ou non.

Outre une liste de rôles, on associe à une étiquette donnée un ensemble de *patrons définitionnels* , définitions BDéf sous-spécifiées qui permettent de définir des sous-classes de l'étiquette donnée. On aura par exemple une sous-classe d'animaux correspondant aux animaux qui ont une fonction (`LIMIER#1`, `COBAYE#1`, `MULE#1`) et dont la définition sera structurée selon le patron de la figure 2.

<p>DÉFINITION</p> <p>Composante Centrale</p> <p>1 : animal</p> <p>Composantes Périphérique</p> <p>/*fonction*/</p> <p>2 : *1 _être utilisé pour_ *3</p> <p>3 : ...</p>

Figure 2. Un patron définitionnel associé à l'étiquette *animal*

Il s'agit donc d'associer aux étiquettes sémantiques des patrons de définitions, qui, en plus de structurer le lexique et de systématiser son développement, vont nous fournir un mode de représentation des relations lexicales, elles aussi éléments structurants du lexique. Présentons-les maintenant.

2.3. Les règles lexicales

Les règles lexicales dont il sera question ici représentent des relations sémantiques paradigmatiques comme l'hypéronymie, l'antonymie, la synonymie, etc. Ces règles prennent, dans notre étude, la forme de paires de définitions BDéf sous-spécifiées, telles que définies dans (Barque et Nasr, 2005). Des définitions sous-spécifiées sont des définitions dans lesquelles seule l'information pertinente est spécifiée. Nous dirons que, pour qu'une paire de lexies puisse être associée à une règle donnée, il faut que la définition des deux lexies soit compatible avec la partie qui lui correspond dans la règle. Autrement dit, de manière pour le moment informelle, il faut que les informations contenues dans les définitions soient compatibles avec les informations contenues dans la règle.

Prenons, comme premier exemple, la relation d'antonymie. Celle-ci peut être définie de la manière suivante : deux lexies L1 et L2 sont en relations d'antonymie si elles appartiennent à la même partie du discours et si certaines composantes de la définition de L1 sont niées, ou plus généralement « mises en opposition » dans la définition de L2, ou le contraire (Polguère, 2003a). Le fait d'encoder les relations sous forme de paires de définitions sous-spécifiées nous permet de distinguer formelle-

ment plusieurs familles de relation d'antonymie. On peut par exemple distinguer deux sous-types de relations d'antonymie selon la place occupée par la négation dans la composante centrale de la définition : l'antonymie inclusive et l'antonymie intersective⁵. La première est représentée dans la figure 3, au moyen d'une règle indiquant que toute paire de lexies dont l'une inclut l'autre dans sa composante principale modifiée par la négation (*ne pas*) est une instance d'antonymie inclusive. La coréférence est indiquée dans la figure par le soulignement : la lexie qui instancie L1 doit être la même que celle qui l'instancie dans la définition de L2. Par exemple, les verbes RÉUSSIR#1[*Paul a réussi l'examen*] et ÉCHOUER#1[*Paul a échoué à l'examen*] sont en relation d'antonymie inclusive puisque le verbe ÉCHOUER#1 sera défini par '*ne pas réussir#1*'.

Relation : antonymie inclusive	
<u>L1</u>	L2
DÉFINITION	DÉFINITION Composante Centrale 1 : X <i>_ne pas_</i> <u>L1</u> Y

Figure 3. Règle lexicale pour la relation d'antonymie inclusive

La seconde règle d'antonymie, présentée ci-dessous, se fonde non plus sur une inclusion de sens entre L1 et L2 mais sur le partage d'une composante définitionnelle qui est niée dans la définition de L2, comme on peut le voir dans la figure 4. Là encore, la coréférence est indiquée par le soulignement : la lexie qui instancie *pred* doit être la même dans la définition de L1 et dans celle de L2. Les verbes PROVOQUER#1[*Paul a provoqué l'accident*] et EMPÊCHER#1[*Paul a empêché l'accident*] sont par exemple en relation d'antonymie intersective puisque le premier se définit par '*X cause que Y a lieu*' et le second par '*X cause que Y n'a pas lieu*'.

On voit, avec cet exemple de l'antonymie, que le recours à des paires de définitions BDéf sous-spécifiées nous permet de distinguer formellement des sous-types de relations sémantiques, tout en offrant une décomposition (et donc une définition plus fine) de cette relation.

Ce mode de représentation des relations a également l'intérêt de contrôler le pouvoir génératif d'une règle donnée, problème souvent évoqué dans la littérature,

5. Cette typologie de relations d'antonymie s'inspire principalement de (Apresjan, 1992) et de (Milićević, 2003). Notons bien que la relation d'antonymie n'est choisie ici que pour illustrer notre modélisation des relations sémantiques paradigmatiques. Pour une littérature détaillée sur la relation d'antonymie, voir entre autres (Lehrer et Lehrer, 1982 ; Apresjan, 1992 ; Lyons, 1995).

Relation : antonymie intersective	
L1	L2
DÉFINITION Composante Centrale 1 : X causer *2 2 : Y <u>pred</u>	DÉFINITION Composante Centrale 1 : X causer *2 2 : Y <u>_ne pas_ pred</u>

Figure 4. Règle lexicale pour la relation d'antonymie intersective

notamment dans la littérature consacrée aux relations de polysémie (Copestake et Briscoe, 1995 ; Kleiber, 1999). Illustrons maintenant ce second point avec une règle de métonymie. Rappelons que la relation de métonymie est une relation de polysémie dans laquelle le référent de L1 est en relation de contiguïté avec le référent de L2 (Polguère, 2003a). Pour reprendre un exemple classique, le référent de VERRE#1[*Ce verre est en plastique*] est en relation de contiguïté avec le référent de VERRE#2[*Il bu son verre d'une traite*] puisque contenant et contenu se touchent.

Relation : métonymie	
<u>L1</u>	L2
Étiquette sémantique lieu	Étiquette sémantique ensemble d'animaux
DÉFINITION Composante centrale 1 : lieu Composantes périphériques /*fonction*/ 2 : *1 _être destiné à_ *3 3 : *1 abriter <u>ensemble d'animaux</u>	DÉFINITION Composante centrale 1 : <u>ensemble d'animaux</u> Composantes périphériques /*origine*/ 2 : *1 provenir de <u>L1</u>

Figure 5. Règle lexicale pour la relation de métonymie entre un lieu et un ensemble d'animaux

La règle lexicale présentée dans la figure 5 décrit un lien de métonymie entre une lexie dénotant un lieu destiné à abriter un ensemble d'animaux

et une autre lexie (de même forme que la première) dénotant ce même ensemble d'animaux provenant de ce même lieu (la coréférence est indiquée dans la règle par le soulignement). L'existence d'une telle règle nous permet de dire par exemple que le vocable ÉCURIE, qui contient la lexie ÉCURIE#1 dénotant un bâtiment destiné à abriter un ensemble de chevaux contiendra aussi probablement une lexie ÉCURIE#2 dénotant un ensemble de chevaux provenant d'une écurie. En revanche, les lexies qui ne dénotent pas un lieu destiné à abriter un ensemble d'animaux ne sont *a priori* pas concernées par cette alternance. La précision des règles proposées permet donc, pour nombre de cas, de limiter la surgénération.

3. Ce que permet une formalisation poussée des descriptions

La forme graphique des données présentées jusqu'ici n'offre qu'un reflet de la structure sous-jacente des définitions et des règles BDéf (voir figures 1, 3, 4 et 5). Cette forme graphique n'est guère adaptée aux calculs que l'on peut envisager d'effectuer sur de telles définitions. Il nous faut par conséquent un outil qui représente de façon plus explicite et plus uniforme la structure des définitions et des règles qui en sont dérivées. Nous allons présenter, dans cette section, les avantages d'une formalisation poussée des descriptions BDéf. Celle-ci permet en effet de s'assurer de la bonne formation syntaxique et sémantique des descriptions et de créer des patrons définitionnels à partir desquels développer le lexique de manière rapide et cohérente. Elle permet également de développer progressivement données et métadonnées, par allers-retours successifs entre l'écriture des définitions et les outils descriptifs qui leur sont liés (hiérarchie d'étiquettes, lien entre étiquettes et rôles de composantes périphériques, métalexique, patrons définitionnels, règles lexicales).

3.1. Le contrôle de la bonne formation des données

La formalisation nous permet de contrôler automatiquement la bonne formation syntaxique et sémantique des définitions. Présentons maintenant ces deux points.

Nous disons qu'une définition BDéf est *bien formée syntaxiquement* si l'organisation de ses composantes respecte les règles de la grammaire définitionnelle. Ci-dessous figurent quelques-unes de ces règles, énoncées pour le moment indépendamment de tout formalisme :

- une définition BDéf est composée d'un bloc central suivi d'une liste de blocs périphériques ;
- un bloc est constitué d'un rôle suivi d'une liste non vide de propositions ;
- une proposition est constituée d'une lexie prédicative qui peut être modifiée par une autre lexie ;
- etc.

Nous disons qu'une définition BDéf est *bien formée sémantiquement* si elle respecte les trois contraintes suivantes :

1) l'étiquette sémantique de la lexie définie doit figurer dans la hiérarchie des étiquettes sémantiques développée à l'OLST (Polguère, 2003b) ;

2) les rôles de blocs périphériques utilisés dans la définition doivent être compatibles avec l'étiquette sémantique de la lexie définie, c'est-à-dire qu'ils doivent avoir été déclarés comme étant des rôles possibles en regard du type de la lexie définie (cf. section 2.2). Par exemple, à un moment donné de la description du lexique, les lexies de type *animal* seront définissables au moyen des blocs ayant pour rôles */*caractéristiques physiques*/* et */*fonction*/*. Les définitions de lexies *animal* qui utilisent d'autres blocs seront rejetées lors du contrôle sémantique ;

3) les lexies utilisées dans la définition doivent figurer dans le métalexique et respecter la définition qui en est faite, notamment pour ce qui est de leur structure actancielle. Imaginons par exemple que le prédicat *nourrir#1*, utilisé dans la définition de *GAVER#1*, soit décrit dans le métalexique comme ayant deux actants, le premier représentant la personne qui donne à manger et le second celle qui mange (X *nourrir#1* Y). Si le lexicographe construit une proposition BDéf avec un troisième actant Z servant à exprimer ce qui est donné à manger, une erreur se produira car la définition de la lexie *nourrir#1* ne correspond pas à l'utilisation qui en est faite par le lexicographe.

3.2. L'aide à l'encodage des données

Rapellons que, dans notre modèle de lexique sémantique formalisé, les lexies sont organisées autour de patrons définitionnels prenant la forme de définitions sous-spécifiées, patrons eux-mêmes organisés autour d'étiquettes sémantiques provenant d'une même hiérarchie. La précision des informations contiennues dans ces patrons (cf. *supra* sections 2.2 et 2.3) nous permet de développer rapidement de solides ébauches de définitions. Cet avantage, déjà mis en avant dans la littérature consacrée aux extensions de sens (Ostler et Atkins, 1991), est d'autant plus vrai que les patrons définitionnels sont susceptibles d'être très précis. Considérons, pour s'en convaincre, le patron définitionnel pour les lexies dénotant un ensemble d'animaux provenant d'un lieu donné ainsi que deux définitions construites à partir de ce patron. Ils sont présentés tous les trois dans la figure 6.

Ce patron définitionnel permet de produire rapidement, de façon homogène et systématique, la définition des lexies pouvant y être associées, comme *ÉCURIE#2* et *RUCHE#2*. Il n'a fallu faire, pour ces deux définitions, qu'assez peu d'aménagement par rapport au patron qui a servi à leur confection : la composante centrale 'ensemble d'animaux', dans le patron, a été remplacée par 'ensemble de chevaux' dans la définition d'*ÉCURIE#2* et par 'ensemble d'abeilles' dans celle de *RUCHE#2*. La variable 'L1' dans la composante périphérique du patron a été instanciée par *ÉCURIE#1* dans la définition d'*ÉCURIE#2* et par *RUCHE#1* dans celle de *RUCHE#2*.

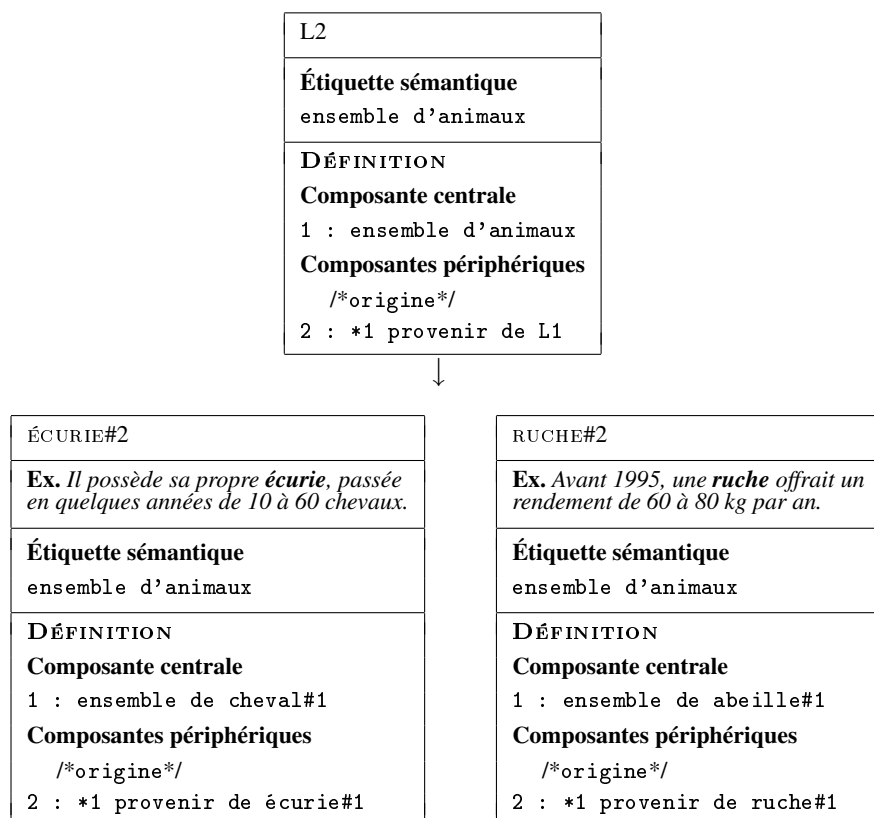


Figure 6. *Écriture de définitions à partir de patrons*

L'utilisation de patrons définitionnels garantit ainsi le développement rapide de définitions homogènes.

3.3. *La description progressive des données lexicales*

Les trois contraintes de bonne formation sémantique d'une définition énoncées plus haut (*cf.* section 3.1) nous permettent d'effectuer des allers-retours entre les spécifications formelles et l'enrichissement de la base. Nous avons déjà signalé plus haut dans l'article que toutes les spécifications ne couvrent pas, dès le début de la description, l'ensemble du lexique. Les étiquettes sémantiques, les liens entre étiquettes sémantiques et rôles, le métalexique, les règles lexicales, tous ces outils descriptifs sont définis au fur et à mesure de la description des unités lexicales.

Concernant tout d'abord le développement de la hiérarchie d'étiquettes sémantiques, nous avons vu plus haut que l'une des contraintes de bonne formation voulait que l'étiquette sémantique de la lexie définie figure dans la hiérarchie. Si ça n'est pas le cas et si aucune des étiquettes existant déjà dans la hiérarchie ne convient pour l'étiquetage de la lexie à définir, alors seulement on crée une nouvelle étiquette. Par exemple, l'étiquette de la lexie *GAVER#1* est *nourrir*. Si la lexie *ALLAITER#1* n'a pas encore été décrite, il est probable que l'étiquette *nourrir* ne figurera pas dans la hiérarchie et il faudra donc la créer.

La liste des rôles de composantes périphériques associés à une étiquette donnée se stabilise elle-aussi au fur et à mesure de la description du lexique. Si par exemple, en écrivant la définition de *TAUPE#1*, le lexicographe décide de faire un bloc ayant pour rôle */*comportement*/* (pour décrire le fait que la taupe se déplace sous terre) mais que ce rôle ne figure pas dans la liste de ceux associés à l'étiquette *animal*, il lui faudra modifier cette liste pour que la définition produite satisfasse la règle de bonne formation sémantique.

La troisième contrainte sémantique, selon laquelle une lexie utilisée dans une définition doit avoir déjà été définie, nous permet, quant à elle, de construire progressivement le métalexique. Si le lexicographe veut utiliser une lexie dans une définition mais que celle-ci ne figure pas dans le métalexique, il faudra dans un premier temps qu'il s'assure que le sens qu'il souhaite exprimer n'est pas exprimable par une lexie du métalexique. L'idée est de contrôler le métalexique et d'éviter le plus possible d'exprimer un sens définitionnel de plusieurs manières, autrement dit d'éviter la paraphrase.

Outre ces contraintes sémantiques, qui nous permettent de définir progressivement les métadonnées (hiérarchie d'étiquettes sémantiques et métalexique), nous disposons d'un mécanisme capable de calculer de manière automatique l'ensemble des paires de lexies compatibles avec une règle donnée (voir la notion d'unification, section 4), permettant ainsi une définition progressive des règles lexicales. Imaginons, par exemple, une règle très simple représentant un lien de métonymie entre une lexie de type *meuble* et une lexie de type *lieu*. Imaginons maintenant qu'à l'issue d'une recherche automatique, le système renvoie la liste des paires de lexies associées à cette règle. On y trouverait, entre autres, les trois paires de lexies suivantes (ici caractérisées sémantiquement par leur seule étiquette sémantique, indiquée entre parenthèses, et illustrées d'un exemple) :

- *BIBLIOTHÈQUE#1* (*lieu*) [*Il entra dans la bibliothèque.*] ~ *BIBLIOTHÈQUE#2* (*meuble*) [*Les clés sont sur l'étagère de la bibliothèque.*]
- *BUREAU#1* (*meuble*) [*Des livres traînaient sur son bureau.*] ~ *BUREAU#2* (*lieu*) [*Il me raconta que les meubles du bureau et de la cuisine s'étaient soudain mis à l'agresser.*]
- *BERCEAU#1* (*meuble*) [*Elle fait la surprise à son mari d'acheter un berceau, ce qu'il trouve prématuré.*] ~ *BERCEAU#2* (*lieu*) [*Pourquoi ne pas se rendre dans le berceau même de cette civilisation ?*]

De ces trois paires de lexies associées à notre règle (trop simple), seules les deux premières sont effectivement des instances de métonymie entre un *meuble* et un *lieu*

qui contient ce meuble. La troisième paire ne devrait pas être associée à cette règle puisque BERCEAU#2 dénote certes un lieu, mais non un lieu qui contient un BERCEAU#1. De fait, le lien entre BERCEAU#1 et BERCEAU#2 n'est pas un lien de métonymie mais un lien de métaphore, puisqu'il y a une analogie (et non contiguïté) entre le petit lit destiné à l'enfant qui vient de naître et l'endroit où apparaît un nouveau phénomène. Il faudra par conséquent que cette règle soit raffinée jusqu'à ce que le système associe un ensemble d'instances de la règle qui soient toutes pertinentes.

4. Les structures de traits typées comme outil de représentation

Les deux sections précédentes ont postulé l'existence d'un certain nombre d'objets (fiches, règles lexicales, hiérarchie d'étiquettes) ainsi que de contraintes sur la structure de ces objets (contraintes syntaxiques, lexicales et sémantiques). Elles ont aussi laissé entrevoir l'existence d'un mécanisme d'application des règles lexicales à des fiches pouvant être interprété dans un sens génératif (génération, à partir d'une fiche et d'une règle, d'une nouvelle fiche) ou prédicatif (est-ce que deux fiches sont reliées par une règle lexicale donnée ?).

Les différents objets et processus mentionnés n'ont pas reçu de description formelle, étape préalable à leur implémentation. C'est l'objectif de cette section. Elle débutera, en 4.1, par la présentation du formalisme des structures de traits typées. Nous décrirons ensuite, en 4.2, comment les structures de traits typées nous permettent de représenter tous les objets manipulés (fiches, règles lexicales, hiérarchie d'étiquettes) ainsi que les contraintes de bonne formation qu'elles doivent vérifier. Nous verrons aussi comment l'opération d'unification permet la vérification des contraintes de bonne formation mais aussi l'application des règles lexicales. Nous terminerons, en 4.3 par un résumé des avantages que nous offre ce mode de représentation.

4.1. Présentation des structures de traits typées

Nous nous contenterons ici d'une brève introduction au formalisme des structures de traits typées. Nous renvoyons le lecteur à (Jurafsky et Martin, 2000) pour une présentation générale, dont nous nous sommes inspirés ici, et à (Carpenter, 1992 ; Copestake, 2002) pour une présentation plus détaillée. Les trois notions les plus importantes du formalisme sont les notions de *structures de traits*, d'*unification* et de *types*. Nous les présentons successivement.

Une structure de traits (ST) est un ensemble de paires composées d'une *attribut* et d'une *valeur* où les attributs sont des symboles non décomposables. Les valeurs peuvent être soit des symboles non décomposables, on parle alors de *valeurs simples*, soit une structure de traits, on a alors affaire à des *valeurs complexes*. Les attributs seront typographiés en petites majuscules (ATTRIBUT).

Les structures de traits sont conventionnellement représentées sous la forme d'une *matrice attribut-valeur* (ou AVM, de l'anglais *attribute value matrix*). Un exemple de

structure de traits a été représenté sous la forme d'un AVM dans la partie gauche de la figure 4.1, il est composé de deux paires ayant respectivement pour attribut LEXIE et ETIQ. L'attribut LEXIE est associé à une valeur complexe, composée à son tour de deux paires (ORTH, *nourrir*) et (NUM, 4) tandis que l'attribut ETIQ est associé à une valeur simple, le symbole *nourrir*. Il est parfois utile d'accéder à un attribut d'une ST complexe à l'aide du *chemin d'attributs* permettant d'y mener. Ce dernier est représenté par une séquence d'attributs séparés par des points. On peut ainsi accéder à l'attribut NUM de la ST de gauche de la figure 4.1 par le chemin LEXIE.NUM.

Deux attributs peuvent partager la même valeur. On parle dans ce cas de *structures réentrantes*. On trouvera dans la partie droite de la figure 4.1 un exemple de structure réentrante. Dans cet exemple, les deux attributs ETIQ (pour « étiquette sémantique ») et CC.TITRE (pour « titre de la composante centrale (de la définition) ») partagent la même valeur, le symbole *nourrir*. Ce partage est indiqué par l'ajout d'un index numérique, représenté dans un carré, au niveau de la valeur partagée. Le partage peut porter sur des valeurs simples comme sur des valeurs complexes.

$$\left[\begin{array}{l} \text{LEXIE} \\ \left[\begin{array}{ll} \text{ORTH} & \text{gaver} \\ \text{NUM} & 1 \\ \text{ETIQ} & \text{nourrir} \end{array} \right] \end{array} \right] \quad \left[\begin{array}{l} \text{LEXIE} \\ \left[\begin{array}{ll} \text{ORTH} & \text{gaver} \\ \text{NUM} & 1 \\ \text{ETIQ} & \boxed{1} \text{nourrir} \end{array} \right] \\ \text{DEF} \\ \left[\begin{array}{ll} \text{CC.TITRE} & \boxed{1} \end{array} \right] \end{array} \right]$$

Figure 7. Deux exemples de structures de traits

Les structures de traits peuvent être combinées à l'aide de l'opération d'*unification* (notée \sqcup) qui consiste à combiner deux structures de traits pour obtenir une structure qui contienne toute l'information des deux structures de traits.

L'opération d'unification se comprend mieux à l'aide de la notion de *subsumption*, qui détermine un ordre partiel sur les structures de traits, de la moins spécifique (celle qui représente le moins d'information) à la plus spécifique (qui en représente le plus). On dit qu'une structure ST_2 est *subsumée* par une autre structure ST_1 (noté $ST_2 \sqsubseteq ST_1$) si ST_2 est plus spécifique que ST_1 , c'est-à-dire si elle contient au moins autant d'information que ST_1 . Le résultat de l'unification de deux structures de traits typées ST_a et ST_b est la structure de traits la plus générale qui est subsumée aussi bien par ST_a que par ST_b . Si cette structure de traits n'existe pas, l'unification échoue. Considérons les deux exemples ci-dessous :

$$\begin{aligned} 1) & \left[\begin{array}{ll} \text{LEXIE.ORTH} & \text{gaver} \end{array} \right] \sqcup \left[\begin{array}{ll} \text{ETIQ} & \text{nourrir} \end{array} \right] = \left[\begin{array}{ll} \text{LEXIE.ORTH} & \text{gaver} \\ \text{ETIQ} & \text{nourrir} \end{array} \right] \\ 2) & \left[\begin{array}{ll} \text{LEXIE.ORTH} & \text{gaver} \\ \text{ETIQ} & \text{nourrir} \end{array} \right] \sqcup \left[\begin{array}{ll} \text{ETIQ} & \text{bâtiment} \end{array} \right] = \perp \end{aligned}$$

Dans le premier exemple, les deux structures s'unifient et donnent comme résultat une nouvelle structure qui réunit les informations des deux premières. Dans le second exemple, l'unification des deux structures échoue car il n'existe pas de ST qui puisse

contenir les informations des deux structures, la valeur du trait ETIQ de la première ST (*nourrir*) n'étant pas compatible avec celle du trait ETIQ de la seconde ST (*bâtiment*).

Les structures de traits peuvent être vues comme un moyen de représenter de l'information que l'on possède sur un objet ou encore comme un moyen de représenter des contraintes sur la structure d'un objet, une sorte de patron ou de spécification en quelque sorte. L'opération d'unification permet alors de vérifier qu'un objet est conforme à une spécification.

Le formalisme des structures de traits tel que décrit ci-dessus a connu de nombreuses évolutions visant à enrichir son pouvoir descriptif. Nous citerons ici trois évolutions utiles pour notre travail, les *attributs à valeur de liste*, les *attributs à valeur d'ensemble*, que nous décrivons rapidement dans le paragraphe suivant et les *structures de traits typées* que nous développons un peu plus en détail dans les paragraphes suivants.

Les attributs à valeur de liste permettent, comme leur nom l'indique, d'associer à un attribut une liste de valeurs. Les listes de valeurs sont notées conventionnellement entre chevrons. Les attributs à valeur d'ensemble, introduits par (Pollard et Moshier, 1990), permettent, on l'aura compris, d'associer à un attribut un ensemble de valeurs, notées entre accolades. L'opération d'unification est enrichie pour prendre en compte ces deux extensions du formalisme.

Dans les structures de traits décrites ci-dessus, il est impossible de contraindre les valeurs que peut prendre un attribut ni les attributs que l'on peut trouver au sein d'une même structure. On n'a donc aucun moyen de distinguer parmi toutes les structures que l'on peut engendrer à partir d'un ensemble d'attributs et de valeurs simples, un sous-ensemble de structures licites. Afin de résoudre ce problème, on ajoute aux structures de traits la notion de *type* pour former les structures de traits typées (STT).

De la même manière que l'on avait distingué des valeurs simples et des valeurs complexes, on distingue les *types simples* et les *types complexes*. Les types simples seront typographiés en minuscules grasses (**type**) et les types complexes en majuscules grasses (**TYPE**).

Les types simples sont simplement des symboles qui remplacent les valeurs simples dans les structures de traits. Les types simples sont organisés en une *hiérarchie de types* qui définit un ordre partiel sur l'ensemble des types simples. Cette hiérarchie permet de définir, pour un type simple, un ou plusieurs sous-types. Elle autorise l'héritage multiple : un type peut avoir plusieurs sur-types. On trouvera, figure 8, un exemple de hiérarchie de types simples. Le type le plus général est noté de manière conventionnelle **top**. Il possède dans notre exemple trois sous-types, le type **string**, le type **nom_actant** et le type **etiq**, qui possède à son tour deux sous-types, etc.

L'opération d'unification est étendue à la notion de type simple. L'unification de deux types simples t_1 et t_2 est le type le plus général qui est plus spécifique que t_1 et que t_2 . Un tel type peut ne pas être défini, auquel cas l'unification échoue. Voici un exemple d'unification mettant en jeu des types simples **nourrir** et **action** (où **nourrir**

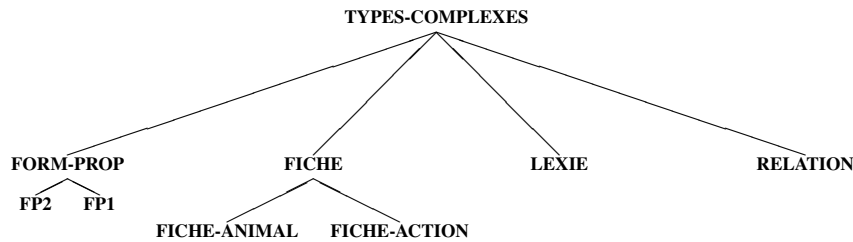


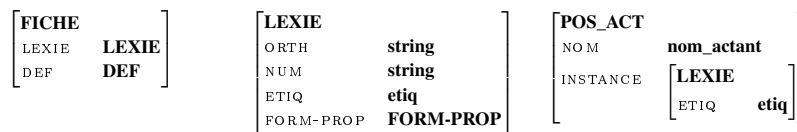
Figure 10. *Un exemple de hiérarchie de types complexes*

L'opération d'unification est étendue aux structures de traits typées en ajoutant comme contrainte, pour que deux structures puissent s'unifier, que leurs types le puisse aussi.

4.2. Les descriptions BDéf en format STT

La présentation du formalisme des structures de traits typées étant faite, il est temps de décrire les différents types qui vont nous permettre d'encoder les fiches et les règles BDéf dans ce formalisme⁶.

Le type **FICHE**, présenté ci-dessous à gauche, est défini par l'attribut **LEXIE**, qui représente la lexie définie dans la fiche, et l'attribut **DEF**, qui représente sa définition. Le type **LEXIE** (au centre) a, quant à lui, quatre attributs : **ORTH** représente la forme orthographique de la lexie, **NUM** correspond au numéro permettant d'identifier la lexie parmi les autres lexies du vocable si celui-ci est polysémique (par exemple, le sens 1 du vocable **GAVER**), l'attribut **ETIQ** indique quelle est l'étiquette sémantique de la lexie et enfin **FORM-PROP** représente sa forme propositionnelle. Le type **POS_ACT** (pour position actantielle), à droite, est défini par deux attributs : l'attribut **NOM**, qui prend comme valeur **X**, **Y**, **Z** ou **W** (un prédicat ayant rarement plus de quatre actants), et l'attribut **INSTANCE** qui prend comme valeur une lexie dont l'étiquette sémantique est spécifiée, typant ainsi la position actantielle.



Le type **FORM-PROP** a quatre sous-types (**FP1**, **FP2**, **FP3** et **FP4**), qui définissent la forme propositionnelle des lexies selon qu'elles ont 1, 2, 3 ou 4 actants. Nous présentons ici trois de ces sous-types :

6. Les questions liées à l'implémentation du processus permettant de passer d'une représentation à l'autre (*i. e.* d'une fiche BDéf à une fiche STT) ne seront pas abordées ici.

$\left[\begin{array}{l} \mathbf{FP1} \\ \text{ACT1} \quad \mathbf{POS_ACT} \end{array} \right]$	$\left[\begin{array}{l} \mathbf{FP2} \\ \text{ACT1} \quad \mathbf{POS_ACT} \\ \text{ACT2} \quad \mathbf{POS_ACT} \end{array} \right]$	$\left[\begin{array}{l} \mathbf{FP3} \\ \text{ACT1} \quad \mathbf{POS_ACT} \\ \text{ACT2} \quad \mathbf{POS_ACT} \\ \text{ACT3} \quad \mathbf{POS_ACT} \end{array} \right]$
---	--	---

Le type simple **etiq**, qui sert de valeur à l'attribut **ETIQ** (*cf.* type **LEXIE**) permet d'intégrer l'ensemble de la hiérarchie des étiquettes sémantiques développée à l'OLST (Polguère, 2003b). Ci-dessous, une branche de cette hiérarchie déjà rencontrée dans nos exemples :

nourrir \sqsubseteq **action** \sqsubseteq **etiq** \sqsubseteq **top**

Le type **DEF**, ci-dessous à gauche, est défini par les attributs **CC** (pour composante centrale) et **CP** (pour composantes périphériques). L'un et l'autre prennent comme valeur des objets de type **BLOC**, défini par l'attribut **TITRE**, qui représente le titre du bloc, et par l'attribut **PROPS**, qui représente l'ensemble des propositions regroupées sous ce bloc. Enfin, le type **PROP** (pour proposition) est défini par les attributs **PRED**, qui représente le prédicat principal de la proposition et qui a pour valeur une lexie et l'attribut **MOD**, qui représente l'éventuel modifieur du prédicat.

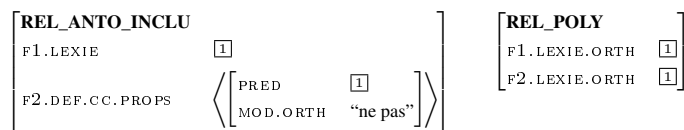
$\left[\begin{array}{l} \mathbf{DEF} \\ \text{CC} \quad \mathbf{BLOC} \\ \text{CP} \quad \{ \text{ENSEMBLE DE } \mathbf{BLOC} \} \end{array} \right]$	$\left[\begin{array}{l} \mathbf{BLOC} \\ \text{TITRE} \quad \mathbf{etiq} \\ \text{PROPS} \quad \langle \text{LISTE DE } \mathbf{PROP} \rangle \end{array} \right]$	$\left[\begin{array}{l} \mathbf{PROP} \\ \text{PRED} \quad \mathbf{LEXIE} \\ \text{MOD} \quad \mathbf{LEXIE} \end{array} \right]$
--	--	--

Nous venons de présenter l'ensemble des types de plus haut niveau⁷ permettant de construire une fiche BDéf dans le formalisme des structures de traits typées. Il nous reste à voir les types permettant de définir les règles lexicales BDéf.

Le type **REL_SEM** (pour relation sémantique) est défini par un attribut **NOM** et deux attributs **F1** et **F2** dont la valeur est de type **FICHE**. Le type **REL_ANTO_INCLU** (pour relation d'antonymie inclusive, *cf. supra*, section 2.3) est un sous-type de **REL_SEM**. Il a ceci de particulier que **L1** doit être inclus dans la définition de **L2** au niveau de la composante centrale (c'est-à-dire, dans une des propositions de ce bloc) et qu'il doit être modifié par la négation (**ne pas**). Le type **REL_POLY** (pour relation de polysémie) hérite également du type **REL_SEM** mais impose que la forme des lexies définies soit de même forme. Cette contrainte est indiquée par un numéro encadré.

$\left[\begin{array}{l} \mathbf{REL_SEM} \\ \text{NOM} \quad \mathbf{string} \\ \text{F1} \quad \mathbf{FICHE} \\ \text{F2} \quad \mathbf{FICHE} \end{array} \right]$
↓

7. Certains de ces types peuvent avoir en effet des sous-types. C'est le cas par exemple de **FICHE**, qui a pour sous-types **FICHE-ACTION**, **FICHE-ENTITÉ**, etc., qui ont eux-mêmes pour sous-types **FICHE-COMMUNICATION**, **FICHE-ANIMAL**, etc.

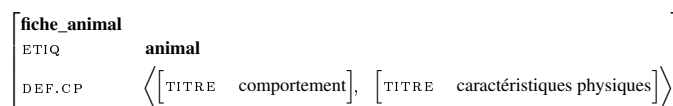


4.3. Intérêts du formalisme

Les intérêts du formalisme des structures de traits typées sont nombreux pour notre étude. Énumérons-les maintenant.

Tout d’abord, les structures de traits permettent, comme leur nom l’indique, de manier des objets structurés. Or, les définitions BDéf et les règles lexicales construites à partir de ces définitions sont des objets structurés : les informations qu’elles contiennent ne sont pas indépendantes les unes des autres mais prennent place au sein d’une structure définitionnelle. La représentation sous forme de structure de traits offre du reste une analyse de la structure définitionnelle qui n’était pas totalement explicite dans la représentation BDéf. La STT ci-dessous (figure 11) montre une partie de la définition BDéf de *GAVER#1*, dans laquelle on voit par exemple que ‘nourrir’ est le prédicat principal de la première proposition de la composante centrale, information qui demeurerait implicite dans la définition BDéf du même verbe (*cf.* section 2.1, figure 1). Or, plus l’information est structurée, plus il est possible d’exprimer des contraintes fortes et plus il est possible, par conséquent, d’effectuer des calculs élaborés.

Les structures de traits nous permettent par ailleurs d’effectuer les contrôles de bonne formation syntaxique et sémantique des définitions (*cf. supra*, section 3.1). Revenons pour commencer sur la contrainte de bonne formation sémantique selon laquelle les étiquettes de blocs d’une définition doivent être compatibles avec l’étiquette sémantique de la lexie : la hiérarchie de types oblige à respecter cette contrainte, sans quoi une erreur se produit lors de l’unification. Nous déclarons en effet des sous-types du type **FICHE** (voir *supra*, section 4.2) qui expriment la dépendance entre l’étiquette sémantique de la lexie définie et les étiquettes de blocs de sa définition. Ci-dessous, à titre d’exemple, le type **FICHE_ANIMAL**, qui hérite du type **FICHE** et qui indique que la définition d’une lexie étiquetée *animal* peut contenir des blocs définitionnels de type **comportement** ou **caractéristiques physiques**. Si on crée une fiche de type **FICHE_ANIMAL** qui contient un bloc de type **fonction** (par exemple, pour la lexie *LIMIER#1*, chien de chasse dont la fonction est de localiser le gibier), une erreur se produira car **fonction** ne correspond pas à un des blocs déclarés. Il faudra alors, dans ce cas de figure, compléter la définition du type **FICHE_ANIMAL**.



Une autre contrainte de bonne formation sémantique veut que les prédicats utilisés dans les propositions correspondent à la spécification qui en est faite dans le lexique.

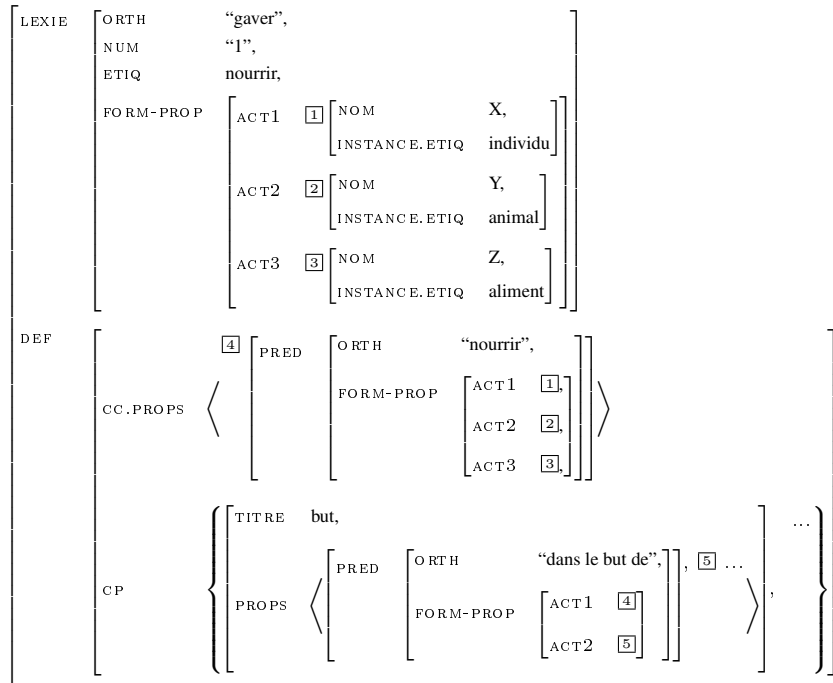


Figure 11. La définition de *gaver#1* au format STT (extrait)

Notre modélisation des données dans le formalisme des structures de traits typées nous permet, là encore, de vérifier cette contrainte. Les informations associées à une lexie tenant lieu de prédicat principal d’une proposition doivent en effet s’unifier avec les informations contenues dans la fiche de cette lexie. Par exemple, la lexie “nourrir#1” utilisée dans la définition de “gaver#1” a trois actants, qui correspondent aux trois actants de “gaver#1” (voir figure 11). Le système ira voir dans la fiche de la lexie “nourrir#1” si ces trois actants s’unifient avec la définition des trois positions actantielles associées à cette lexie.

Enfin, le fait que les structures de traits soient typées nous permet d’établir un ordre sur l’ensemble des structures et de pouvoir ainsi exprimer des règles plus ou moins générales. Prenons l’exemple d’une règle représentant un lien de polysémie entre une lexie étiquetée *animal* et une lexie étiquetée *individu*. Si les structures de traits n’étaient pas typées, nous ne pourrions associer à cette règle que les paires de lexies étiquetées *animal* et *individu*, laissant de côté toutes les paires de lexies étiquetées par des étiquettes plus spécifiques que *animal* (comme par exemple *oiseau*) ou que *individu* (comme par exemple *individu peu intelligent*). Le typage nous permet d’obtenir des ensembles de règles lexicales hiérarchisées, synonyme d’une plus grande organisation des données. À titre d’exemple, voici trois rè-

gles de type `relation_de_polysémie` organisées de la plus spécifique à la moins spécifique.

$$\left[\begin{array}{ll} \text{NOM} & \text{métaphore,} \\ \text{F1.ETIQ} & \text{individu peu intelligent,} \\ \text{F2.ETIQ} & \text{oiseau} \end{array} \right] \sqsubseteq \left[\begin{array}{ll} \text{NOM} & \text{métaphore,} \\ \text{F1.ETIQ} & \text{individu peu intelligent,} \\ \text{F2.ETIQ} & \text{animal} \end{array} \right] \sqsubseteq \left[\begin{array}{ll} \text{NOM} & \text{métaphore,} \\ \text{F1.ETIQ} & \text{individu} \\ \text{F2.ETIQ} & \text{animal} \end{array} \right]$$

Ces règles représentent toutes les trois un lien de métaphore reposant sur une analogie entre une caractéristique associée à un `animal` et une caractéristique associée à un `individu`. Seul varie, entre ces trois règles, le degré de spécificité des informations représentées (en l'occurrence le degré de spécificité de l'étiquette sémantique). Considérons par exemple les trois paires de lexie suivantes :

BÉCASSE

- BÉCASSE1 (`oiseau`)

Ex. *Seules la palombe et la **bécasse** des bois pourront être chassées jusqu'au 10 février.*

- BÉCASSE2 (`individu peu intelligent`)

Ex. *Il faudra m'expliquer pourquoi Navarro et cette **bécasse** de Lescaut cartonnent tant.*

ÂNE

- ÂNE1 (`animal`)

Ex. *Désormais, l'**âne** n'est plus considéré comme une bête de travail.*

- ÂNE2 (`individu peu intelligent`)

Ex. *Me prend pas pour un **âne**, je l'ai vécu, je sais donc de quoi je parle.*

TAUPE

- TAUPE1 (`animal`)

Ex. *Pour se débarrasser d'une **taupe**, le mieux est de l'envoyer chez le voisin.*

- TAUPE2 (`individu (qui a une fonction)`)

Ex. *Il ne savait pas que son collègue était une **taupe** au sein du FBI.*

On voit que les deux lexies du vocable TAUPE ne peuvent être associées qu'à la règle la plus générale, celles du vocable ÂNE peuvent être associées aux deux règles les plus générales et celles de BÉCASSE peuvent l'être au trois. Or, il est intéressant de regrouper ces instances, car elles ont bien en commun de toutes présenter une analogie entre l'homme et l'animal.

En bref, le formalisme des structures de traits typées se révèle tout à fait adéquat pour mettre en œuvre les calculs envisagés dans la section 3 : la vérification de la bonne formation des définitions (qui offre une aide à l'encodage de ces définitions) et le calcul des instances des règles (qui offre une aide à l'encodage de ces règles). Ce sont ces caractéristiques qui nous ont fait adopter les structures de traits typées comme formalisme de représentation. Les définitions sous forme de STT ne sont pas destinées à être manipulées par des lexicographes mais par un ordinateur. Le format de traitement de définitions BDéf par le lexicographe sera quant à lui un format XML (Barque & Polguère 2009).

5. Conclusion

Le travail proposé ici prend sa source dans la lexicologie explicative combinatoire qui propose une décomposition du sens lexical sous forme de réseaux sémantiques (Polguère, 1997). Bien que séduisant, ce mode de description s'est heurté à des difficultés méthodologiques de description ainsi qu'au caractère très abstrait des réseaux sémantiques, qui peut se révéler inadéquat pour le lexicographe. Le développement du format BDéf a permis d'atténuer ces difficultés en proposant un outil descriptif de plus haut niveau permettant de se rapprocher des descriptions lexicographiques plus conventionnelles. Cependant, les principes qui sous-tendent ce format n'ont pas été totalement explicités ni représentés de manière formelle, empêchant de réaliser tout un ensemble de calculs que permettent d'envisager ce format. C'est dans ce contexte que s'inscrit le travail présenté ici.

La formalisation des données sous forme de structures de traits typées nous a permis non seulement de mettre en œuvre les calculs envisagés (vérification automatique de la bonne formation des définitions, indication de la relation lexicale existant entre deux définitions, etc.), mais aussi d'envisager un développement progressif des données (les définitions et les règles lexicales) et des métadonnées (la hiérarchie d'étiquettes sémantiques, les patrons définitionnels et le métalexique), en autorisant la répercussion automatique des nouvelles décisions lexicographiques sur l'ensemble du lexique.

Bien que s'inscrivant dans le cadre de la lexicologie explicative et combinatoire, notre étude pourra servir à l'élaboration d'outils pour l'analyse des définitions de dictionnaires existants à large couverture ainsi qu'à l'élaboration d'outils d'aide à l'encodage, pour une reformulation contrôlée de ces définitions.

6. Bibliographie

- Altman J., Polguère A., « La BDéf : base de définitions dérivée du Dictionnaire Explicatif Combinatoire », *International Conference on Meaning-Text Theory*, Paris, p. 43-54, 2003.
- Apresjan J., *Lexical semantics*, Karoma Publisher, Ann Arbor, 1992.
- Barque L., Nasr A., « Computing semantic relations on structured lexical definitions », *Second International Conference on Meaning-Text Theory*, Moscou, p. 41-49, 2005.
- Barque L., Polguère A., « Structuration et balisage sémantique des définitions du *Trésor de la Langue Française informatisé (TLFi)* », *Fourth International Conference on Meaning-Text Theory*, Montréal, 2009.
- Carpenter R., *The logic of typed feature structures*, Cambridge University Press, 1992.
- Copestake A., *Implementing Typed Feature Structure Grammars*, CSLI publications, 2002.
- Copestake A., Briscoe T., « Semi-productive Polysemy and Sense Extension », *Journal of Semantics*, vol. 1, p. 15-67, 1995.
- Francoπούλο G., *Extended examples of lexicons using LMF (auxiliary working paper for LMF)*, vol. INRIA-Loria, Rapport technique, 2005.

- Jurafsky D., Martin J., *Speech and Language Processing*, Prentice Hall, Upper Saddle River, New Jersey, Etats Unis, 2000.
- Kleiber G., *Problèmes de sémantique. La polysémie en question*, Presses Universitaires du Septentrion, Villeneuve d'Ascq, 1999.
- Lehrer A., Lehrer K., « Antonymy », *Linguistics and philosophy*, vol. 5, p. 483-501, 1982.
- LMF, *Language resource management - Lexical markup framework (LMF)*, ISO/TC 37/SC 4 N453 (N330 Rev.16), 2008.
- Lyons J., *Linguistics Semantics. An introduction*, Cambridge University Press, London, 1995.
- Mel'čuk I., Arbatchewsky-Jumarie N., Lenitsky L., Iordanskaja L., Lessard A., *Dictionnaire explicatif et combinatoire du français contemporain - Recherches lexico-sémantiques I, II, III, IV*, Les Presses de l'Université de Montréal, Montréal, 1984, 1988, 1992, 1999.
- Mel'čuk I., Clas A., Polguère A., *Introduction à la lexicologie explicative et combinatoire*, Duculot, Louvain La Neuve, 1995.
- Milićević J., Étude des aspects sémantiques et syntaxiques de la paraphrase : application à la génération automatique des phrases françaises, PhD thesis, Université de Montréal, 2003.
- Ostler N., Atkins B., « Predictable Meaning Shift : Some Linguistic Properties of Lexical Implication Rules », in J. Pustejovsky, S. Bergler (eds), *Lexical Semantics and Knowledge Representation : First SIGLEX Workshop Proceedings*, Springer-Verlag, Berlin, 1991.
- Polguère A., « Meaning-Text Semantic Networks as a Formal Language », in L. Wanner (ed.), *Recent Trends in Meaning-Text Theory*, John Benjamins, Amsterdam/Philadelphia, p. 1-24, 1997.
- Polguère A., *Lexicologie et sémantique lexicale*, Presses de l'université de Montréal, 2003a.
- Polguère A., « Étiquetage sémantique des lexies dans la base de donnée DiCo », *Traitement Automatique des Langues*, vol. 44, n° 2, p. 39-68, 2003b.
- Pollard C., Moshier M. A., « Unifying Partial Descriptions of Sets », in P. P. Hanson (ed.), *Information, Language, and Cognition*, University of British Columbia Press, Vancouver, p. 285-322, 1990.