Université Sorbonne Paris Cité

Thèse de doctorat
de l'Université Sorbonne Paris Cité
Préparée à l'Université Paris Diderot
École doctorale de Sciences du Langage (ED 132)
Laboratoire de linguistique formelle (CNRS – Université Paris Diderot)

université PARIS DIDEROT PARIS 7

MARIANNE DJEMAA

STRATÉGIE DOMAINE PAR DOMAINE POUR LA CRÉATION D'UN FRAMENET DU FRANÇAIS : ANNOTATIONS EN CORPUS DE CADRES ET RÔLES SÉMANTIQUES

Thèse de doctorat de Linguistique Théorique, Descriptive et Automatique

Dirigée par Marie CANDITO

Présentée et soutenue publiquement à Paris le 14 juin 2017

JURY

Dr.	Marie CANDITO	Université Paris Diderot (Paris 7)	directrice
Pr.	Laurence DANLOS	Université Paris Diderot (Paris 7)	présidente du jury
Pr.	Sylvain KAHANE	Université Paris Ouest (Paris 10)	rapporteur
Pr.	Marie-Claude L'HOMME	Université de Montréal	rapporteure
Pr.	Alexis NASR	Université Aix-Marseille	examinateur

RÉSUMÉ

Dans cette thèse, nous décrivons la création du French FrameNet (FFN), une ressource de type FrameNet pour le français créée à partir du FrameNet de l'anglais (Baker *et al.*, 1998) et de deux corpus arborés : le French Treebank (Abeillé *et al.*, 2003) et le Sequoia Treebank (Candito et Seddah, 2012). La ressource séminale, le FrameNet de l'anglais, constitue un modèle d'annotation sémantique de situations prototypiques et de leurs participants. Elle propose à la fois :

a) un ensemble structuré de situations prototypiques, appelées *cadres*, associées à des caractérisations sémantiques des participants impliqués (les rôles) ;
b) un lexique de déclencheurs, les lexèmes évoquant ces cadres ;
c) un ensemble d'annotations en cadres pour l'anglais.

Pour créer le FFN, nous avons suivi une approche «par domaine notionnel» : nous avons défini quatre «domaines» centrés chacun autour d'une notion (cause, communication langagière, position cognitive ou transaction commerciale), que nous avons travaillé à couvrir exhaustivement à la fois pour la définition des cadres sémantiques, la définition du lexique, et l'annotation en corpus. Cette stratégie permet de garantir une plus grande cohérence dans la structuration en cadres sémantiques, tout en abordant la polysémie au sein d'un domaine et entre les domaines. De plus, nous avons annoté les cadres de nos domaines sur du texte continu, sans sélection d'occurrences : nous préservons ainsi la distribution des caractéristiques lexicales et syntaxiques de l'évocation des cadres dans notre corpus. À l'heure actuelle, le FFN comporte 105 cadres et 873 déclencheurs distincts, qui donnent lieu à 1 109 paires déclencheur-cadre distinctes, c'est-à-dire 1 109 sens. Le corpus annoté compte au total 16 167 annotations de cadres de nos domaines et de leurs rôles.

La thèse commence par resituer le modèle FrameNet dans un contexte théorique plus large. Nous justifions ensuite le choix de nous appuyer sur cette ressource et motivons notre méthodologie en domaines notionnels. Nous explicitons pour le FFN certaines notions définies pour le FrameNet de l'anglais que nous avons jugées trop floues pour être appliquées de manière cohérente. Nous introduisons en particulier des critères plus directement syntaxiques pour la définition du périmètre lexical d'un cadre, ainsi que pour la distinction entre rôles noyaux et non-noyaux.

Nous décrivons ensuite la création du FFN : d'abord, la délimitation de la structure de cadres utilisée pour le FFN, et la création de leur lexique. Nous présentons alors de manière approfondie le domaine notionnel des positions cognitives, qui englobe les cadres portant sur le degré de certitude d'un être doué de conscience sur une proposition. Puis, nous présentons notre méthodologie d'annotation du corpus en cadres et en rôles. À cette occasion, nous passons en revue certains phénomènes linguistiques qu'il nous a fallu traiter pour obtenir une annotation cohérente ; c'est par exemple le cas des constructions à attribut de l'objet.

Enfin, nous présentons des données quantitatives sur le FFN tel qu'il est à ce jour et sur son évaluation. Nous terminons sur des perspectives de travaux d'amélioration et d'exploitation de la ressource créée.

Mots clefs : Traitement automatique des langues, linguistique, FrameNet, French FrameNet, sémantique lexicale, annotation sémantique, rôles sémantiques, corpus, français

ABSTRACT

This thesis describes the creation of the French FrameNet (FFN), a French language FrameNet type resource made using both the Berkeley FrameNet (Baker *et al.*, 1998) and two morphosyntactic treebanks: the French Treebank (Abeillé *et al.*, 2003) and the Sequoia Treebank (Candito et Seddah, 2012). The Berkeley FrameNet allows for semantic annotation of prototypical situations and their participants. It consists of:

a) a structured set of prototypical situations, called *frames*. These frames incorporate semantic characterizations of the situations' participants (*Frame Elements*, or FEs);

b) a lexicon of *lexical units* (LUs) which can evoke those frames;

c) a set of English language frame annotations.

In order to create the FFN, we designed a "domain by domain" methodology: we defined four "domains", each centered on a specific notion (cause, verbal communication, cognitive stance, or commercial transaction). We then sought to obtain full frame and lexical coverage for these domains, and annotated the first 100 corpus occurrences of each LU in our domains. This strategy guarantees a greater consistency in terms of frame structuring than other approaches and is conducive to work on both intra-domain and inter-domains frame polysemy. Our annotating frames on continuous text without selecting particular LU occurrences preserves the natural distribution of lexical and syntactic characteristics of frame-evoking elements in our corpus. At the present time, the FFN includes 105 distinct frames and 873 distinct LUs, which combine into 1,109 LU-frame pairs (i.e. 1,109 senses). 16,167 frame occurrences, as well as their FEs, have been annotated in our corpus.

In this thesis, I first situate the FrameNet model in a larger theoretical background. I then justify our using the Berkeley FrameNet as our resource base and explain why we used a domain-by-domain methodology. I next try to clarify some specific BFN notions that we found too vague to be coherently used to make the FFN. Specifically, I introduce more directly syntactic criteria both for defining a frame's lexical perimeter and for differentiating core FEs from non-core ones.

Then, I describe the FFN creation itself first by delimitating a structure of frames that will be used in the resource and by creating a lexicon for these frames. I then introduce in detail the Cognitive Stances notional domain, which includes frames having to do with a cognizer's degree of certainty about some particular content. Next, I describe our methodology for annotating a corpus with frames and FEs, and analyze our treatment of several specific linguistic phenomena that required additional consideration (such as object complement constructions).

Finally, I give quantified information about the current status of the FFN and its evaluation. I conclude with some perspectives on improving and exploiting the FFN.

Keywords: Natural Language Processing, Linguistics, FrameNet, French FrameNet, Lexical Semantics, Semantic Annotation, Semantic Roles, Corpus, French

REMERCIEMENTS

Merci tout d'abord à ma directrice de thèse, Marie Candito. Merci à elle de m'avoir intégrée au sein du projet ASFALDA et résolument encadrée toutes ces années. Son attention, sa rigueur, son souci du détail, sa disponibilité et ses maintes relectures ont infiniment amélioré ce manuscrit et m'ont obligée à progresser. Merci enfin pour sa patience et sa persévérance dans mes (nombreux) moments de creux.

Je remercie également les membres de mon jury. À Marie-Claude L'Homme et Sylvain Kahane, qui ont accepté d'être rapporteurs de cette thèse, merci d'avoir lu mon travail avec autant d'attention, et merci pour toutes leurs remarques. Je remercie aussi Laurence Danlos et Alexis Nasr d'avoir accepté de faire partie du jury de cette soutenance.

Le French FrameNet est le fruit d'un effort très collectif. Merci du fond du cœur à toutes les personnes qui y ont contribué. Pascal Amsili, Lucie Barque, Farah Benamara, Marie Candito, Gaël de Chalendar, Pauline Haas, Richard Huyghe, Yannick Mathieu, Philippe Muller, Benoît Sagot et Laure Vieu ont consacré beaucoup de temps et de travail à la première partie du projet, et je les en remercie. Un merci particulier à Philippe Muller et Laure Vieu, qui ont investi énormément de temps et d'efforts dans le projet. Merci encore à Marie Candito, porteuse du projet, pour tout le temps et l'énergie qu'elle y a consacré. Merci aussi à Lucie Barque, Pauline Haas et Richard Huyghe pour nos discussions sur les déclencheurs nominaux et les ambiguïtés évènement/participant. Merci enfin à Pascal Amsili pour les nombreuses heures passées ensemble à scruter les cadres de positions cognitives pour tenter d'y discerner une logique. Merci aussi à lui pour sa disponibilité et sa gentillesse.

Un très grand merci aux annotatrices du projet : Vanessa Combet, Noémie Faivre, Virginie Mouilleron, Marjorie Raufast, Anny Soubeille et Émilia Verzeni. Merci particulièrement aux quatre parisiennes, Émilia, Noémie, Virginie et Vanessa, avec lesquelles j'ai passé beaucoup de temps et que j'ai personnellement vu souffrir pour l'avènement du French FrameNet.

Merci également aux anciens membres d'Alpage, en particulier aux doctorantes et doctorants avec qui j'ai partagé entre autres un bureau : Charlotte, Chloé, Corentin, Emmanuel, Enrique, Marion, Maximin, Pierre, Sacha, Timothée et Valérie. En vrac, merci pour les séminaires du jeudi, les cheveux, le support technique, Boby Lapointe, la bonne humeur malgré tout, le soutien moral, les questions existentielles, les supports de cours déjà préparés, la confiance, la politique, le café, la carte de cantine, l'exemple à suivre, les contrepèteries, les croquettes et les astuces contre le hoquet.

Merci aussi aux amies et amis hors fac qui sont pour beaucoup dans ma survie. Celles et ceux du lycée pour leur amitié qui résiste au temps et ce kit de survie à la thèse dont je n'ai pas assez fait usage. Celles et ceux du Paris TPBG pour le soutien moral et la communauté. Merci en particulier à Guilhem, mon voisin de toujours ; Sacha, mon collègue de muscles ; Camille, ma partenaire de *witty banter* ; Valérie, mon maître en toutes choses ; et Simon, mon jumeau maléfique.

Merci enfin à ma famille, qui m'ayant connue avant la thèse a pourtant continué à m'aimer pendant, et à Aurélie qui ne m'a connue qu'en thèse et qui j'espère continuera à m'aimer après. Sans votre soutien indéfectible, je n'aurais sans doute jamais terminé. Merci d'ailleurs d'avoir, dans un moment d'égarement, accepté de relire mon travail pour en extraire le plus de coquilles possible.

TABLE DES MATIÈRES

Résumé i

Abstract iii

Remerciements v

Table des figures x

Liste des tableaux xii

1 Introduction 1

I État de l'art 5

2 Classements sémantiques 7
2.1 Conception des rôles sémantiques . . . 8
2.1.1 Introduction . . . 8
2.1.2 Conception « traditionnelle » des rôles sémantiques . . . 9
2.1.3 Rôles sémantiques généralisés . . . 13
2.1.4 Théories incluant une classification sémantique des prédicats . . . 15
2.2 Ressources . . . 17
2.2.1 Wordnet . . . 17
2.2.2 PropBank . . . 18
2.2.3 VerbNet . . . 20
2.2.4 Tables du lexique-grammaire . . . 21
2.2.5 Les verbes français (LVF) . . . 24
2.2.6 FrameNet . . . 24
2.2.7 Comparaison des ressources présentées . . . 28

3 FrameNet 31
3.1 Introduction . . . 31
3.2 FrameNet : structure, annotations, lexique . . . 33
3.2.1 Réseau de cadres . . . 33
3.2.2 Annotations . . . 42
3.2.3 Lexique . . . 43
3.3 Méthodologie du projet FrameNet . . . 44
3.3.1 Organisation du travail . . . 45
3.3.2 Création et délimitation des cadres . . . 47
3.3.3 Pratiques d'annotation . . . 50
3.4 Projets d'adaptation de FrameNet dans d'autres langues que l'anglais . . . 53

3.4.1 Espagnol : Spanish FrameNet (SFN) . . . 53
3.4.2 Japonais : Japanese FrameNet (JFN) . . . 54
3.4.3 Allemand : le projet SALSA . . . 55
3.4.4 Suédois : Swedish FrameNet++ (SweFN++) . . . 55
3.4.5 Une diversité d'autres projets . . . 56
3.5 Motivation de notre choix d'utiliser FrameNet . . . 58

II Création d'un FrameNet du français 61

4 Le projet ASFALDA 63
4.1 Un projet de création de ressource . . . 63
4.1.1 Un objectif différent de celui du FrameNet de Berkeley . . . 64
4.1.2 Une stratégie en domaines guidée par le corpus . . . 65
4.2 Précision de notions pour le French FrameNet . . . 68
4.2.1 De la nécessité d'introduire explicitement un ancrage syntaxique des remplisseurs de rôles . . . 68
4.2.2 Notions nécessaires à nos éclaircissements . . . 71
4.2.3 Nucléarité et rôles à annoter au niveau du déclencheur . . . 75
4.2.4 Délimitation d'un cadre . . . 78
4.2.5 Point sur les relations entre cadres dans le French FrameNet . . . 83

5 Cadres et lexique 85
5.1 La polysémie . . . 86
5.1.1 Les tests . . . 87
5.1.2 Application à des cas d'école . . . 93
5.2 Travail conjoint cadres et lexique . . . 97
5.2.1 Modélisation d'un domaine . . . 98
5.2.2 Lexique des cadres . . . 102

6 Le domaine des positions cognitives 107
6.1 Lien avec la littérature . . . 108
6.1.1 Modes de connaissance . . . 110
6.1.2 Factivité . . . 112
6.2 Modélisation du domaine en cadres : vue d'ensemble . . . 113
6.3 Positions cognitives : le cœur du domaine . . . 113
6.3.1 Position cognitive sur un contenu . . . 115
6.3.2 Inchoatif de position cognitive . . . 117
6.3.3 Position cognitive sur la probabilité d'un évènement futur . . . 118
6.3.4 Réactivation mentale du passé . . . 118
6.3.5 Influence . . . 120
6.3.6 Accord . . . 121
6.4 Positions cognitives portant sur des actions . . . 122
6.4.1 Influence portant sur une action à effectuer . . . 123
6.4.2 Accord portant sur une action à effectuer . . . 124
6.4.3 Intégration du but au domaine . . . 125
6.5 Positions cognitives de légitimité . . . 126
6.5.1 Cadres statifs de position sur la légitimité . . . 126
6.5.2 Cadres causatifs de position sur la légitimité . . . 127
6.5.3 Cadres portant sur la légitimité d'éventualités passées . . . 128
6.6 Par extension : cadres de jugement . . . 129

6.7 Relations entre les cadres du domaine des positions cognitives 130

7 Annotation 135
7.1 Principes généraux pour l'annotation 136
7.1.1 Annotation des cadres 136
7.1.2 Annotation des rôles et de leurs remplisseurs 139
7.1.3 Constructions à verbes supports 151
7.2 Déroulement du travail d'annotation 153
7.2.1 Corpus utilisés 153
7.2.2 Pré-annotation automatique 154
7.2.3 Annotation 157
7.2.4 Adjudication 160
7.2.5 Post-traitement 160
7.2.6 Documents et outils utilisés tout au long du cycle d'annotation 162
7.3 Points spécifiques 163
7.3.1 Constructions à attribut de l'objet 163
7.3.2 Déclencheurs nominaux 174

III Résultats et synthèse 181

8 Le French FrameNet 183
8.1 Évaluation de la qualité des annotations 183
8.2 Présentation quantitative du corpus obtenu 185
8.2.1 Extraction d'un lexique quantifié 185
8.2.2 Analyse quantitative 186
8.3 Les macro-rôles 190

9 Conclusion 193
9.1 Synthèse 193
9.2 Bilan 195
9.3 Perspectives 195

A Exemples de guides de polysémie 199

B Guide d'annotation 201

C Lexique quantifié extrait des annotations 245

D Hiérarchie des macro-rôles 249

Bibliographie 259

TABLE DES FIGURES

2.1 Exemples de classes de Levin . . . 17
2.2 Extrait d'un synset WordNet et de ses relations à d'autres synsets . . . 18
2.3 Extrait du frame file PropBank du verbe *kill* . . . 20
2.4 Classe VerbNet `murder-42.1` et sous-classe `murder-42.1-1` . . . 22
2.5 extrait de la table 32H du lexique-grammaire . . . 23
2.6 Extrait des entrées du sous-type syntaxique LVF `F1b.1` : *frapper quelqu'un, un animal à mort* . . . 25
2.7 Extrait du cadre KILLING : définition et rôles Core . . . 26
2.8 Extrait du cadre KILLING : relations aux autres cadres et unités lexicales . . . 26
2.9 Extrait de l'entrée lexicale de *kill.v*, unité lexicale du cadre KILLING . . . 27

3.1 Extrait de la hiérarchie de types ontologiques FrameNet . . . 34
3.2 Exemple de relations de généralisation : les transactions commerciales . . . 41
3.3 Exemples de relations de structure évènementielle : le sommeil et l'éveil . . . 41
3.4 Extrait des annotations lexicographiques du déclencheur *kill* (verbe) du cadre KILLING . 43
3.5 Extrait des annotations de plein texte . . . 44
3.6 Entrée lexicale du déclencheur *kill* (verbe) du cadre KILLING : extrait des réalisations possibles des rôles . . . 44
3.7 Entrée lexicale du déclencheur *kill* (verbe) du cadre KILLING : extrait de liste de patrons de valence . . . 45

5.1 Ensemble des cadres examinés pour le domaine des transactions commerciales . . . 99
5.2 Capture d'écran de la définition du cadre COMING_TO_BELIEVE et de ses caractéristiques distinctives . . . 102
5.3 Capture d'écran de l'adjudication du lexique de CONTAINING à l'aide de l'outil Asfalinks 104

6.1 Représentation des relations au sein du domaine des positions cognitives . . . 132
6.2 Suite de la représentation des relations au sein du domaine des positions cognitives . . . 133

7.1 Le rôle est rempli par une phrase : la tête syntaxique et la tête sémantique concordent . 142
7.2 Le rôle est rempli par un NP prédicatif : la tête syntaxique et la tête sémantique concordent 143
7.3 Le rôle est rempli par une infinitive : on sature le remplisseur à l'aide du sujet de l'infinitive 144
7.4 La tête sémantique du remplisseur est un attribut du sujet : on annote toute la phrase à copule . . . 145
7.5 La tête sémantique du remplisseur est un attribut de l'objet : on intègre son sujet dans le remplisseur . . . 146
7.6 La tête sémantique du remplisseur est un nom : on annote tout le NP sans marquage spécifique. . . . 146
7.7 La tête sémantique du remplisseur est un adjectif épithète : on annote tout le NP en ajoutant un marquage spécial sur l'adjectif . . . 146

7.8 Représentation syntaxique de la phrase *Ils ont été reçus à la boulangerie Leroy pour visiter le fournil [...]* . . . 150
7.9 Représentation syntaxique de la phrase *Ces affaires se distinguent des autres parce qu'elles ne sont pas offertes en souscription publique.* . . . 151
7.10 Astuce pour annoter les phrases précédentes . . . 151
7.11 Astuce pour annoter les phrases avec incise. . . . 152
7.12 Cas d'annotation d'une phrase avec incise ne permettant pas d'utiliser le raccourci vu plus haut. . . . 152
7.13 Pré-annotation d'une phrase où un déclencheur potentiel composé et un de ses composants sont proposés pour l'annotation. . . . 157
7.14 Sélection du cadre déclenché . . . 158
7.15 Rôles noyaux disponibles pour l'annotation . . . 159
7.16 Annotation des remplisseurs des rôles . . . 159
7.17 Annotation d'un rôle rempli par un élément anaphorique . . . 160
7.18 Adjudication du travail de deux annotatrices . . . 161
7.19 Représentation syntaxique de la phrase *[...] les soeurs AFFIRMÈRENT avoir été ensorcelées par le curé Urbain Grandier.* . . . 161
7.20 Représentation syntaxique de la phrase *Je **TROUVE** scandaleux qu'il n'ait rien fait pour attirer l'attention des dirigeants de la chaîne [...]* . . . 162

LISTE DES TABLEAUX

2.1 Ensemble de rôles sémantiques largement acceptés . . . 10
2.2 Propriétés des proto-rôles de Dowty (1991) . . . 14
2.3 Valence sémantique et syntaxique (voix active) des verbes du cadre TRANSACTION COMMERCIALE . . . 16
2.4 Comparaison des différentes ressources présentées . . . 29

3.1 Répartition des rôles non-noyaux les plus nombreux . . . 37
3.2 Répartition des instances de cadres annotés par type d'annotation . . . 42
3.3 Comparaison des différents projets de type FrameNet présentés . . . 56

4.1 Exemples de dépendants sémantiquement obligatoires, sous-catégorisés, les deux à la fois, ou aucun des deux. . . . 75

5.1 Application des tests de polysémie . . . 93
5.2 Nombre de cadres dans la structure du French FrameNet, répartis par domaine notionnel, avec comptes de nouveaux cadres et de cadres modifiés par rapport au FrameNet de Berkeley . . . 101

6.1 Vue d'ensemble des cadres lexicaux du domaine des Positions cognitives . . . 114

8.1 Accord inter-annotateurs pour toutes les occurrences de déclencheurs annotées indépendamment par deux annotateurs . . . 185
8.2 Taille du lexique et volume d'annotations par catégorie morpho-syntaxique . . . 187
8.3 Taille du lexique et volume d'annotations par catégorie morpho-syntaxique, ventilée par proportion de *Other_sense* dans les annotations . . . 188
8.4 Couverture du French FrameNet : découpage par domaine et par catégorie . . . 189
8.5 Couverture et polysémie de la ressource . . . 189
8.6 Polysémie des lemmacats du French FrameNet : répartition des nombres de sens . . . 189

CHAPITRE 1

INTRODUCTION

Dans un énoncé décrivant une situation, certains éléments peuvent référer aux participants de ladite situation. Ces éléments, qui entretiennent une relation avec le prédicat dénotant la situation, peuvent être typés à l'aide de ce qu'on appelle des *rôles sémantiques*. Le repérage des participants à l'aide de ces rôles sémantiques présente deux grands intérêts. D'abord, il permet de rendre compte de la similarité sémantique entre les phrases présentées en (1) — on parle pour ce type de cas d'*alternance syntaxique* — en mettant en évidence la similarité du rôle de leurs participants.

(1) a. Sam a cassé le miroir avec ce marteau.
b. Le miroir a été cassé par Sam.
c. Ce marteau a cassé le miroir.
d. Le miroir a cassé.

Dans une situation qui peut être totalement ou partiellement décrite par toutes ces phrases, *le miroir* est l'objet qui était auparavant d'une pièce et est à présent en plusieurs morceaux (traditionnellement le patient), *Sam* est la personne volontairement responsable de cet état de fait (l'agent) et *ce marteau* est l'instrument utilisé par Sam pour casser le miroir. Assigner la même étiquette de rôle à chacun de ces éléments dans les phrases de (1) dans lesquelles il apparaît permet de donner une représentation unifiée de *casser* comme un prédicat pouvant prendre en arguments un agent, un patient et un instrument. Ainsi on peut rendre compte de l'alternance en (1).

Le deuxième intérêt de l'utilisation des rôles sémantiques pour étiqueter les participants à une situation est qu'elle permet de construire une représentation sémantique des textes. Pouvoir repérer *qui a cassé quoi avec quel instrument* peut profiter à plusieurs tâches de traitement automatique des langues, comme l'extraction d'informations et les systèmes de questions-réponses. Pradet (2015, p. 6) recense un grand nombre d'applications pour lesquelles l'annotation en rôles sémantiques a pu être utilisée, comme la fouille d'opinion et la traduction automatique.

De nombreuses théories tentent de rendre compte de ces rôles sémantiques, et du lien qu'ils entretiennent avec les arguments syntaxiques des prédicats. Une de ces approches, la sémantique des cadres (Fillmore, 1982), considère que la compréhension du sens des mots nécessite obligatoirement l'appel à des connaissances d'arrière-plan, structurées en *cadres* (*frames*). Un cadre décrit une situation prototypique pouvant être évoquée par un ensemble de mots. Ainsi, l'emploi du verbe *vendre* nécessite la connaissance de la notion de transaction commerciale. Les *éléments du cadre* (*frame elements*) nomment les participants de la situation prototypique (pour une transaction commerciale l'acheteur, le vendeur, l'argent et les biens échangés) et décrivent les relations qu'ils entretiennent entre eux. Ces éléments de cadre correspondent à des rôles sémantiques valables pour un groupe de lexèmes, les *unités lexicales* (*lexical units*) pouvant évoquer le cadre. Pour la suite du travail, nous appellerons ces unités lexicales *déclencheurs*, et les éléments de cadres *rôles*

du cadre. La sémantique des cadres a été implémentée par le projet FrameNet (Baker *et al.*, 1998), développé à l'ICSI de l'UC Berkeley. On peut voir cette ressource comme possédant trois volets :

a) un ensemble structuré de cadres — situations prototypiques associées à des caractérisations sémantiques des participants et propriétés y intervenant — liés entre eux par des relations (comme l'héritage et la causalité) ;
b) un lexique de déclencheurs, les lexèmes évoquant ces cadres ; et
c) un ensemble d'annotations en cadres pour l'anglais.

FrameNet permet à la fois de définir des regroupements sémantiques de lexèmes, grâce à son lexique de déclencheurs, et de typer les arguments desdits lexèmes en contexte, grâce aux rôles de leurs cadres.

La présente thèse s'insère dans le cadre du projet ASFALDA — un projet financé par l'ANR (ANR-12-CORD-023) lancé en septembre 2012 et terminé en juin 2016 — qui a pour objectif général de développer un corpus du français annoté sémantiquement et un analyseur sémantique prédisant le même type d'annotations. Pour ce faire, l'équipe du projet a choisi de s'appuyer sur des ressources préexistantes, à savoir FrameNet et deux corpus arborés : le French Treebank (Abeillé *et al.*, 2003) et le Sequoia Treebank (Candito et Seddah, 2012). Cette thèse s'attache à décrire le travail de création du French FrameNet, ressource en français adaptée du FrameNet de Berkeley. Nous avons tiré parti de la structure FrameNet, qui s'était déjà avérée largement transposable à d'autres langues (Boas, 2009a), pour construire cette ressource. Le French FrameNet est composé d'un sous-ensemble de cadres, d'un lexique indiquant les lexèmes français pouvant évoquer les cadres de ce sous-ensemble, et de plus de 16 000 occurrences de cadres annotées (comprenant le déclencheur du cadre et les éléments qui remplissent les rôles de ce cadre) sur les corpus arborés cités ci-dessus.

La spécificité du French FrameNet, par rapport aux autres projets d'adaptation de FrameNet à d'autres langues que l'anglais, est sa méthodologie de travail en domaines notionnels et guidée par le corpus. Originellement [1], le projet FrameNet de Berkeley a été créé suivant une méthodologie dite *cadre à cadre*. Il s'agit de concevoir des cadres en identifiant pour eux des ensembles de déclencheurs, puis de fournir des exemples annotés des différentes réalisations syntaxiques des rôles de ces cadres, et ce pour tous leurs déclencheurs. Cette méthodologie permet en théorie une couverture lexicale totale à l'intérieur des cadres : en créant chaque cadre, on tâche d'ajouter tous les déclencheurs susceptibles de l'évoquer. Elle permet aussi, toujours en théorie, une couverture totale des possibles variations syntaxiques, à la fois pour un couple déclencheur-cadre et pour l'ensemble des déclencheurs d'un cadre. En revanche, comme les phrases annotées ont été choisies pour leur exemplarité, cette approche induit une mauvaise représentativité à la fois de la distribution des sens en corpus et de celle des constructions syntaxiques pour un sens donné.

À l'inverse, le projet SALSA de création d'un FrameNet de l'allemand (Burchardt *et al.*, 2006b) a utilisé une méthodologie *lemme à lemme* : pour un ensemble de verbes et de déverbaux choisis dans différentes plages de fréquence en corpus, il s'agit d'annoter toutes ses occurrences en cadre et en rôles dans le corpus. Cette méthodologie fournit une couverture parfaite de la polysémie des lemmes annotés dans le corpus, avec une représentativité parfaite de la distribution des sens et des constructions syntaxiques des lemmes annotés (toujours pour le corpus). En revanche, à moins que tous les lemmes du corpus aient été annotés, une telle approche ne peut prétendre à une bonne couverture lexicale au sein des cadres.

Étant donné nos ressources et le temps imparti, une annotation de plein texte comme celle pratiquée par le projet FrameNet dans sa seconde phase (cf. note 1) n'était pas envisageable : la trop grande diversité de cadres aurait conduit à des données *in fine* très parcellaires. Nous venons

1. On distingue dans FrameNet deux ensembles d'annotations nés de deux procédures distinctes. Objectif originel du projet, l'annotation lexicographique, pratiquée sur un choix de phrases piochées dans le British National Corpus pour servir d'exemples aux couples déclencheur-cadre, a pour but de décrire exhaustivement la diversité des réalisations syntaxiques des rôles. Par opposition, les annotations de plein texte (*full-text annotation*) consistent en un ensemble de textes continus où toutes les occurrences de paires déclencheur-cadre ont été annotées.

de le voir, une approche lemme à lemme est attractive pour une application de type désambiguïsation lexicale, mais suppose également de couvrir seulement partiellement un grand nombre de cadres. Une approche cadre à cadre ne présente pas cet inconvénient mais induit une mauvaise représentativité des annotations. Nous avons opté pour une approche hybride où l'on considère des *domaines notionnels*, définis chacun comme un ensemble de cadres recouvrant un certain concept. Nous visons l'exhaustivité non seulement du lexique d'un cadre, mais du lexique de tous les cadres d'un même domaine notionnel. D'autre part, en annotant sur du texte continu, sans sélection d'occurrences, nous préservons la distribution des caractéristiques lexicales et syntaxiques de l'évocation des cadres en corpus. Le premier point oblige à travailler sur des cadres proches (les cadres appartenant à un même domaine) et à en dégager les caractéristiques distinctives. Le deuxième point garantit que les données annotées respectent la distribution naturelle des phénomènes linguistiques (dans le corpus visé). Cela les rend utilisables comme données d'apprentissage pour un analyseur sémantique — restreint cela dit aux domaines notionnels couverts.

La thèse commence par resituer FrameNet dans un contexte théorique plus large. Nous justifions ensuite le choix de cette ressource comme base pour le projet et expliquons le choix de notre méthodologie en domaines. Nous explicitons pour le French FrameNet certaines notions définies par FrameNet que nous avons jugées trop floues pour être appliquées de manière cohérente. Puis, nous décrivons la création du French FrameNet : la délimitation de la sous-structure de cadres utilisée pour le French FrameNet, et la création de leur lexique, ainsi que le travail d'annotation du corpus en cadres et en rôles. Nous présentons ensuite de manière approfondie le domaine notionnel des positions cognitives, qui englobe les cadres portant sur le degré de certitude d'un être doué de conscience sur une proposition. Nous analysons aussi le cas de phénomènes linguistiques pour lesquels fournir un traitement cohérent nous a demandé une certaine réflexion. Enfin, nous présentons des données quantitatives sur le French FrameNet tel qu'il est à ce jour et sur son évaluation. Nous terminons sur des perspectives de travaux d'amélioration et d'exploitation de la ressource créée.

Organisation de la thèse

Plus précisément, la thèse est structurée en neuf chapitres (dont ce premier chapitre introductif) regroupés en trois parties. La première partie présente l'état de l'art dans lequel s'inscrit ce travail : la ressource FrameNet, ainsi que les cadres théoriques dans lesquels et par rapport auxquels elle s'inscrit. Elle est composée de deux chapitres :

— le chapitre 2, qui s'attache à placer FrameNet dans un cadre plus général. La ressource s'appuyant sur la sémantique des cadres de Fillmore (1982), une des nombreuses théories s'intéressant aux rôles sémantiques, la section 2.1 tente de définir la notion de rôles sémantiques et explore rapidement plusieurs conceptions typiques de ces rôles. Dans la section 2.2, nous nous attachons à présenter des ressources lexicales qui nous semblent proches de FrameNet par leur manière de généraliser sur les lexèmes ou sur les rôles sémantiques ;

— le chapitre 3, qui présente le modèle FrameNet de manière plus détaillée. Nous commençons par y décrire la structure du FrameNet de Berkeley, sur lequel s'appuie notre French FrameNet, et explorer la méthodologie du projet originel — ce qui nous permettra plus tard de pointer ses différences avec la nôtre. Nous passons ensuite en revue une variété de projets d'adaptation de FrameNet à d'autres langues que l'anglais, ainsi que d'autres types de projets s'appuyant sur le modèle FrameNet. Enfin, nous utilisons les points déjà explorés pour mieux motiver notre choix d'appuyer notre ressource sur FrameNet plutôt que sur un autre modèle.

La deuxième partie présente la création du French FrameNet : nous commençons par décrire le cadre de cette création (le projet ANR ASFALDA), puis nous redéfinissons certaines notions FrameNet qui se sont avérées trop floues. Nous décrivons ensuite les grandes étapes de la construction du French FrameNet. Cette partie se découpe en quatre chapitres :

— le chapitre 4, qui part de la présentation rapide du projet ASFALSDA dans lequel s'insère ce travail. Nous y présentons notre stratégie en domaines, puis cherchons à préciser pour le French FrameNet certaines notions définies par FrameNet, principalement en tâchant de les ancrer dans la syntaxe.

— le chapitre 5, qui commence par une première partie destinée à expliciter la notion de polysémie, à laquelle nous avons souvent fait appel. Une deuxième partie présente ensuite les étapes concomitantes de modélisation des domaines et du lexique des cadres ;

— le chapitre 6, dans lequel nous nous intéressons au domaine notionnel des positions cognitives — domaine sur lequel nous avons particulièrement travaillé, qui englobe les cadres portant sur le degré de certitude d'un être doué de conscience sur une proposition. Nous décrivons ce domaine, tâchons de le rattacher à des notions présentes dans la littérature, et l'utilisons comme exemple d'application des méthodes décrites au chapitre précédent ;

— le chapitre 7, qui traite de la phase d'annotation du French FrameNet. Ses deux premières sections s'intéressent à notre méthodologie générale d'annotation : nous présentons les principes généraux qui nous ont guidés lors de cette phase, ainsi que le déroulement précis du travail d'annotation. Une troisième section vient présenter deux points spécifiques, pour lesquels nous fixer une ligne de conduite a demandé une réflexion particulière : d'abord, les constructions à attribut de l'objet, qui fournissent un cas intéressant de divergence syntaxe sémantique ; puis les déclencheurs nominaux, qui peuvent présenter plusieurs types de lectures nécessitant des traitements différents.

La troisième partie s'intéresse à la ressource qui résulte du travail déjà décrit, et à la synthèse de ce travail. Elle se compose de deux chapitres :

— le chapitre 8, qui se penche sur le French FrameNet lui-même. La première section propose une évaluation de la tâche d'annotation, la suivante une présentation quantitative de la ressource, et la troisième section présente les macro-rôles que nous avons ajoutés à la ressource pour permettre de mieux généraliser sur les rôles ;

— et enfin, le chapitre 9, qui conclut cette thèse en récapitulant l'ensemble du travail effectué pour la création du French FrameNet. Nous y proposons ensuite des perspectives de travail futur pour l'amélioration et l'exploitation de la ressource créée.

PREMIÈRE PARTIE

ÉTAT DE L'ART

CHAPITRE 2

CLASSEMENTS SÉMANTIQUES : CLASSES DE PRÉDICATS ET RÔLES SÉMANTIQUES

Sommaire

2.1 Conception des rôles sémantiques **8**
2.1.1 Introduction 8
2.1.2 Conception « traditionnelle » des rôles sémantiques 9
2.1.2.1 Points communs des approches « traditionnelles » 10
2.1.2.2 Limites 11
2.1.3 Rôles sémantiques généralisés 13
2.1.4 Théories incluant une classification sémantique des prédicats 15
2.1.4.1 Sémantique des cadres 15
2.1.4.2 Classes de Levin 16
2.2 Ressources **17**
2.2.1 Wordnet 17
2.2.2 PropBank 18
2.2.3 VerbNet 20
2.2.4 Tables du lexique-grammaire 21
2.2.5 Les verbes français (LVF) 24
2.2.6 FrameNet 24
2.2.7 Comparaison des ressources présentées 28

Ce travail de thèse portant sur la création d'un FrameNet du français, nous cherchons dans ce chapitre à placer FrameNet dans un cadre plus général. La ressource s'appuie sur la sémantique des cadres de Fillmore, une des nombreuses théories s'intéressant aux rôles sémantiques : nous allons donc d'abord dans la section 2.1 définir la notion de rôles sémantiques et explorer rapidement plusieurs conceptions typiques de ces rôles. Dans la section 2.2, nous nous attacherons à présenter des ressources lexicales qui nous semblent proches de FrameNet par leur manière de généraliser sur les lexèmes ou sur les rôles sémantiques.

2.1 Conception des rôles sémantiques

2.1.1 Introduction

Déjà aux alentours de -500 (± 2 siècles), dans sa grammaire du sanskrit, Pāṇini tentait d'expliquer la relation entre marquages morpho-syntaxiques de type **cas** et régularités sémantiques. Pour ce faire, il a défini les rôles kāraka qui permettent de classer les participants d'un évènement, et dont l'assignation aux arguments d'un verbe dépend de la classe à laquelle ce verbe appartient [1]. Ces rôles kāraka sont les lointains ancêtres des **rôles sémantiques** (ou encore **cas profonds**, **relations thématiques**, **rôles thématiques**, **théta-rôles**...), redécouverts par Gruber (1965) et Fillmore (1968). Ces derniers ont en effet ouvert la discussion qui continue encore aujourd'hui sur les arguments en faveur ou défaveur de l'existence de rôles sémantiques, ainsi que sur les manières de décrire ces rôles.

Que sont ces rôles sémantiques, et à quoi servent-ils ? Nous tentons dans cette introduction de répondre à ces questions de manière assez générale, afin de n'entrer en contradiction avec aucune des multiples théories.

Considérons les exemples en (2).

(2) a. Alex a cassé la vitre.
b. La vitre a été cassée par Alex.
c. Alex a cassé la vitre avec un marteau.
d. Un marteau a cassé la vitre.
e. La vitre a cassé.

(2a) et (2b) sont de même sens, et présentent une forte similarité sémantique avec les autres exemples. Cette similarité peut être exprimée au moyen de l'ajout à ces phrases d'une étiquette informant sur la relation entre l'action prédiquée par le verbe et chaque constituant participant à cette action. Ce sont ces étiquettes qu'on appelle **rôles sémantiques**. Si on annote les phrases (2) en rôles sémantiques, on obtient les exemples en (3).

(3) a. Alex a cassé la vitre.
CASSEUR CASSÉ
b. La vitre a été cassée par Alex.
CASSÉ CASSEUR
c. Alex a cassé la vitre avec un marteau.
CASSEUR CASSÉ INSTRUMENT
d. Un marteau a cassé la vitre.
INSTRUMENT CASSÉ
e. La vitre a cassé.
CASSÉ

L'ajout de ces étiquettes permet donc de donner une représentation unifiée du sens de *casser* comme un prédicat à (au moins) trois arguments : le CASSEUR, le CASSÉ et l'INSTRUMENT, valable pour toutes ces phrases. Typer la relation entre évènement et participants est utile en tant que tel : disposer de cette information facilite la construction de représentations sémantiques des textes, et savoir repérer dans un texte *qui a fait quoi à qui*, ici *qui a cassé quoi avec quel instrument* peut profiter à certaines tâches de traitement automatique des langues, par exemple l'extraction d'information. Pradet (2015, p. 6) recense un grand nombre d'applications pour lesquelles l'annotation en rôles sémantiques a pu être utilisée.

1. Pāṇini propose pour les verbes du sanskrit dix classes définies à partir de leur forme au présent.

Nous venons d'évoquer l'intérêt des rôles sémantiques comme outils de généralisation sémantique, mais historiquement, les rôles sémantiques sont une proposition de généralisation sur la syntaxe au vu de régularités d'alternances syntaxiques comme celles présentées en (2). À l'examen des phrases en (3), on observe aisément que des constituants de fonctions différentes peuvent remplir un même rôle, par exemple le CASSÉ est sujet en (3b) mais objet en (3a). De même, des constituants occupant la même fonction peuvent remplir des rôles différents : le sujet de (3a) est CASSEUR tandis que celui de (3b) est CASSÉ, et celui de (3c) INSTRUMENT. Il existe cependant de très fortes régularités dans la correspondance entre arguments sémantiques et fonctions grammaticales : le V-EUR d'un verbe transitif décrivant une action V à la voix active est souvent son sujet, et le V-É son objet. Les diverses théories des rôles sémantiques tentent toutes (avec plus ou moins de succès) de rendre compte de ces régularités avec ce qu'on appelle des *théories du linking*[2]. Une théorie du linking doit rendre compte des régularités observées dans la correspondance entre arguments sémantiques et fonctions grammaticales ; par exemple étant donnés les arguments sémantiques d'un verbe, elle doit pouvoir rendre compte de quelle fonction grammaticale est occupée par chaque argument sémantique. Le linking permet donc de mettre forme et sens en correspondance.

À partir de rôles comme ceux-ci, c'est-à-dire entièrement dépendants du sens du verbe de la phrase, on peut généraliser pour obtenir des rôles moins spécifiques. Un CASSEUR et un MANGEUR ont pour points communs d'être des entités animées, volitionnelles et à l'initiative de l'évènement décrit par leurs verbes respectifs. Elles peuvent donc être regroupées en un rôle plus général : l'AGENT. Ce passage à des rôles de granularité plus large rend la tâche d'établir des règles de linking entre rôles et fonctions, ou même des règles d'inférence basées sur les rôles dans le cadre d'autres tâches, plus aisée.

Nous avons à présent introduit les rôles sémantiques et établi leur intérêt en deux axes : ils permettent à la fois de typer la relation sémantique qu'un prédicat entretient avec chacun de ses arguments et d'exprimer la correspondance probable de ces arguments sémantiques avec les constituants sous-catégorisés par le prédicat. Nous allons maintenant passer en revue quelques théories des rôles sémantiques qui nous paraissent représentatives. Nous étudierons d'abord les approches « traditionnelles », qui ont ouvert la discussion sur les rôles sémantiques, avant de nous intéresser à d'autres approches, regroupées par degré de granularité des rôles relatif à celui des approches traditionnelles.

2.1.2 Conception « traditionnelle » des rôles sémantiques

Nous évoquons ici sous le nom de « traditionnelles » l'ensemble des premières théories développées parallèlement à partir de la fin des années 1960. Ces approches s'attachent à décrire le sens d'un prédicat par une liste de rôles : des étiquettes identifiant le rôle que chacun des arguments du verbe joue dans l'évènement dénoté par ce prédicat. L'ensemble de ces étiquettes, ou rôles sémantiques, est défini à l'avance. Ces « listes de rôles sémantiques » (terminologie Levin et Rappaport Hovav (2005)) sont aussi souvent appelées grilles thématiques.

On trouve dans la littérature de nombreuses propositions d'ensembles de rôles sémantiques conçues pour permettre de rendre compte des alternances syntaxiques et caractériser les relations entre tout verbe et ses participants. Gruber (1965) et Fillmore (1968) en sont les exemples les plus anciens.

Gruber (1965) critique la grammaire transformationnelle de Chomsky, qui ne permet de rendre compte d'alternances comme celles présentées en (2) qu'en postulant des règles spécifiques à chaque verbe, ce qui oblige à créer un grand nombre de règles applicables à faible pouvoir prédictif. Fillmore formule le même type de critiques et propose des **cas profonds** pour caractériser les arguments des verbes et ainsi pouvoir généraliser sur les variations de surface.

2. Butt (2006) attribue la première utilisation du terme à Carter (1976) et sa reformulation et popularisation à Ostler (1979).

Nous n'examinerons pas plus avant les propriétés spécifiques des propositions traditionnelles d'ensembles de rôles émanant de la littérature mais étudierons plutôt leurs points communs, avant de nous intéresser à leurs limites qui expliquent l'émergence d'autres conceptions des rôles que nous présenterons par la suite.

Les auteurs utilisent des termes différents (*cas profonds, relations thématiques, rôles thématiques, thêta-rôles*...) pour référer à la notion de rôles sémantiques. À moins de vouloir évoquer une spécificité d'une théorie particulière, c'est ce dernier terme que nous continuerons à utiliser ci-après.

2.1.2.1 Points communs des approches « traditionnelles »

Stalmaszczyk (1996) repère quatre points communs à ces approches dites traditionnelles, qui considèrent globalement que :

— les rôles sémantiques sont des étiquettes atomiques
— les étiquettes appartiennent à une liste fixée
— les rôles sont liés à des positions syntaxiques
— chaque argument a exactement un rôle sémantique

Nous examinons ici très sommairement l'ensemble de ces points.

Nous l'avons vu plus haut, Levin et Rappaport Hovav (2005) parlent de « listes de rôles sémantiques ». Cela est dû au fait que toutes les approches considérées dans cette section ont pour objet d'associer à chaque verbe une liste de rôles sémantiques qui servent d'étiquettes aux participants de la situation dénotée par le verbe. Chacune des approches utilise pour définir cette liste un ensemble prédéfini de noms et définitions de rôles. Pour toutes, cet ensemble vise à être discret et fini, idéalement de petite taille, et ses rôles doivent avoir une définition indépendante des verbes qui les assignent. De plus, ils doivent être atomiques, et donc tous au même niveau sans relation de subsomption.

La figure 2.1, empruntée à Palmer *et al.* (2010), représente un ensemble de rôles généralement acceptés dans la littérature. Nous verrons par la suite que ce type d'ensemble n'est en réalité pas complètement satisfaisant, et ce pour un certain nombre de raisons.

Rôle	Description	Exemples
Agent	Initiateur de l'action, capable de volition	**Alex** a posé l'avion sans aucun problème.
Patient	Affecté par l'action, subit un changement d'état	Alex a cassé **la vitre**.
Thème	Entité déplacée ou localisée	Alex a jeté **le frisbee**. **La photo** se trouve sur le mur.
Expérienceur	Perçoit l'action mais ne la contrôle pas	**Alex** a vu Sam sortir.
Bénéficiaire	Pour le bénéfice de qui l'action est accomplie	Alex a loué un appartement pour **Sam**.
Instrument	Intermédiaire ou moyen utilisé pour accomplir une action	Sam a procédé à l'incision avec **un bistouri**.
Localisation	Lieu de l'objet ou de l'action	Il y a des monstres **sous le lit**.
Source	Point de départ	L'avion est parti de **Nairobi**. Nous avons entendu la rumeur par **un ami**.
But	Point d'arrivée	Le ballon est arrivé dans **la cage**.

TABLE 2.1: Ensemble de rôles sémantiques largement acceptés

Dowty (1991) regroupe les théories que nous avons qualifiées de « traditionnelles » sous le nom d'*indexation d'arguments*. Il s'agit en particulier des théories dans la lignée de la grammaire des cas

de Fillmore (1968) et des théories des thêta-rôles s'insérant dans le cadre de la théorie syntaxique du gouvernement et du liage (GB) (Chomsky, 1981, 1982, 1986).

Ces approches visent globalement à une correspondance spécifique entre arguments syntaxiques et arguments sémantiques : tout argument (syntaxique) d'un prédicat est supposé avoir exactement un rôle, et chaque rôle ne peut être assigné qu'à un seul argument du prédicat. Certaines théories du linking vont même plus loin et postulent une correspondance 1:1 entre arguments syntaxiques et sémantiques : tout rôle disponible pour un prédicat (c'est-à-dire tout argument sémantique d'un prédicat) doit être assigné à un de ses arguments syntaxiques (thêta-critère de Chomsky (1981)).

Dowty (1991) souligne qu'avec ce thêta-critère, l'utilisation de thêta-rôles dans le cadre GB a deux utilités. D'une part, elle permet de distinguer les arguments syntaxiques d'un verbe ayant un contenu sémantique (c'est-à-dire étant aussi un argument sémantique et donc recevant un rôle) des arguments purement syntaxiques (comme le *il* explétif). D'autre part, elle permet de garder une trace de l'identité des arguments sémantiques d'un prédicat à travers toutes ses réalisations syntaxiques. Pour l'auteur, l'indexation d'arguments est aussi une des fonctions des cas profonds de Fillmore (1968), puisque chaque argument syntaxique a une et une seule étiquette de cas et que ce sont ces étiquettes qui sont utilisées (et pas des étiquettes syntaxiques par exemple) pour référer aux arguments qui subissent des transformations syntaxiques.

Il pointe aussi que pour qu'un tel système d'indexation d'arguments fonctionne, on doit pouvoir choisir sans hésitation quel rôle sémantique assigner à chaque argument d'un verbe. Cela présuppose que le sens de chaque prédicat de la langue inclut clairement des arguments remplissant des rôles distincts et prédéfinis, ce qui n'est en réalité pas le cas puisqu'on trouve des contre-exemples dès lors que l'on essaye de définir précisément un ensemble de rôles.

Il s'agit là d'une des limites des approches traditionnelles que nous allons maintenant examiner.

2.1.2.2 Limites

Nous avons dit qu'il existait un accord général sur les rôles listés table 2.1, mais cet accord est tout relatif. En effet, il est difficile de s'accorder sur des définitions précises pour ces rôles. Il est donc aussi difficile de déterminer les cas précis dans lesquels chaque rôle devrait être utilisé, et de quels rôles supplémentaires on pourrait avoir besoin.

Dowty (1991) s'interroge sur ce qui permet de décider d'un degré de granularité pour les rôles sémantiques. Il pointe l'exemple du rôle Agent, sur l'existence duquel beaucoup s'accordent mais qui n'est pourtant pas forcément aussi intuitif qu'on pourrait le penser.

L'agent est généralement défini comme l'initiateur de l'action, typiquement capable de volition, et autorise la question *Qu'est-ce que X a fait ?* (où *X* est l'agent) et l'ajout de *délibérément* à la phrase. Pourtant, Cruse (1973) identifie quatre traits distincts pour ce rôle, qui peuvent être considérés comme des nouveaux rôles :

— *Volitif* : « un acte de volonté est exprimé ou impliqué » [traduction]
— *Effectif* : « quelque chose qui exerce une force (littéralement ou métaphoriquement) pas à cause d'une énergie interne propre mais à cause de sa position, son mouvement, etc. » [traduction]
— *Initiatif* : « initiation d'une action par formulation d'un ordre » [traduction]
— *Agentif* : « action effectuée par un objet dont on considère qu'il utilise son énergie propre pour effectuer cette action » [traduction]

Levin et Rappaport Hovav (2005) citent d'autres exemples de rôles pouvant être fragmentés, comme le rôle Instrument pour lequel Nilsen (1973) identifie quatre sous-rôles.

Dowty (1991) souligne le dilemme présenté par cette fragmentation des rôles : adopter des rôles fragmentés revient à perdre du pouvoir de généralisation, mais ignorer l'existence de possibles sous-rôles équivaut à perdre en capacité de distinction.

L'auteur ajoute au problème de fragmentation des rôles celui de la définition de frontières entre différents rôles. La sémantique suffit-elle à déterminer ces frontières ? Ou peut-on au contraire s'appuyer uniquement sur les constructions syntaxiques sans se soucier des parallélismes sémantiques si la syntaxe est suffisamment différente ?

(4) a. J'ai marché un kilomètre.
J'ai couru une heure.

b. Cette route mesure un kilomètre.
Ce film dure une heure.

Dans les exemples en (4a), Dowty identifie *un kilomètre* et *une heure* comme des ajouts, et ne considère pas comme évidente l'idée que les ajouts puissent avoir un rôle sémantique. En (4b), les mêmes syntagmes sont sous-catégorisés par les verbes des phrases dans lesquelles ils s'insèrent (cf. section 4.2.2.2), et à ce titre, il paraît certain qu'un rôle sémantique leur est assigné. Pour Dowty, dire que les mêmes syntagmes nominaux peuvent recevoir des rôles dans certains cas mais pas dans d'autres revient à ignorer le parallélisme sémantique entre (4a) et (4b) et à dire que c'est la syntaxe et pas le sens des phrases qui détermine quels rôles sont en jeu.

L'auteur utilise ces exemples pour montrer que l'assignation de rôles sémantiques n'est pas toujours transparente. La syntaxe et la sémantique doivent toutes deux jouer un rôle, et il est souvent difficile de les concilier de manière cohérente et de motiver ou pas la création d'un nouveau rôle. Cela pose pour lui un problème méthodologique : l'utilisation de la syntaxe pose un risque de falsification empirique, celui de créer des rôles pas réellement motivés sémantiquement, ou incohérents, et de justifier cela par des particularités syntaxiques.

Ces deux premiers problèmes montrent que l'idée d'une liste fixée et finie de rôles n'est pas aussi évidente qu'il n'y paraît. D'autres arguments ont été utilisés pour critiquer les autres points centraux des approches traditionnelles décrits en 2.1.2.1.

Nous avons mentionné supra que les approches traditionnelles voyaient les rôles comme atomiques, et donc tous au même niveau sans relation de subsomption. Le problème de fragmentation des rôles que nous avons décrit plus haut — à savoir le fait de pouvoir fragmenter les rôles traditionnels en plusieurs sous-rôles ayant des caractéristiques communes — contredit le postulat d'atomicité puisqu'on a affaire à des subsomptions de rôles. Cette idée d'atomicité des rôles est aussi contredite par le fait que comme pointé par Levin et Rappaport Hovav (2005), certains rôles sont fréquemment remplis par des syntagmes partageant un même marquage morpho-syntaxique (et ce souvent dans plusieurs langues), ce qui semble suggérer que les rôles se regroupent en sortes de « classes naturelles » et présentent donc probablement des traits communs.

Dowty (1991), à la suite de Jackendoff (1987), présente aussi des arguments qui vont à l'encontre du postulat des théories de l'approche traditionnelle selon lequel un et un seul rôle sémantique est assigné à chaque argument du verbe, et à l'encontre aussi de l'autre partie du thêta-critère de Chomsky (1981) qui postule que tout rôle est assigné à un et un seul argument syntaxique.

Jackendoff (1987) présente d'abord des cas où un même syntagme nominal doit porter plusieurs rôles sémantiques. Il indique qu'une phrase comme (5) intègre à la fois les évènements (5a) et (5b). *X* et *Z* ont donc chacun deux rôles sémantiques. Levin et Rappaport Hovav (2005) ajoutent à cet argument l'exemple (6), dans lequel Alex joue à la fois les rôles de Thème (ici entité subissant le mouvement) et Agent (initiateur de l'action).

(5) X achète Y à Z.

a. *Y passe du possesseur Z au possesseur X.*

b. *De l'argent passe du possesseur X au possesseur Z.*

(6) Alex a roulé jusqu'en bas de la colline.

Réciproquement, Jackendoff (1987) propose l'exemple (7), où pour lui deux syntagmes nominaux distincts, *Bill* et *lui*, remplissent un même rôle sémantique.

(7) Bill a pris des livres avec lui.

Dowty (1991) arrive aussi à cette conclusion. Pour lui, l'existence de prédicats comme *ressembler*, *similaire*, *près*, dont les deux arguments sont symétriques ((8a) a approximativement le même sens que (8b)) et donc pour lesquels on n'a aucune raison sémantique d'assigner un rôle sémantique différent à chacun de ces arguments, montre qu'un même rôle peut être assigné à deux arguments différents.

(8) a. Alex ressemble à Sam.
b. Sam ressemble à Alex.

Les approches traditionnelles se heurtent principalement à des problèmes de définitions : des rôles eux-même, et d'un ensemble fini de rôles satisfaisant. Il apparaît que ces problèmes seraient peut-être traitables si les rôles sémantiques avaient un degré de granularité différent de celui envisagé par les approches vues jusqu'ici. Nous nous pencherons donc dans un premier temps sur les approches dites de « rôles sémantiques généralisés », aux rôles bien plus larges. Enfin, nous nous intéresserons à des théories qui envisagent les rôles comme bien plus spécifiques et qui reposent sur, ou permettent, des groupements de prédicats.

2.1.3 Rôles sémantiques généralisés

Nous avons examiné plus haut les critiques des approches traditionnelles formulées par Dowty (1991). Ces approches tentent de définir empiriquement un ensemble de rôles sémantiques bien motivé, mais rencontrent certains problèmes. Tout d'abord la granularité idéale des rôles n'est pas claire puisque ceux de l'ensemble généralement admis peuvent être fragmentés en sous rôles et que les limites entre eux sont souvent vagues ; ensuite plusieurs contre-exemples montrent qu'on ne peut pas postuler de correspondance 1:1 entre les arguments d'un prédicat et les rôles sémantiques qu'il leur assigne.

Pour tenter de rendre compte de manière satisfaisante de la correspondance entre les arguments sémantiques d'un prédicat et les constituants qu'il sous-catégorise tout en évitant les problèmes rencontrés par les approches traditionnelles, Dowty (1991) propose une approche très différente. En s'inspirant des prototypes de Rosch et Mervis (1975), il propose l'idée de rôles prototypiques. Un proto-rôle est un ensemble de propriétés qui caractérisent un exemple prototypique de ce rôle. Tout Agent n'est pas nécessairement un être doué de raison qui cause volontairement un évènement, mais on considèrera un argument qui possède assez de ces propriétés comme un Agent. L'auteur considère ces propriétés comme des implications logiques pouvant être imposées par un verbe à un de ses arguments.

Dowty pose que dès lors qu'on accepte que les arguments n'ont pas tous le même degré d'appartenance à un rôle, on a seulement besoin de deux rôles de ce type pour formuler une théorie du linking efficace (l'auteur parle de *sélection argumentale* plutôt que de linking). Il propose donc deux proto-rôles : *proto-agent* et *proto-patient*, chacun caractérisé par les implications présentées table 2.2.

L'idée est qu'un verbe attribuera certaines de ces implications à chacun de ses arguments, et que chaque argument sera considéré proto-agent ou proto-patient à un degré différent en fonction du nombre d'implications de chaque rôle que le verbe lui attribue. Ces implications sont censées être sémantiquement indépendantes les unes des autres, et il existe aussi bien des verbes dont les arguments vérifient toutes les implications de leurs proto-rôles que d'autres dont ils n'en vérifient qu'une.

Proto-agent
- implication volitionnelle dans l'évènement ou l'état
- sentience (et/ou perception)
- causation d'un évènement ou du changement d'état d'un autre participant
- mouvement (relatif à la position d'un autre participant)
- existe indépendemment de l'évènement nommé par le verbe
Proto-patient
- subit un changement d'état
- thème incrémental
- affecté causalement par un autre participant
- stationnaire relativement au mouvement d'un autre participant
- n'existe pas indépendamment de l'évènement nommé par le verbe, ou n'existe pas du tout

TABLE 2.2: Propriétés des proto-rôles de Dowty (1991, p. 572, (27) et (28))

Dowty utilise les proto-rôles pour proposer son principe de sélection argumentale, ainsi que ses corollaires (Dowty, 1991, p. 576).

> **Principe de sélection argumentale :** Dans les prédicats avec sujet et objet grammaticaux, l'argument pour lequel le prédicat implique le plus de propriétés de proto-agent sera lexicalisé comme sujet du prédicat ; celui pour lequel le prédicat implique le plus de propriétés de proto-patient sera lexicalisé comme objet direct du prédicat.
>
> **Corollaire 1 :** Si un prédicat implique grossièrement le même nombre de propriétés de proto-agent et proto-patient pour ses deux arguments, chacun de ces derniers (l'un, l'autre ou les deux) peut être lexicalisé comme sujet du prédicat (même chose pour l'objet direct).
>
> **Corollaire 2 :** Dans un prédicat à trois places, l'argument non-sujet pour lequel le prédicat implique le plus de propriétés de proto-patient sera lexicalisé comme objet direct et celui pour lequel il en implique le moins sera réalisé comme oblique ou objet indirect. Si les deux arguments présentent le même nombre de propriétés de proto-patient, l'un, l'autre ou les deux pourront être réalisés comme objet direct.
>
> **Caractère non discret :** Les proto-rôles ne classent pas les arguments exhaustivement (certains arguments n'ont aucun des rôles) ni uniquement (certains arguments partagent un même rôle), ni encore de manière discrète (certains arguments peuvent être adaptés aux deux rôles à degré égal).

Le concept de proto-rôles permet de contourner les problèmes rencontrés par les approches traditionnelles. Il ne s'agit pas vraiment d'une décomposition en traits discrets : il est parfois difficile de départager quelle implication est déclenchée, et il n'y a pas de raison de fixer de limites discrètes puisque les arguments peuvent remplir des implications de proto-rôles différents. En évitant cette discrétisation, l'auteur contourne le problème de limites peu claires. Par ailleurs, la fragmentation de rôles n'est plus un problème : on peut tirer différentes classes d'arguments à différents niveaux de granularité des implications de chaque proto-rôle réalisées, et ainsi aisément obtenir une hiérarchie de rôles. En revanche, le fait que l'approche de Dowty (1991) ne permette pas d'assigner de rôles à tous les arguments des verbes qui en possèdent plus de deux pose le problème de la couverture des données : une annotation en proto-rôles sera nécessairement moins informative qu'une annotation dans laquelle tous les arguments sont étiquetés en rôles.

2.1.4 Théories incluant une classification sémantique des prédicats

Nous nous pencherons ici brièvement sur deux approches dont la première, la sémantique des cadres, inclut explicitement une conception des rôles sémantiques, et la deuxième, les classes de Levin (1993), a aussi comme idée sous-jacente l'existence de ces rôles. Les deux approches ont en commun de déboucher sur une classification sémantique des prédicats, mais aussi d'être actuellement utilisées en TAL sous forme de ressources que nous présenterons section 2.2.3 pour VerbNet et 2.2.6, puis chapitre 3 pour FrameNet.

2.1.4.1 Sémantique des cadres

À l'inverse de la proposition de Dowty (1991), la sémantique des cadres de Fillmore (1976, 1977b, 1982, 1985) évoque un modèle dans lequel la granularité de l'équivalent des rôles sémantiques est bien plus fine que dans les conceptions traditionnelles vues plus tôt.

Cependant, la redéfinition du concept de rôles n'est pas la motivation originelle de la sémantique des cadres : Fillmore (1982) s'intéresse à la compréhension de textes. Il propose une manière d'étudier le sens en mettant en avant la continuité entre langue et connaissance du monde. L'auteur ne présente pas de proposition formelle, mais plutôt un programme de recherche, un ensemble de points à considérer avant de pouvoir élaborer une théorie plus solide à partir des idées proposées.

Il introduit le concept de *cadre* (*frames*), qu'il rapproche de celui des *« case frames »* qu'il évoquait déjà dans (Fillmore, 1968), et qui correspondent à ce que nous avons décrit plus haut comme des listes de rôles associés à chaque verbe. Fillmore (1977a) décrit ces *« case frames »* comme permettant d'assurer le lien entre description de situations et représentations syntaxiques sous-jacentes. Il précise dans (Fillmore, 1982) qu'il entendait aussi que la compréhension de la structure sémantique d'un verbe nécessitait la compréhension des propriétés de cette scène ou situation.

L'auteur redéfinit les cadres comme « tout système de concepts organisés de manière à ce que pour comprendre n'importe lequel d'entre eux, il faille comprendre toute la structure dans laquelle il s'insère. Quand un des éléments de la structure est évoqué dans un texte [...], tous les autres sont automatiquement rendus disponibles » (Fillmore, 1982). Ainsi on ne peut comprendre l'idée « pension alimentaire » si on n'a pas connaissance du concept de divorce, et on ne peut concevoir ce concept sans idée de celui de mariage. Des mots permettent (et existent pour permettre) de donner accès à la connaissance de ces cadres aux participants de la communication : un locuteur applique un cadre à une situation et montre qu'il veut utiliser ce cadre en employant des mots connus pour l'évoquer.

Fillmore (1976, 1982) donne l'exemple (maintes fois repris) de l'ensemble des mots comme *acheter*, *vendre*, *payer*, *coût*, *monnaie*, etc. Pour comprendre ces mots, il faut avoir connaissance de et mentalement accéder à un cadre qui correspondrait à une situation prototypique dans laquelle « une personne acquiert contrôle ou possession d'une chose appartenant à une seconde personne, par accord mutuel, après avoir donné à cette personne une somme d'argent » (Fillmore et Atkins, 1992). C'est ce cadre qui motive l'existence des mots que nous venons d'évoquer. Ces mots, eux, indexent ou évoquent chacun une portion du cadre, puisqu'ils mettent chacun à l'avant-plan une partie différente de la situation décrite.

Les participants à la situation prototypique décrite par un cadre (l'*acheteur*, le *vendeur*, les *biens* et l'*argent* pour notre exemple) sont appelés *éléments* du cadre [3], et sont constitutifs de la définition d'un cadre. Ces éléments de cadre correspondent à des rôles sémantiques, amenés à être remplis par les arguments des mots évoquant un cadre. Comme ils sont spécifiques à un cadre, ils constituent des rôles à granularité bien plus fine que celle des approches traditionnelles présentées section 2.1.2.

3. On trouve des mentions du terme dès (Fillmore, 1976) mais exception notable, Fillmore et Atkins (1992) utilisent à la place le terme *catégories*.

La sémantique des cadres ne propose pas comme les approches évoquées plus haut de théorie du linking. Cela dit, Fillmore et Atkins (1992) proposent d'intégrer des descriptions lexicales en sémantique des cadres à une théorie du lexique mêlant information sémantique et syntaxique, où « chaque élément lexical [...] pourrait être associé à [...] la description de sa valence [...] qui spécifierait en termes à la fois sémantiques et syntaxiques ce que l'expression exige de ses constituants et son contexte et ce qu'elle apporte aux structures qui la contiennent ». Des informations ressemblant à la table 2.3 seraient donc disponibles dans le cadre TRANSACTION COMMERCIALE décrit plus haut. Fillmore et Atkins définissent un cadre pour les nom et verbe anglais *risk* et confrontent leur idée aux données en examinant un large corpus d'occurrences de ces mots pour observer et répertorier les réalisations de leurs arguments.

C'est de cette idée de ressource qu'est né FrameNet (Baker *et al.*, 1998), ressource sur laquelle porte notre travail, qui sera décrite section 2.2.6 et chapitre 3.

2.1.4.2 Classes de Levin

Levin (1993) postule que « le comportement d'un verbe, particulièrement en ce qui concerne l'expression et l'interprétation de ses arguments est en grande partie déterminé par son sens ». Pour elle, c'est le sens des verbes qui permet aux locuteurs de prédire leur comportement syntaxique, et les verbes qui partagent un même comportement syntaxique partagent aussi des composantes de sens.

Partant de cette idée, elle regroupe les verbes qui présentent les mêmes alternances syntaxiques, qu'elle définit comme des alternances dans l'expression des arguments d'un verbe, parfois accompagnées de variations de sens.

Elle reprend ainsi l'exemple classique de l'*alternance locative* qui caractérise par exemple le verbe *charger*[4] : ce verbe permet une réalisation de ses arguments aussi bien comme (9a) que (9b) [5]. Pour contraste, comme on peut le voir en (10) et en (11) et bien que le sens de ces verbes soit proche de celui de *charger*, *remplir* ne permet que la première option et *verser* que la deuxième. Ces verbes ne présentent donc pas l'alternance locative qui caractérise *charger*.

	Buyer	*Seller*	*Goods*	*Money*
BUY	Subj	(from)	D-Obj	(for)
SELL	(to)	Subj	D-Obj	(for)
CHARGE	(I-Obj)	Subj	(for)	D-Obj
SPEND	Subj	NULL	for/on	D-Obj
PAY	Subj	[I-Obj]	[for]	D-Obj
PAY	Subj	(to)	for	D-Obj
COST	(I-Obj)	NULL	Subj	D-Obj

TABLE 2.3: Valence sémantique et syntaxique (voix active) des verbes du cadre TRANSACTION COMMERCIALE (Fillmore et Atkins, 1992, p.79)

4. Les exemples sont transposés en français puisque l'alternance locative y existe comme en anglais.

5. De nombreux auteurs ont pointé que le sens de (9a) et (9b) n'est pas exactement le même : (9a) implique qu'à la fin de l'opération, le chariot est entièrement rempli, tandis que (9b) implique qu'à la fin de l'opération, toutes les pommes sont dans le chariot.

(9) a. Alex charge le chariot de pommes.
b. Alex charge les pommes dans le chariot.

(10) a. Alex remplit le chariot de pommes.
b. *Alex remplit les pommes dans le chariot.

(11) a. *Alex verse le chariot de pommes.
b. Alex verse les pommes dans le chariot.

Levin fait valoir qu'étant donné le postulat mentionné au premier paragraphe, regrouper les verbes présentant les mêmes alternances de diathèse permet d'obtenir des classes sémantiques de verbes qui partagent des composantes de sens (on voit effectivement figure 2.1 que les verbes de chaque classe sont sémantiquement proches). Elle suggère que les alternances de diathèse sont sensibles à ces composantes de sens : ce sont ces composantes qui déterminent si un verbe autorise ou non un certain argument. Elle répertorie donc un grand nombre d'alternances de diathèse pour l'anglais, et les utilise pour répartir 3 100 verbes distincts dans une hiérarchie de 240 classes.

L'auteure ne mentionne pas explicitement l'idée de rôles sémantiques mais cette idée est sous-jacente à son travail : Palmer *et al.* (2010, p. 16) soulignent que les alternances, qui sont supposées préserver le sens, correspondent en fait au mouvement d'un rôle sémantique spécifique d'un constituant à un autre. On ne trouve néanmoins dans les classes aucune mention d'étiquettes de rôles, ni de généralisations sur les propriétés communes des arguments de verbes appartenant à différentes classes.

Nous verrons section 2.2.3 que VerbNet, ressource qui étend les classe de Levin, fournit des rôles explicites pour les arguments des verbes de la ressource.

2.2 Ressources

Nous allons le voir, FrameNet, la ressource sur laquelle porte le reste de ce travail, permet à la fois des généralisations sur les lexèmes et sur les rôles sémantiques. Afin de mieux situer FrameNet, que nous examinerons plus en détail dans le chapitre suivant, nous passons donc ici brièvement en revue un certain nombre de ressources découlant d'entreprises de structuration du lexique avec généralisation sur les lexèmes et/ou sur les rôles.

2.2.1 Wordnet

Wordnet (Miller, 1995; Fellbaum, 1998), développé à la Princeton University, est une base de données lexicales rassemblant noms, verbes, adjectifs et adverbes de l'anglais dans une hiérarchie

Class	Syntactic Frames	Members
break 45.1	John broke the jar. / The jar broke. / Jars break easily.	*break, chip, crack, crash, crush, fracture, rip, shatter, smash, snap, splinter, snip, tear*
cut 21.1	John cut the bread. / *The bread cut. / Bread cuts easily.	*chip, chop, clip, cut, hack, hew, rip, saw, scrape, scratch, slash, slice, snip*
hit 18.1	John hit the wall. / *The wall hit. / *Walls hit easily.	*bang, bash, click, dash, squash, tamp, thump, thwack, whack, batter, beat, bump, butt, drum, hammer, hit, jab, kick, knock, lash, pound, rap, slap, smack, smash, strike, tap*

FIGURE 2.1: Exemples de classes de Levin (figure empruntée à Palmer *et al.* (2010, p. 16))

de concepts sous la forme d'ensemble de synonymes, ou *synsets*. En plus de leur ensemble de synonymes, les synsets comportent une brève définition, quelques phrases d'exemples, et des cadres de sous-catégorisation possibles pour les membres du synset. De plus, ces synsets sont liés les uns aux autres par des relations sémantiques comme l'antonymie, l'hyperonymie, la méronymie, ou encore la cause. La figure 2.2 montre un exemple de synset avec ses liens à d'autres synsets et leurs cadres syntaxiques.

WordNet 3.0 comporte 155 287 mots distincts, organisés en 117 659 synsets pour créer en tout 206 941 paires mot-sens. La ressource est utilisée par de nombreuses applications TAL, des WordNets ont été développés pour de nombreuses autres langues (par exemple les WordNet multilingues de Vossen (1998); Stamou *et al.* (2002); Bhattacharyya (2010)). En français, plusieurs adaptations différentes de WordNet coexistent : EuroWordNet (Vossen, 1998) puis (Jacquin *et al.*, 2006), WOLF (Sagot et Fišer, 2008), et JAWS (Mouton et de Chalendar, 2010) (étendu sous le nom de WoNeF par Pradet *et al.* (2013)).

2.2.2 PropBank

La Proposition Bank (ou PropBank) de Palmer *et al.* (2005) dans sa première version est une ressource qui ajoute une couche d'annotations sémantiques au Penn Treebank II (Marcus *et al.*, 1993, 1994), corpus arboré d'un million de mots du Wall Street Journal. PropBank se focalise au départ sur les verbes.

Les auteurs partent du postulat de Levin (1993) selon lequel la variation syntaxique s'explique par la sémantique sous-jacente des verbes et de leurs arguments, mais ne cherchent pas à cela d'explication théorique. Ils choisissent plutôt de proposer un corpus annoté qui permettra l'étude quantifiée des phénomènes de variation syntaxique et pourra servir de données d'entraînement pour les systèmes automatiques d'étiquetage en rôles sémantiques. Comme les auteurs souhaitent aussi s'abstraire de toute théorie spécifique, l'étiquetage se veut simple, générique et consistant sur

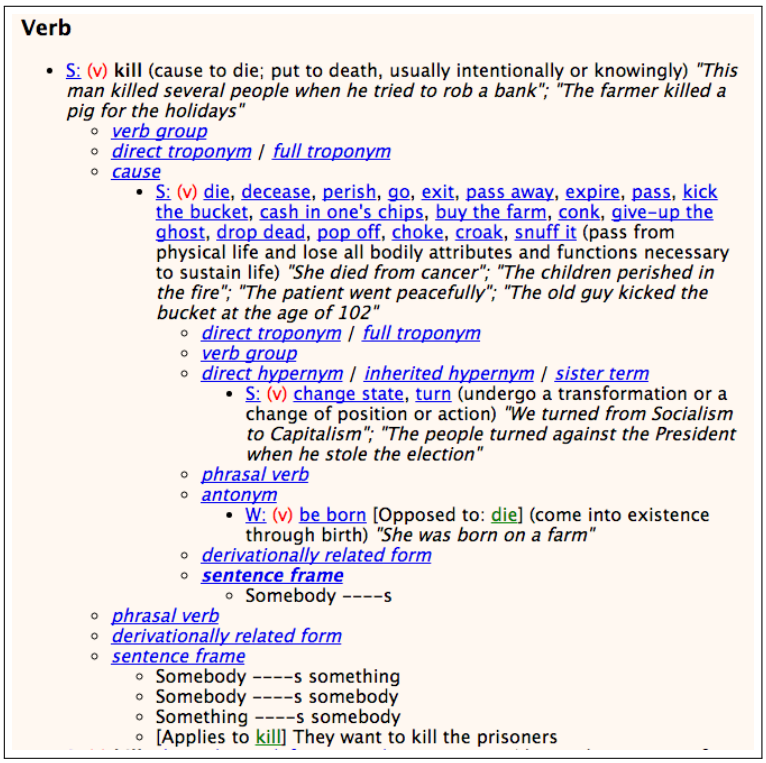


FIGURE 2.2: Extrait d'un synset WordNet et de ses relations à d'autres synsets

une grande quantité de données.

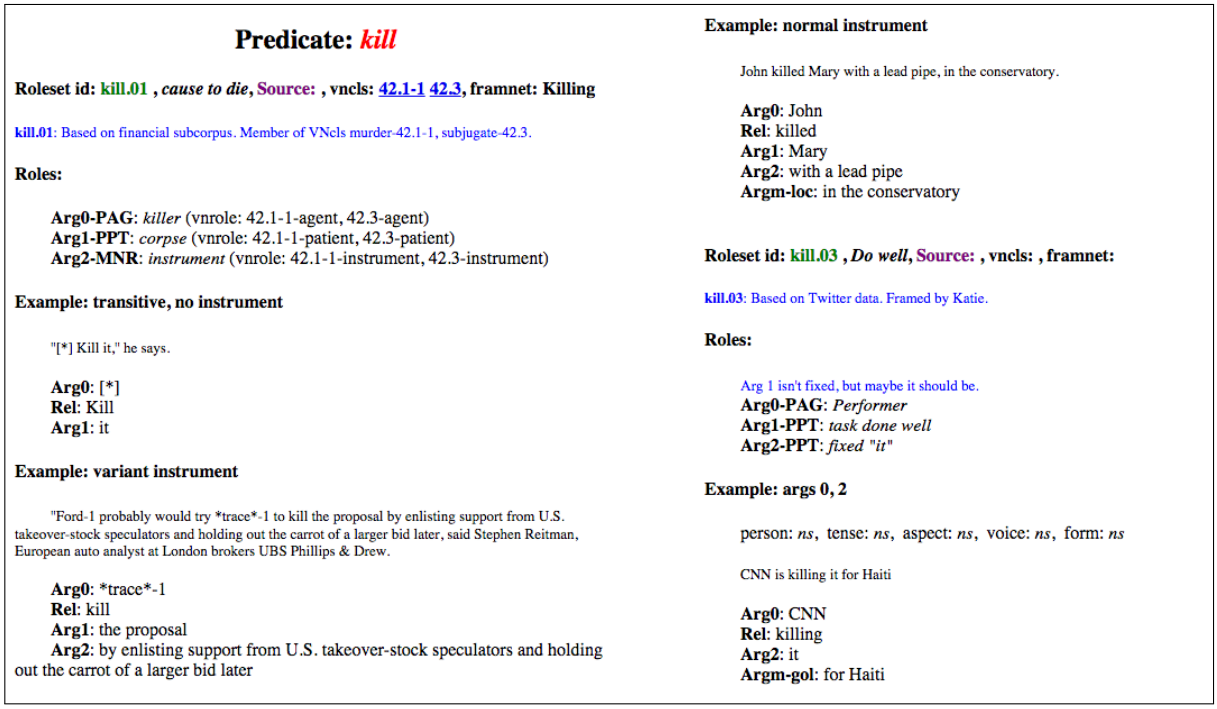
PropBank ajoute sur le Penn Treebank II une annotation des structures prédicat-argument avec étiquetage en rôles sémantiques des arguments. Les rôles sont spécifiques à chaque verbe, même si toutes les étiquettes sont les mêmes pour tous (Arg0, Arg1, Arg2 et ainsi de suite). Ainsi, les étiquettes permettent d'annoter simplement et de manière consistante un même rôle sémantique dans toutes les variations syntaxiques d'un même verbe alors même que les auteurs s'affranchissent des théories et de tout besoin de définir un ensemble universel de rôles. À côté de ces rôles numérotés spécifiques à chaque verbe, PropBank définit un ensemble de rôles génériques appelés **ArgM**. Ces rôles correspondent en fait aux ajouts comme le lieu, la durée ou la cause, exprimables pour une grande variété de verbes.

Notons qu'en ce qui concerne les étiquettes des arguments, elles semblent grosso modo numérotées selon un ordre d'oblicité des fonctions grammaticales canoniques des arguments du verbe. Cependant, l'Arg0 d'un verbe est généralement réservé à son argument qui possède le plus de propriétés d'agent prototypique (à la manière des proto-agents de Dowty (1991)) et son Arg1 à celui qui possède le plus de propriétés de Patient ou de Thème prototypique. C'est ce qu'on peut voir avec l'exemple du verbe *decline* présenté par Palmer *et al.* (2005, p. 6) : dans son sens *refuser quelque chose*, il a pour Arg0 la personne qui refuse et pour Arg1 la chose refusée, tandis que dans son sens *descendre incrémentalement*, il n'y a pas d'Arg0 (puisque pas d'agent) et l'Arg1 est l'entité subissant le mouvement. Les Arg0 et Arg1 sont l'exception à la règle puisque les arguments qui les suivent ne permettent pas de généralisations consistantes. Ainsi pour le verbe *say*, l'Arg0 est la personne qui parle (agent), l'Arg1 le message (thème), et l'Arg2 l'interlocuteur ; pour le verbe *break*, l'Arg0 est aussi l'agent (personne qui casse quelque chose), l'Arg1 est aussi le thème (objet cassé), mais l'Arg2 correspond à l'instrument utilisé. On voit bien que les Arg2 de ces deux verbes n'ont sémantiquement rien en commun.

PropBank fournit un lexique qui rassemble des **framesets**, sens à gros grain des verbes définis sur la base de toutes les possibles réalisations syntaxiques d'un verbe mettant en jeu le même **roleset**, c'est-à-dire la même liste d'arguments sémantiques étiquetés de rôles. Pour un même lemme verbal, on crée donc deux framesets différents dès lors que l'on repère des usages fonctionnant avec des ensembles de rôles différents. Ainsi les deux sens de *decline* présentés au paragraphe précédent appartiennent à deux framesets différents. Les alternances syntaxiques et possibilités d'omission qui n'affectent pas le sens du verbe, comme par exemple l'alternation causatif/inchoatif, sont considérées comme appartenant au même frameset. De plus, le type syntaxique des arguments n'entre pas en jeu dans la distinction en plusieurs framesets. Dans le guide d'annotation, un descripteur spécifique (e.g. *baker* pour le verbe *bake*) est fourni pour chaque rôle de chaque sens de chaque verbe, mais Palmer *et al.* (2010) précisent que ces descripteurs n'ont aucune valeur théorique et existent uniquement à des fins de documentation. L'ensemble des framesets d'un même verbe est regroupé dans un **frame file**.

La figure 2.3 présente un extrait du frame file du verbe *kill*, qui regroupe des framesets associés à ce verbe. On y voit bien pour chaque roleset une glose de chaque rôle ; comme nous l'avons mentionné plus haut, elles ne servent que de documentation et d'aide à l'annotation. Les codes en majuscules à côté de l'étiquette de chaque rôle semblent fournir un nom alternatif pour ce rôle (on suppose que PAG, PPT et MNR signifient respectivement proto-agent, proto-patient et manière), mais le dernier guide d'annotation PropBank en date (Bonial *et al.*, 2015) ne fait aucune mention de ces étiquettes alternatives, à part dans le cas des argM. Dans chaque frameset sont ensuite présentés des exemples de réalisations des rôles disponibles. Sont aussi indiqués lorsque cela s'applique les liens entre rolesets, classes VerbNet et cadres FrameNet, ainsi que les correspondances entre les rôles proposés par les trois ressources (nous verrons un peu plus bas quand ces correspondances ont été établies).

Comme nous l'avons mentionné, PropBank est constitué à la fois de ce lexique et d'un corpus d'annotations. Toutes les occurrences de verbes du Penn TreeBank II — à l'exception de *be* et des



Predicate: *kill*

Roleset id: kill.01 , *cause to die*, Source: , vncls: 42.1-1 42.3, framnet: Killing

kill.01: Based on financial subcorpus. Member of VNcls murder-42.1-1, subjugate-42.3.

Roles:

Arg0-PAG: *killer* (vnrole: 42.1-1-agent, 42.3-agent)
Arg1-PPT: *corpse* (vnrole: 42.1-1-patient, 42.3-patient)
Arg2-MNR: *instrument* (vnrole: 42.1-1-instrument, 42.3-instrument)

Example: transitive, no instrument

"[*] Kill it," he says.

Arg0: [*]
Rel: Kill
Arg1: it

Example: variant instrument

"Ford-1 probably would try *trace*-1 to kill the proposal by enlisting support from U.S. takeover-stock speculators and holding out the carrot of a larger bid later, said Stephen Reitman, European auto analyst at London brokers UBS Phillips & Drew.

Arg0: *trace*-1
Rel: kill
Arg1: the proposal
Arg2: by enlisting support from U.S. takeover-stock speculators and holding out the carrot of a larger bid later

Example: normal instrument

John killed Mary with a lead pipe, in the conservatory.

Arg0: John
Rel: killed
Arg1: Mary
Arg2: with a lead pipe
Argm-loc: in the conservatory

Roleset id: kill.03 , *Do well*, Source: , vncls: , framnet:

kill.03: Based on Twitter data. Framed by Katie.

Roles:

Arg 1 isn't fixed, but maybe it should be.
Arg0-PAG: *Performer*
Arg1-PPT: *task done well*
Arg2-PPT: *fixed "it"*

Example: args 0, 2

person: *ns*, tense: *ns*, aspect: *ns*, voice: *ns*, form: *ns*

CNN is killing it for Haiti

Arg0: CNN
Rel: killing
Arg2: it
Argm-gol: for Haiti

FIGURE 2.3: Extrait du frame file PropBank du verbe *kill*

emplois en tant qu'auxiliaires de *do* et *have* — ont été annotées : leur frameset a été identifié et leurs arguments ont été étiquetés en rôles. Palmer *et al.* (2010) indiquent que PropBank couvrait en 2010 5 213 verbes distincts donnant lieu à plus de 6 204 framesets. Nous n'avons pas trouvé de mention du nombre total d'annotations mais il s'agit grosso modo de tous les verbes d'un corpus de 1,75 millions de mots.

L'utilisation de rôles spécifiques à chaque verbe et sans relation les uns avec les autres rend en théorie impossible toute généralisation sur ces rôles (hormis celle citée supra sur les Arg0 et les Arg1) [6], ainsi que, forcément, l'utilisation de telles généralisations pour définir des regroupements de verbes.

Le projet SemLink (Palmer, 2009; Bonial *et al.*, 2013) vise à mettre en vis-à-vis Propbank, VerbNet (Kipper Schuler (2005), cf. section 2.2.3), FrameNet (Baker *et al.* (1998), cf. section 2.2.6 et chapitre 3), et les groupements de sens OntoNotes (Pradhan *et al.*, 2007). Des correspondances ont été établies entre instances annotées pour Propbank, VerbNet et FrameNet ainsi que rôle à rôle pour VerbNet et Propbank et pour VerbNet et FrameNet.

Par ailleurs, Bonial *et al.* (2014) ont étendu Propbank à de nouveaux types de prédicats : les noms, les adjectifs, et les expressions multi-mots. Des propbanks ont aussi été développés pour plusieurs autres langues, comme le chinois (Xue et Palmer, 2003), le coréen (Palmer *et al.*, 2006), l'hindi (Ashwini *et al.*, 2011) et l'arabe (Palmer *et al.*, 2008).

2.2.3 VerbNet

VerbNet (Kipper *et al.*, 2000; Kipper Schuler, 2005; Kipper *et al.*, 2008) est une hiérarchie de classes verbales modélisant et étendant les classes de Levin (1993). Cette extension porte à la fois sur les classes elles-mêmes (VerbNet compte 487 classes dont 273 de premier niveau contre 240 au total pour les classes de Levin ; la ressource propose aussi un niveau de hiérarchie supplémentaire) et sur leur couverture, puisque VerbNet recense 4 402 verbes distincts (contre 2 100 pour les classes de

6. Cela n'a pas empêché de nombreux travaux proposant des analyseurs PropBank de passer par un apprentissage automatique utilisant les étiquettes de rôles de manière « trans-verbes ».

Levin). VerbNet ajoute un typage des arguments des verbes utilisant 32 rôles sémantiques généraux, ainsi qu'une représentation du sens de chaque classe en termes de formules logiques.

Chaque classe VerbNet présente des cadres syntaxiques qui décrivent les possibles réalisations de surface de la structure argumentale des verbes de la classe. Chaque cadre syntaxique est défini au moyen du verbe, de rôles sémantiques, et du reste du matériel lexical nécessaire pour chaque construction. Ainsi figure 2.4, le cadre syntaxique AGENT *V* PATIENT décrit une des réalisations possibles de la structure argumentale des verbes de la classe `murder-42.1`. Dans ce cadre syntaxique, l'agent est le sujet du verbe et le patient son objet. Sont aussi indiquées des restrictions de sélection sur les rôles sémantiques, qui doivent être considérées comme des préférences plutôt que des contraintes rigides. Ainsi pour la classe `murder-42.1`, l'agent et le patient doivent tous deux être animés. Les cadres imposent parfois aussi des contraintes sur les prépositions autorisées et la nature syntaxique des constituants remplissant les rôles. Ainsi, le cadre AGENT *V* PATIENT {WITH} INSTRUMENT indique que l'instrument utilisé par l'agent des verbes de la classe se réalise comme oblique et n'autorise que la préposition *with*.

À chaque cadre syntaxique est associée une représentation en sémantique des prédicats qui dénote les relations entre évènements et participants et permet ainsi de représenter l'essentiel du sens de chaque cadre. Cette représentation est constituée d'une conjonction de prédicats sémantiques, par exemple `motion`, `contact` ou `cause` (VerbNet 3.2 compte 146 prédicats). Elle utilise une variable évènementielle `E` et des prédicats temporels `start(e)`, `during(e)` et `end(e)` pour pouvoir spécifier l'état ou le comportement de participants à différents moments de l'évènement. Palmer *et al.* (2010) pointent que ces prédicats sémantiques permettent d'obtenir des interprétations des rôles spécifiques à chaque classe.

On peut voir figure 2.4 (classe murder-42.1) qu'une classe VerbNet se compose des éléments suivants :

— liste des verbes membres de la classe ;
— liste des rôles sémantiques s'appliquant aux arguments des verbes de la classe ;
— liste des cadres syntaxiques valables pour les verbes de la classe, avec pour chacun un exemple et une représentation sémantique du cadre ;
— en haut à droite, liste des éventuelles sous-classes, qu'on retrouve décrites plus bas dans la classe.

Des VerbNets ont aussi été développés ou traduits pour plusieurs autres langues, comme l'italien (Merlo *et al.*, 2002), l'arabe (Mousser, 2010), le portugais brésilien (Scarton *et al.*, 2014), mais aussi plus récemment le français, avec Verb∋Net : Danlos *et al.* (2016) ont semi-automatiquement adapté VerbNet au français en utilisant deux ressources lexicales existantes, LVF (Dubois et Dubois-Charlier (1997), voir section 2.2.5) et le Lexique-Grammaire (Gross (1975), voir section 2.2.4).

2.2.4 Tables du lexique-grammaire

Le lexique-grammaire, initié par Gross (1975), est une base de données lexicales, syntaxiques et sémantiques pour les verbes, noms prédicatifs, expressions figées et adverbes du français. Les tables du lexique-grammaire comportent :

— **verbes :** 13 872 entrées (5 738 verbes distincts) réparties en 67 tables
— **noms prédicatifs :** 14 271 entrées (10 112 noms distincts) réparties en 81 tables
— **expressions figées :** 39 628 entrées (38 658 expressions distinctes) réparties en 69 tables
— **adverbes :** 10 488 entrées (9 326 adverbes distincts) réparties en 32 tables

La création des tables du Lexique-Grammaire part d'une critique des génératives, à qui Gross (1975) reproche d'élaborer des hypothèses à partir d'un nombre restreint d'exemples choisis, et de ne pas les confronter aux données. Gross cherche donc à accumuler des données empiriques aussi complètes et détaillées que possible, en répertoriant les phrases simples et en étudiant les transformations qu'elles peuvent subir. Il étudie essentiellement les propriétés formelles reposant

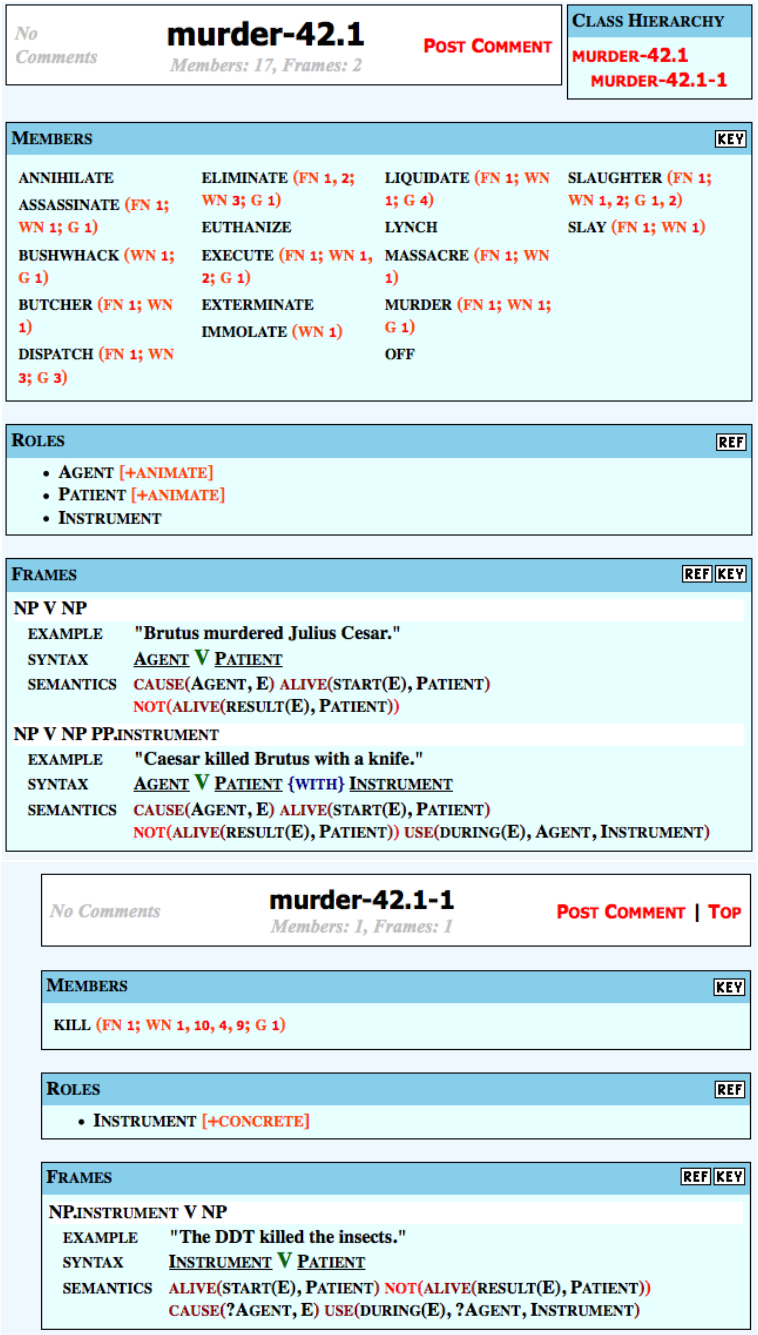


FIGURE 2.4: Classe VerbNet `murder-42.1` et sous-classe `murder-42.1-1`

sur la syntaxe plutôt que la sémantique, afin de garantir la reproductivité des tests mais certaines propriétés sémantiques ont été prises en compte lorsqu'elles pouvaient être testées de façon claire et surtout reproductible.

Chaque table du lexique-grammaire est en fait une classe regroupant des éléments lexicaux de même catégorie grammaticale qui partagent certaines propriétés définitoires. Pour les verbes, on distingue principalement les propriétés suivantes (Leclère, 1990) :

— **propriétés de structure :** pour un sens d'un verbe, sa construction la plus significative, généralement celle dans laquelle il prend le plus de compléments (comme N_0 *V Prép* N_1, par exemple *obéir* (table 33))

— **propriétés distributionnelles :** restrictions de sélection sur les compléments ou spécification des prépositions autorisées. Les restrictions de sélection peuvent porter sur la syntaxe (possibilité pour un complément d'être une complétive ou infinitive), ou sur le type sémantique d'un complément (humain, concret, etc.)...

— **propriétés sémantiques :** interprétation sémantique du procès (comme *V = transformer en nom dérivé de V*, par exemple *caraméliser* (table 32CV))

— **propriétés de constructions associées :** relation de paraphrase entre deux constructions (comme en (12), table 35S) ou alternance régulière bloquée (comme en (13), table 32NM)

(12) N_0 *V avec* N_1 (par exemple *Sam flirte avec Alex*)
= N_0 *et* N_1 *V* (par exemple *Sam et Alex flirtent*)

(13) N_0 *V* N_1 (par exemple *Le sac pèse 3 kg*)
**N*$_1$ *être V-pp par* N_0 (par exemple **3kg sont pesés par le sac*)

Les différentes classes des éléments lexicaux d'autres catégories grammaticales sont définies sur les mêmes types de critères. Toutes ces propriétés définitoires, ainsi que les traits définis ci-après, s'appliquent à une phrase simple caractéristique de l'emploi de l'élément lexical étudié.

Chaque table prend la forme d'une matrice, dont les lignes correspondent aux lexèmes de la classe et les colonnes aux traits syntaxico-sémantiques qui peuvent ou non s'appliquer à ces lexèmes et à leurs arguments dans une phrase. À la croisée des lignes et colonnes, un signe + ou - code le statut (autorisée ou non) de la propriété décrite pour le lexème concerné. Pour exemple, la figure 2.5 montre un extrait de la table 32H, qui regroupe les verbes dont la construction de base est N_0 *V* N_1 avec N_1 obligatoirement humain.

Des lexiques-grammaires sur le même principe ont aussi été développés pour d'autres langues, comme l'italien (D'Agostino *et al.*, 2004), le coréen (Nam, 2004), le polonais (Vetulani *et al.*, 2007), et le grec (Voskaki, 2011).

N0 =: Nhum	N0 =: Nnr	N0 être V-n	N1 V	<ENT>Ppv	Ppv =: se figé	Ppv =: Neg	<ENT>V	Neg	Aux =: avoir	Aux =: être	N1 être V-n	N0 V Nhum sur ce point	N1 être Vpp W	N1 =: Nconc métaphore	N1 =: Nabs métaphore	N0 être V-n (forme V-n)	N1 être V-n (forme V-n)	<OPT>
+	+	-	-	<E>	-	-	abandonner	-	+	-	-	-	-	-	-	<E>	<E>	La chance abandonne Max
-	+	-	-	<E>	-	-	abasourdir	-	+	-	-	-	+	-	-	<E>	<E>	Le bruit a abasourdi Max
+	-	-	-	<E>	-	-	trucider	-	+	-	-	-	+	-	+	<E>	<E>	Max a trucidé un ours
+	+	-	-	<E>	-	-	tuer	-	+	-	-	-	-	-	+	<E>	<E>	Max a tué Azor
+	-	-	-	<E>	-	-	tutoyer	-	+	-	-	-	-	-	-	<E>	<E>	Max tutoie Léa

FIGURE 2.5: extrait de la table 32H du lexique-grammaire

2.2.5 Les verbes français (LVF)

Les verbes français (ou *LVF*) de Dubois et Dubois-Charlier (1997) est, dans les mots de François *et al.* (2007), « un thésaurus de classes syntaxico-sémantiques, c'est-à-dire de classes sémantiques définies par la syntaxe ». La ressource, qui possède à la fois une version papier et une version électronique, compte 12 310 verbes distincts donnant lieu à 25 610 entrées structurées en classes sémantiques définies par la syntaxe.

Dubois et Dubois-Charlier posent le concept de *schème syntaxique*, défini par a) la nature des constituants de la phrase, leurs propriétés et leurs relations ; et b) les mots du lexique qui entrent dans les types de constituants définis. Les auteurs font l'hypothèse d'une correspondance entre ces schèmes syntaxiques et l'interprétation sémantique qu'en font les locuteurs. Ils cherchent donc à inventorier ces schèmes et à en établir une classification visant à rendre compte de cette adéquation.

Chaque entrée verbale est définie par son *opérateur*, forme codée du *schéma syntaxico-sémantique* qui synthétise l'ensemble de ses propriétés. C'est cet opérateur qui permet de faire le lien entre la syntaxe du verbe de l'entrée et son interprétation sémantique. Des regroupements successifs d'opérateurs similaires permettent de construire des catégories syntaxico-sémantiques, organisées en cinq niveaux de hiérarchie :

— les classes génériques (par exemple `F` : *verbes de type frapper ou toucher*)
— les classes sémantico-syntaxiques (par exemple `F1` : *frapper ou toucher quelqu'un*)
— les sous-classes syntaxiques (par exemple `F1b` : *frapper à mort, tuer, se tuer*)
— les sous-types syntaxiques (par exemple `F1a.1` : *frapper quelqu'un ou un animal à mort*)
— les variantes syntaxiques (par exemple `F1a.1 T1108` : *frapper quelqu'un ou un animal à mort ET transitif direct à sujet et objet humains et circonstant instrumental ou moyen*)

Chaque variante syntaxique regroupe un ensemble de verbes partageant le même sous-type syntaxique et autorisant une même construction. Pour chacune des entrées verbales, LVF fournit plusieurs informations (exemples figure 2.6) :

— l'opérateur (schéma syntactico-sémantique) de l'emploi
— son sens
— une ou des phrases d'exemple d'emploi en langue
— ses potentielles dérivations adjectivales et nominales

2.2.6 FrameNet

N.B. Nous présentons ici sommairement FrameNet afin de permettre au lecteur de le comparer aux autres ressources évoquées supra. Pour une description plus détaillée, se référer au chapitre 3.

Le projet FrameNet (Baker *et al.*, 1998), basé sur la sémantique des cadres de Fillmore (cf. section 2.1.4.1) et développé à l'ICSI de l'UC Berkeley, est une ressource en trois volets : a) un ensemble structuré de situations prototypiques, appelées *cadres*, associées à des caractérisations sémantiques des participants et propriétés y intervenant (les *éléments du cadre*) ; b) un lexique de lexèmes évoquant ces cadres ; et c) un ensemble d'annotations en cadres pour l'anglais. Ce que FrameNet appelle les « éléments du cadre » correspond aux rôles sémantiques assignables aux arguments des lexèmes pouvant évoquer le cadre. Notons dès à présent que pour le French FrameNet, nous appellerons ces éléments du cadre *rôles* pour plus de concision.

Les couples lexème-cadre déclenché comportent généralement un choix d'exemples annotés, piochés dans le BNC, dans le but de décrire exhaustivement la diversité des réalisations syntaxiques des éléments de cadre, en particulier pour en dériver des généralisations de linking. On parle pour ces exemples d'**annotations lexicographiques**. FrameNet propose aussi un deuxième type d'annotations : les **annotations full-text**. Il s'agit d'un ensemble de textes continus où toutes les occurrences de lexèmes déclenchant un cadre sont annotées. Ces annotations sont plus adaptées à l'apprentissage automatique puisqu'elles préservent la distribution des sens, ainsi que celle des

Classe générique F : Verbes de type frapper ou toucher (1727 entrées, 4 classes)

Classe sémantico-syntaxique F1 : frapper ou toucher qn (716 entrées, 6 sous-classes)

Sous-classe syntaxique F1b : frapper à mort, tuer, se tuer (141 entrées, 3 sous-types)

Sous-types

- F1b.1 : frapper qn, un animal à mort (123 entrées)
- F1b.2 : se tuer mutuellement (3 entrées)
- F1b.3 : se tuer (15 entrées)

[Toutes les classes]

Sous-type F1b.1 : frapper qn, un animal à mort (123 entrées)
Le truand a assassiné la vieille dame à coups de couteau.

Constructions: T1108; T1108, A10; T1208; T3100

T1108

Verbe	Opérateur	Sens	Phrases	Dérivations
abattre 4	**ict** qn à mort	descendre	Le tueur a~ P d'une balle à la sortie de l'hôtel.	
achever 6	**ict** qn à mort	tuer un blessé	On a~ P d'une balle dans la nuque. On a~ un animal blessé.	
aligner 5	**ict** qn à mort	abattre	Le tueur a~ P à la sortie de l'hôtel.	
zigouiller 1	**ict** qn à mort	stranguler, étrangler	Le truand z~ P à la sortie de l'hôtel.	

T1108, A10

Verbe	Opérateur	Sens	Phrases	Dérivations
frapper 22	**ict** qn à mort	opérer de nouveau	L'assassin a f~ de nouveau dans le quartier.	
tuer 1	**ict** qn à mort	assassiner	Le truand t~ le gardien. On t~ pour de l'argent.	-able, -eur

T1208

Verbe	Opérateur	Sens	Phrases	Dérivations
abattre 3	**ict** animal à mort	tuer	On a~ les boeufs avec une masse dans un abattoir.	-age, -oir
daguer 1	**ict** animal à mort	frapper animal avec couteau	Le chasseur d~ le cerf.	
détruire 3	**ict** animal+pl à mort	massacrer	On d~ les éléphants, les insectes.	-ion, -eur
égorger 1	**ict** animal à mort	couper la gorge	L'éleveur é~ les cochons.	-ment
estoquer 2	**ict** animal à mort	porter l'estocade à	Le torero e~ le taureau.	
flytoxer 1	**ict** animal à mort	tuer avec flytox	On f~ les mites dans le placard.	
massacrer 2	**ict** animal+pl à mort	détruire	Le chasseur m~ les éléphants en Afrique.	-ment
piquer 5	**ict** animal à mort	tuer en faisant piqûre	Le vétérinaire p~ un chien malade.	-ure
sacrifier 3	**ict** animal à mort	tuer en labo	On s~ les rats à la fin de l'expérience.	
tuer 2	**ict** animal à mort	tuer	Le chasseur t~ un lièvre, un faisan.	-age

T3100

Verbe	Opérateur	Sens	Phrases	Dérivations
emporter 7	qc **ict** qn à mort	faire mourir	La maladie a e~ P à vingt ans.	
enlever 10	qc **ict** qn à mort	emporter, tuer	Un cancer a e~ cet homme encore jeune.	
faucher 5	qc **ict** qn à mort	emporter, enlever	La maladie a f~ P à vingt ans.	
foudroyer 3	qc **ict** qn à mort	tuer	La congestion a f~ P dans la rue.	

FIGURE 2.6: Extrait des entrées du sous-type syntaxique LVF `F1b.1` : *frapper quelqu'un, un animal à mort*

réalisations syntaxiques des rôles. Pour les deux types d'annotation, les phrases ont été annotées en cadres et en morpho-syntaxe.

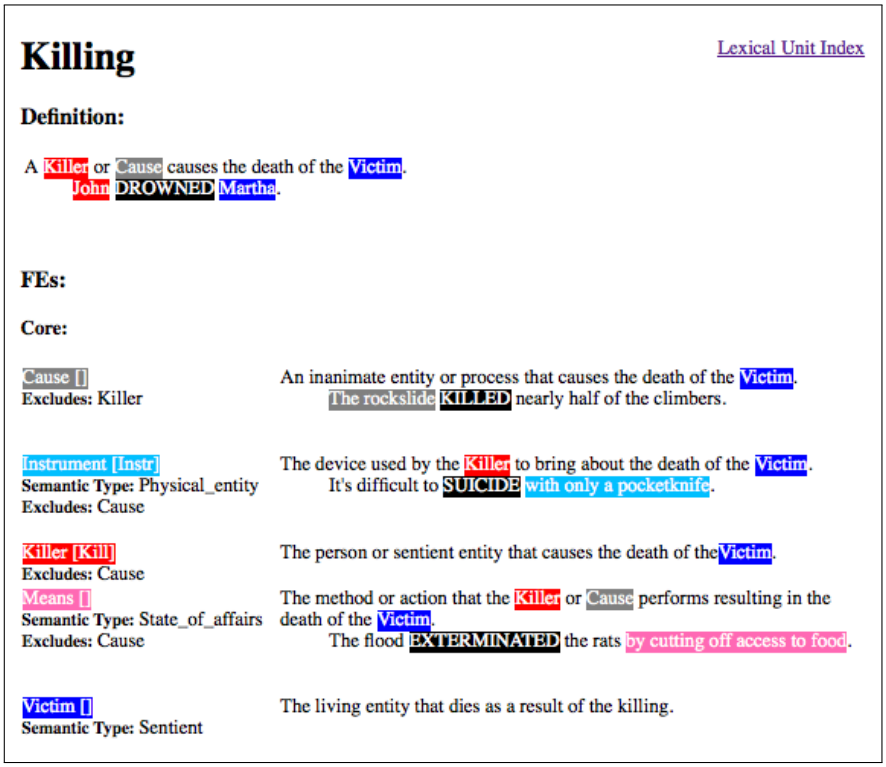
Actuellement, FrameNet comporte 13 311 unités lexicales pouvant déclencher 1 215 cadres. La ressource compte 172 609 cadres annotés dans le corpus lexicographique et 28 157 dans le corpus full-text.

Un cadre, comme KILLING présenté figures 2.7 à 2.9, comporte plusieurs composantes :

— une *définition* en langue naturelle de la situation prototypique décrite par le cadre (figure 2.7)

— un *ensemble d'éléments du cadre*, chacun défini en langue naturelle (figure 2.7). Les éléments peuvent avoir des liens entre eux (la réalisation d'un élément peut exclure celle d'un autre élément du même cadre, par exemple), et chacun d'entre eux peut être typé sémantiquement. De plus, ces éléments sont regroupés en fonction de leur statut dans le cadre (au cœur du sens d'un cadre, périphérique au sens du cadre ou extra-thématique au cadre).

— la liste des cadres avec lesquels il a une *relation* (figure 2.8). Il existe plusieurs types de relations entre cadres (héritage, perspective, antériorité, causativité...) dont nous parlerons plus en détail au chapitre 3.



Killing

Lexical Unit Index

Definition:

A Killer or Cause causes the death of the Victim.
John DROWNED Martha.

FEs:

Core:

Cause []
Excludes: Killer
An inanimate entity or process that causes the death of the Victim.
The rockslide KILLED nearly half of the climbers.

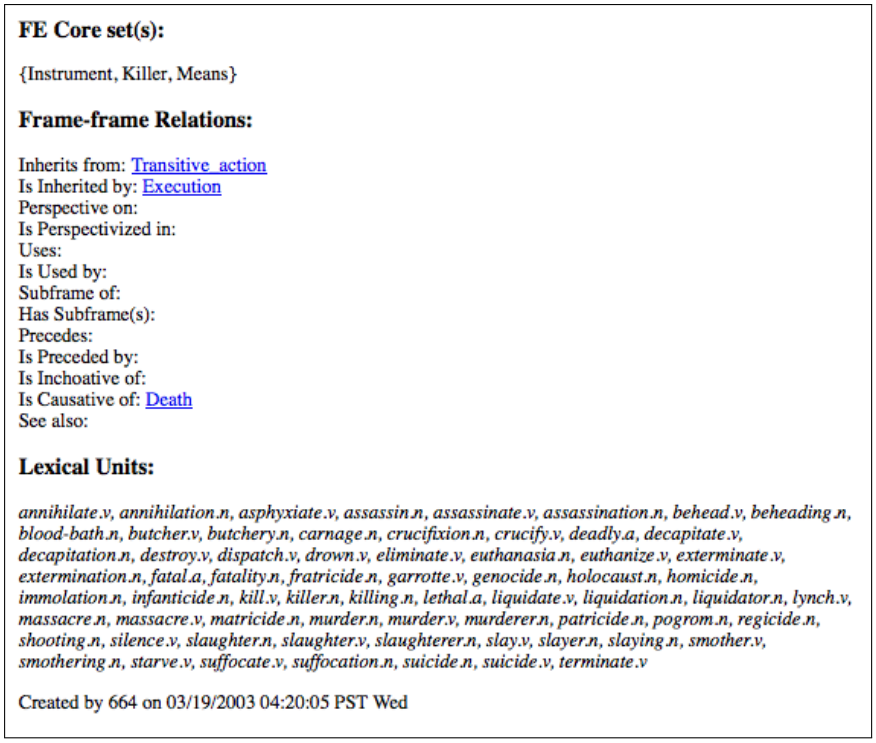
Instrument [Instr]
Semantic Type: Physical_entity
Excludes: Cause
The device used by the Killer to bring about the death of the Victim.
It's difficult to SUICIDE with only a pocketknife.

Killer [Kill]
Excludes: Cause
The person or sentient entity that causes the death of the Victim.

Means []
Semantic Type: State_of_affairs
Excludes: Cause
The method or action that the Killer or Cause performs resulting in the death of the Victim.
The flood EXTERMINATED the rats by cutting off access to food.

Victim []
Semantic Type: Sentient
The living entity that dies as a result of the killing.

FIGURE 2.7: Extrait du cadre KILLING : définition et rôles Core



FE Core set(s):

{Instrument, Killer, Means}

Frame-frame Relations:

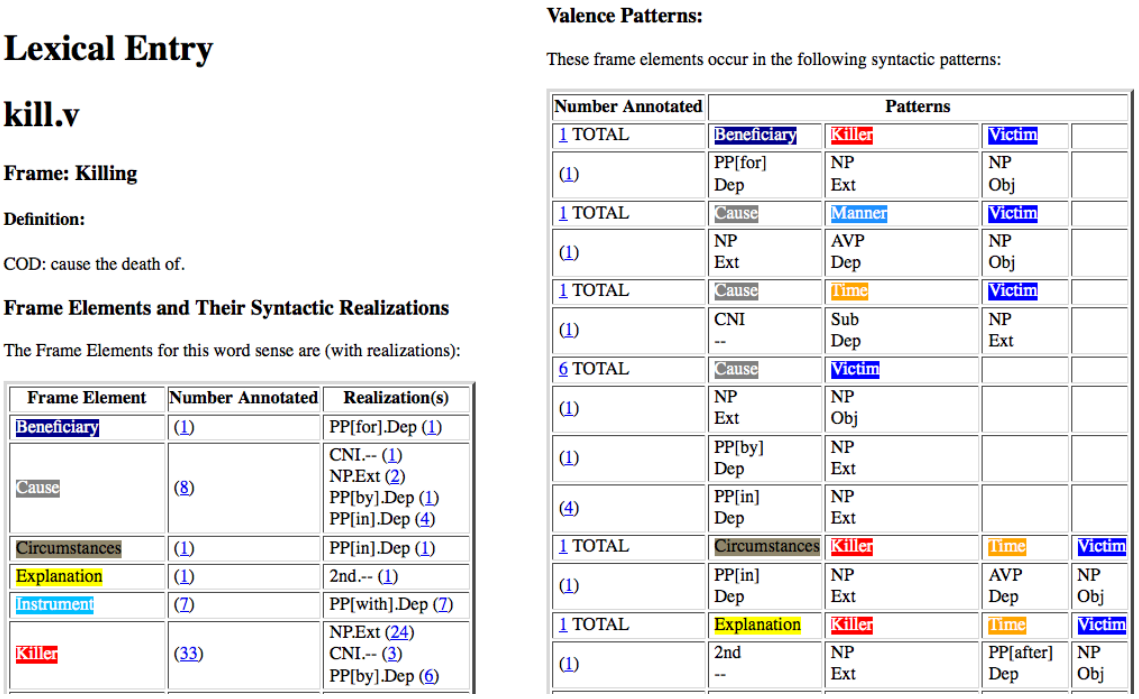
Inherits from: Transitive_action
Is Inherited by: Execution
Perspective on:
Is Perspectivized in:
Uses:
Is Used by:
Subframe of:
Has Subframe(s):
Precedes:
Is Preceded by:
Is Inchoative of:
Is Causative of: Death
See also:

Lexical Units:

annihilate.v, annihilation.n, asphyxiate.v, assassin.n, assassinate.v, assassination.n, behead.v, beheading.n, blood-bath.n, butcher.v, butchery.n, carnage.n, crucifixion.n, crucify.v, deadly.a, decapitate.v, decapitation.n, destroy.v, dispatch.v, drown.v, eliminate.v, euthanasia.n, euthanize.v, exterminate.v, extermination.n, fatal.a, fatality.n, fratricide.n, garrotte.v, genocide.n, holocaust.n, homicide.n, immolation.n, infanticide.n, kill.v, killer.n, killing.n, lethal.a, liquidate.v, liquidation.n, liquidator.n, lynch.v, massacre.n, massacre.v, matricide.n, murder.n, murder.v, murderer.n, patricide.n, pogrom.n, regicide.n, shooting.n, silence.v, slaughter.n, slaughter.v, slaughterer.n, slay.v, slayer.n, slaying.n, smother.v, smothering.n, starve.v, suffocate.v, suffocation.n, suicide.n, suicide.v, terminate.v

Created by 664 on 03/19/2003 04:20:05 PST Wed

FIGURE 2.8: Extrait du cadre KILLING : relations aux autres cadres et unités lexicales



Lexical Entry

kill.v

Frame: Killing

Definition:

COD: cause the death of.

Frame Elements and Their Syntactic Realizations

The Frame Elements for this word sense are (with realizations):

Frame Element	Number Annotated	Realization(s)
Beneficiary	(1)	PP[for].Dep (1)
Cause	(8)	CNI.-- (1) NP.Ext (2) PP[by].Dep (1) PP[in].Dep (4)
Circumstances	(1)	PP[in].Dep (1)
Explanation	(1)	2nd.-- (1)
Instrument	(7)	PP[with].Dep (7)
Killer	(33)	NP.Ext (24) CNI.-- (3) PP[by].Dep (6)

Valence Patterns:

These frame elements occur in the following syntactic patterns:

Number Annotated	Patterns			
1 TOTAL	Beneficiary	Killer	Victim	
(1)	PP[for] Dep	NP Ext	NP Obj	
1 TOTAL	Cause	Manner	Victim	
(1)	NP Ext	AVP Dep	NP Obj	
1 TOTAL	Cause	Time	Victim	
(1)	CNI --	Sub Dep	NP Ext	
6 TOTAL	Cause	Victim		
(1)	NP Ext	NP Obj		
(1)	PP[by] Dep	NP Ext		
(4)	PP[in] Dep	NP Ext		
1 TOTAL	Circumstances	Killer	Time	Victim
(1)	PP[in] Dep	NP Ext	AVP Dep	NP Obj
1 TOTAL	Explanation	Killer	Time	Victim
(1)	2nd --	NP Ext	PP[after] Dep	NP Obj

(a) Réalisations syntaxiques des éléments du cadre (extrait)

(b) Patrons de valence (extrait)

FIGURE 2.9: Extrait de l'entrée lexicale de *kill.v*, unité lexicale du cadre KILLING

— un ensemble d'*unités lexicales* pouvant évoquer le cadre (figure 2.8). Chacune de ces unités renvoie vers un ensemble annoté de phrases dans lesquelles elle déclenche le cadre (exemples (14)), ainsi que vers son entrée lexicale (qui recense les différentes réalisations syntaxiques des éléments du cadre pour cette unité lexicale (figure 2.9a), ainsi que des patrons de valence tirés des phrases annotées figure 2.9b)).

(14) a. 21 **KILLED** in a coach crash.
« 21 tués dans un accident de bus. »

b. She sees red and **KILLS** him with a very convenient weapon.
« Elle voit rouge et le TUE avec une arme bien pratique. »

c. And this figure excludes the recent **MURDER** of 2½ million Cambodian citizens by Pol Pot.
« Et ce chiffre exclut le MEURTRE récent de deux millions et demi de citoyens Cambodgiens par Pol Pot. »

d. Bad luck is as **LETHAL** for a politician as halitosis.
« La malchance est aussi FATALE pour un politicien que la mauvaise haleine. »

e. BUFFY THE VAMPIRE **SLAYER** : Valley girls meet vampires in this horror romp.
« Buffy la TUEUSE de vampires : les midinettes rencontrent les vampires dans cette comédie d'épouvante. »

Légende : Killer, Instrument, Cause, Victim

Les critères mis en avant pour le « découpage » en cadres ou pour le typage des rôles sont avant tout sémantiques, au contraire par exemple des classes de Levin (1993) (voir section 2.2.3) ou des tables du Lexique-Grammaire (Gross (1975), voir section 2.2.4). Cependant des critères syntaxiques sont tout de même utilisés dans FrameNet. Ainsi, sauf dans des cas dûment recensés, les rôles

sémantiques essentiels d'un cadre doivent pouvoir être réalisés localement pour tous les lexèmes évoquant ce cadre [7].

La structure FrameNet s'avère largement transposable à d'autres langues (Boas, 2009a). Des FrameNets ont donc été développés pour de nombreuses autres langues, dont l'espagnol (Subirats et Petruck, 2003), l'allemand (Burchardt *et al.*, 2006b), le chinois (Hao *et al.*, 2007), le suédois (Borin *et al.*, 2009), ou encore le français (Candito *et al.*, 2014a), sur lequel porte le reste de ce travail.

2.2.7 Comparaison des ressources présentées

Comme on peut le voir table 2.4, les ressources présentées sont comparables en certains points. Toutes sauf PropBank proposent une généralisation sur les lexèmes : WordNet regroupe les synonymes en synsets et indique leurs relations aux autres synsets ; VerbNet, le Lexique-Grammaire et LVF proposent des classes sémantiques hiérarchisées basées sur le comportement syntaxique des membres de ces classes ; et FrameNet regroupe des lexèmes se rapportant à une même situation prototypique et partageant des arguments qui correspondent aux participants de cette situation. VerbNet, le Lexique-Grammaire et LVF d'un côté et FrameNet de l'autre s'opposent donc en ce que les trois premiers effectuent des regroupements basés sur la syntaxe tandis que FrameNet met en avant des critères sémantiques.

FrameNet type d'ailleurs ces participants au moyen d'étiquettes de rôles spécifiques à chaque cadre mais sur lesquelles on peut généraliser puisque les cadres s'insèrent dans une structure et que des liens d'identité sont définis entre les éléments de cadres en relation l'un avec l'autre. VerbNet type aussi les arguments en rôles et permet de formuler des généralisations encore plus immédiates puisqu'un même ensemble restreint de rôles est utilisé pour toutes les classes. PropBank type aussi les relations prédicat-argument mais bien que portant les mêmes noms, les rôles sont définis séparément pour chaque lexème [8]. La ressource ne proposant pas de classes de lexèmes, il est impossible de partir des rôles existant pour en créer de plus larges. Quant aux trois autres ressources, elles ne proposent aucun typage spécifique de la relation prédicat-argument.

Enfin, les ressources n'intègrent pas toutes à leur lexique des lexèmes des mêmes catégories morphologiques. Toutes comprennent les verbes, catégorie à laquelle se restreignent VerbNet et LVF. Le Lexique-grammaire propose des entrées pour les noms prédicatifs et les adverbes, mais son mode de fonctionnement fait que ses tables ne peuvent mélanger d'éléments de catégories différentes. Il en va de même pour WordNet, qui intègre en plus des verbes, des noms, adjectifs et adverbes : les synsets ne mélangent pas de lexèmes de catégories différentes. Nous avons vu que PropBank portait au départ uniquement sur les verbes. Les noms et adjectifs y ont plus tard été ajoutés, mais la page actuelle du projet indique que ces derniers ne sont annotés que dans les corpus les plus récents. Quant à FrameNet, son lexique comprend des verbes, noms, adjectifs, adverbes, mais aussi des prépositions et conjonctions. Toutes ces catégories peuvent apparaître dans un même cadre ; ainsi le lexique du cadre CAUSATION comprend à la fois des verbes (*induce*, *cause*...), des noms (*cause*, *result*...), un adjectif (*causative*), des conjonctions (*for*, *since*...), et une préposition (*because of*).

FrameNet est donc la seule des ressources explorées à proposer à la fois une structuration du lexique en classes — basées sur la sémantique plutôt que la syntaxe — et des généralisations sur les rôles sémantiques, tout en intégrant des lexèmes de catégories diverses au sein de ses cadres. Ces propriétés sont les motivations principales du choix du cadre FrameNet pour le projet ASFALDA [9]. Mentionnons cependant dès à présent que nous avons utilisé les autres ressources pour le français

7. Les cas de non-localité autorisés sont recensés, et relèvent de phénomènes syntaxiques comme le contrôle ou les constructions à verbe support (Ruppenhofer *et al.*, 2006, pp. 27-31).

8. à l'exception des Arg0 et Arg1 qui nous l'avons vu correspondent généralement aux proto-agent et proto-patient de Dowty (1991)

9. Ces motivations seront abordées de manière un peu plus détaillée — en particulier pour ce qui concerne le choix de FrameNet plutôt que PropBank ou VerbNet — au chapitre suivant, section 3.5.

présentées supra, à savoir LVF et les tables du lexique-grammaire, pour aider aux distinctions de sens lorsqu'elles nous ont posé problème.

Au chapitre 3, nous nous attacherons à décrire le modèle FrameNet de manière plus approfondie. Nous nous intéresserons d'abord à la structure et à la méthodologie du projet séminal de FrameNet de l'anglais — le FrameNet de Berkeley, ou BFN —, avant de nous pencher sur les projets d'adaptation de FrameNet à d'autres langues que l'anglais, ainsi que d'autres types de projets s'appuyant sur le modèle FrameNet. Enfin, nous utiliserons les points déjà explorés pour mieux motiver notre choix d'appuyer notre ressource sur FrameNet plutôt que sur un autre modèle.

Ressource	**WN**	**PB**	**VN**	**L-G**	**LVF**	**FN**
Typage des relations prédicat-argument	non	oui	oui	non	non	oui
Généralisation sur les rôles	non	non (sauf Arg0 et Arg1)	oui	non	non	oui
Généralisation sur les lexèmes	oui	non	oui	oui	oui	oui
POS traitées	Adj Adv - N - V	Adj - - N - V	- - - - - V	- Adv - N - V	- - - - - V	Adj Adv Conj N Prep V

TABLE 2.4: Comparaison des différentes ressources présentées

CHAPITRE 3

FRAMENET

Sommaire

3.1 Introduction **31**
3.2 FrameNet : structure, annotations, lexique **33**
3.2.1 Réseau de cadres 33
3.2.1.1 Qu'est-ce qu'un cadre? 33
3.2.1.2 Rôles 34
3.2.1.3 Relations entre rôles (relations intra-cadres) 36
3.2.1.4 Relations inter-cadres 39
3.2.2 Annotations 42
3.2.3 Lexique 43
3.3 Méthodologie du projet FrameNet **44**
3.3.1 Organisation du travail 45
3.3.2 Création et délimitation des cadres 47
3.3.2.1 Critères de regroupement des lexèmes en cadres 48
3.3.2.2 Attribution de cadres aux mots polysémiques 49
3.3.3 Pratiques d'annotation 50
3.3.3.1 Zoom sur l'instanciation nulle 51
3.4 Projets d'adaptation de FrameNet dans d'autres langues que l'anglais **53**
3.4.1 Espagnol : Spanish FrameNet (SFN) 53
3.4.2 Japonais : Japanese FrameNet (JFN) 54
3.4.3 Allemand : le projet SALSA 55
3.4.4 Suédois : Swedish FrameNet++ (SweFN++) 55
3.4.5 Une diversité d'autres projets 56
3.4.5.1 Application du modèle FrameNet à des domaines spécialisés 56
3.4.5.2 Création automatique de ressources à partir du FrameNet de Berkeley 57
3.5 Motivation de notre choix d'utiliser FrameNet **58**

3.1 Introduction

Dans le chapitre 2, nous avons rapidement présenté la sémantique des cadres (section 2.1.4.1) et FrameNet, ressource qui en est adaptée (section 2.2.6). Le French FrameNet, ressource sur laquelle portera le reste de ce travail, s'appuie sur le modèle FrameNet; aussi le présent chapitre est-il consacré à présenter ce modèle de manière plus détaillée. Nous commençons par y décrire la structure du FrameNet de Berkeley, ainsi que la méthodologie de ce projet originel. Il s'agit d'une

présentation factuelle : elle nous permettra lorsque nous décrirons dans les chapitres suivants à la fois notre projet et la ressource obtenue de pointer les différences entre le French FrameNet et le FrameNet de Berkeley. Après cette présentation du FrameNet séminal, nous passons en revue une variété de projets d'adaptation de FrameNet à d'autres langues que l'anglais, ainsi que d'autres types de projets s'appuyant sur le modèle FrameNet. Enfin, nous utilisons les points déjà explorés pour mieux motiver notre choix d'appuyer notre ressource sur FrameNet plutôt que sur un autre modèle.

Le projet FrameNet (Baker *et al.*, 1998; Lowe *et al.*, 1997; Ruppenhofer *et al.*, 2006), basé sur la sémantique des cadres de Fillmore (vue au chapitre précédent section 2.1.4.1) et développé à l'ICSI de l'UC Berkeley, a créé une ressource tripartite. Le premier élément de la ressource est un ensemble structuré de situations prototypiques, appelées *cadres* (*frames*), associées à des caractérisations sémantiques des participants et propriétés y intervenant, appelées les éléments du cadre (*frame elements*), que nous appellerons ci-après *rôles* pour plus de concision. L'expression *ensemble structuré* réfère au fait que les cadres sont liés entre eux par des relations, comme l'héritage ou la causalité. La ressource comporte aussi un lexique de lexèmes évoquant ces cadres, les unités lexicales (*lexical units*), auxquelles nous réfèrerons sous le nom de *déclencheurs*. Enfin, FrameNet propose un ensemble d'annotations en cadres pour l'anglais. On distingue deux ensembles d'annotations nés de deux procédures distinctes. Objectif originel du projet, l'annotation lexicographique, sur un choix de phrases piochées dans le BNC pour servir d'exemples aux couples déclencheur-cadre, a pour but de décrire exhaustivement la diversité des réalisations syntaxiques des rôles, en particulier pour en dériver des généralisations de linking. Par opposition, les annotations de plein texte (*full-text annotation*) consistent en un ensemble de textes continus où toutes les occurrences de paires déclencheur-cadre ont été annotées.

La version des données sur laquelle s'appuie le reste de notre travail est la version 1.5, sortie en septembre 2010, dernière version à ce jour. Elle comporte 11 829 déclencheurs distincts pouvant évoquer 1 019 cadres. La ressource compte 149 931 occurrences de déclencheurs annotées dans le corpus lexicographiques et 23 087 dans le corpus de plein texte.

La sémantique des cadres, puis le projet FrameNet, sont nés de plusieurs motivations distinctes. Au tout départ, Fillmore (1982) s'intéressait à la modélisation de la compréhension de textes et concevait les cadres comme un ensemble de concepts organisés en une sorte d'arrière-plan du discours, auxquels les locuteurs auraient accès grâce à l'emploi de mots conventionnels. La sémantique des cadres et le projet FrameNet présentent néanmoins d'autres intérêts. Comme nous l'avons vu au chapitre précédent, les rôles sémantiques permettent à la fois de typer la relation sémantique qu'un prédicat entretient avec chacun de ses arguments et d'exprimer la correspondance probable de ces arguments sémantiques avec les constituants sous-catégorisés par le prédicat. Un cadre donné lie un ensemble de rôles à un ensemble de déclencheurs, et permet ainsi de généraliser à la fois sur les deux points susmentionnés et sur le sens des lexèmes, puisque ces derniers sont regroupés en cadres qui peuvent être vus comme des classes sémantiques. L'existence de relations liant les cadres entre eux, dont nous parlerons en 3.2.1.4, permet de généraliser encore davantage sur ces trois dimensions.

D'un point de vue pratique, obtenir une grande quantité de données annotées en sémantique présente aussi un grand intérêt : ces données sont utilisables pour entraîner des parseurs sémantiques. Fillmore *et al.* (2003) présentent encore d'autres utilités concrètes de la ressource. La première se situe dans le champ de la lexicographie, puisque les exemples en corpus des différents sens d'un mot et des réalisations possibles de leurs participants seront utiles à des lexicographes souhaitant créer une nouvelle entrée de dictionnaire. En ce qui concerne les applications TAL, une représentation du texte en sémantique des cadres peut être utile en extraction d'information, compréhension de la langue, traduction automatique, etc. Enfin, pour Fillmore *et al.* (2003), la ressource peut être utilisée en recherche sémantique pour étudier la polysémie, les collocations, les divergences syntaxe-sémantique, l'omissibilité lexicalement autorisée, l'inférence, etc.

Comme nous l'avons mentionné plus haut, dans ce chapitre, nous présenterons tout d'abord une description plus précise du FrameNet de Berkeley, c'est-à-dire les cadres organisés en réseau, les annotations et le lexique. Nous aborderons ensuite des points spécifiques de la méthodologie du projet FrameNet, en nous intéressant à l'organisation du travail, aux critères de délimitation des cadres et à diverses pratiques d'annotation. Dans une autre section, nous passerons brièvement en revue plusieurs projets de constitution de FrameNets en d'autres langues que l'anglais — en cela similaires à notre projet —, ainsi que d'autres types de projets s'appuyant sur le modèle FrameNet. Enfin, nous reprendrons brièvement certaines des caractéristiques de FrameNet que nous aurons exposées afin d'expliquer les motivations de notre choix de FrameNet comme modèle sur lequel appuyer notre ressource dans le cadre du projet ASFALDA.

3.2 FrameNet : structure, annotations, lexique

La base de données produite par les activités du projet FrameNet comprend trois volets. Le premier est un réseau de cadres, constitué d'un ensemble de cadres et d'un réseau de relations qui les lient les uns aux autres. Viennent ensuite un corpus d'annotations, et un lexique des déclencheurs pouvant évoquer des cadres. Ce lexique comporte des entrées lexicales, qui identifient pour chaque déclencheur (appariement d'un lexème avec un cadre) la variété de manières dont chaque rôle du cadre est syntaxiquement réalisé dans les annotations. Chacun des éléments susmentionnés est décrit dans une des sections qui suivent.

3.2.1 Réseau de cadres

Nous reprenons ici les définitions des notions de cadre, rôle, relations, etc. telles que présentées par FrameNet. Nous allons voir que ces définitions sont majoritairement exprimées en termes sémantiques, avec très peu de références à des concepts syntaxiques. Nous verrons au chapitre 4 que pour le French FrameNet, il nous a parfois été nécessaire de retravailler certaines notions, comme par exemple les critères de nucléarité d'un rôle, pour pouvoir les utiliser de la manière la plus cohérente possible.

3.2.1.1 Qu'est-ce qu'un cadre ?

Nous l'avons dit plus haut, un cadre est une représentation schématique d'une structure conceptuelle décrivant un type particulier de situation, objet ou évènement ainsi que ses participants, parties, ou phases. FrameNet les appelle en anglais *frame elements* (ou FEs), mais nous utiliserons le terme *rôles* pour plus de concision.

Un cadre possède plusieurs éléments constitutifs, à commencer par un *nom*, qui lui est propre. Vient ensuite la *définition* du cadre, sur laquelle ne pèsent pas de restrictions formelles. Il s'agit d'une description de la situation prototypique ou entité à laquelle réfère le cadre, qui mentionne en général les rôles importants et leurs relations les uns aux autres. On y trouve aussi parfois des exemples censés aider les annotateurs à avoir en tête les différentes possibilités de réalisation de ces rôles.

Ces *rôles*, qui permettent d'identifier les participants de la situation prototypique décrite par le cadre, sont aussi partie intégrante du cadre. Ceci est développé section 3.2.1.2.

Les cadres comportent aussi éventuellement un *type sémantique*. Deux sortes de types sémantiques peuvent s'appliquer à un cadre : les *types ontologiques* et les *types de cadres*.

Les *types ontologiques* appartiennent à la hiérarchie de types dont un extrait, emprunté à Lönneker-Rodman et Baker (2009), est donné figure 3.1. Si un type ontologique est assigné à un cadre, tous les déclencheurs de ce cadre doivent être de ce type (ou d'un type plus spécifique). Il en est de même pour les déclencheurs des cadres héritant de ce cadre. En pratique, dans les

données FrameNet 1.5, seuls 31 cadres sur 1 019 au total sont explicitement typés. Bien que tous les cadres héritant d'un cadre typé doivent être considérés comme du même type, seuls certains sont explicitement annotés comme tels. Ainsi LOCATIVE_RELATION et CONTAINMENT_RELATION contiennent tous deux l'information selon laquelle leur type est *locative relation*, mais le type des autres cadres héritant de LOCATIVE_RELATION n'est pas indiqué dans le cadre. Le nombre total de cadres dotés d'un type ontologique est en fait de 45, cadres héritiers inclus.

De leur côté, les *types de cadre* fournissent une meta-information sur le cadre (et pas sur les déclencheurs comme dans le cas des types ontologiques). Il existe deux types de cette catégorie qui peuvent être assignés à un cadre. Le type *non-lexical* s'applique aux cadres ne pouvant être évoqués par aucun lexème et n'existant que pour lier sémantiquement entre eux d'autres cadres. Pour sa part, le type *non-perspectivé* est utilisé pour signaler un cadre contenant une grande diversité de déclencheurs qui partagent un arrière-plan (décrit par le cadre) mais pas le point de vue présenté sur cet arrière-plan. 27 cadres sur 1 019 au total sont signalés comme non-perspectivés, et 111 comme non-lexicaux.

En plus de ces éléments constitutifs de cadre, l'implémentation du réseau de cadres donne accès à d'autres informations. Ainsi, à partir d'un cadre donné, on peut visualiser ses potentielles relations aux autres cadres du réseau (section 3.2.1.4), ainsi que les ensembles d'annotations (section 3.2.2) et entrées lexicales (section 3.2.3) de chaque déclencheur du cadre.

3.2.1.2 Rôles

Un *rôle* nomme et décrit un participant ou une propriété de la situation prototypique décrite par le cadre auquel il appartient. Lönneker-Rodman et Baker (2009) précisent que tous les déclencheurs d'un cadre doivent avoir les mêmes rôles. De plus, chaque rôle est spécifique à son cadre, bien que comme nous le verrons en 3.2.1.4, les rôles de deux cadres puissent être liés par une relation d'identité. L'expression linguistique qui dans le contexte d'un cadre évoqué par un déclencheur remplit un des rôles de ce cadre est appelée le *remplisseur* du rôle.

Chaque rôle de chaque cadre possède plusieurs caractéristiques, à commencer par un *nom*.

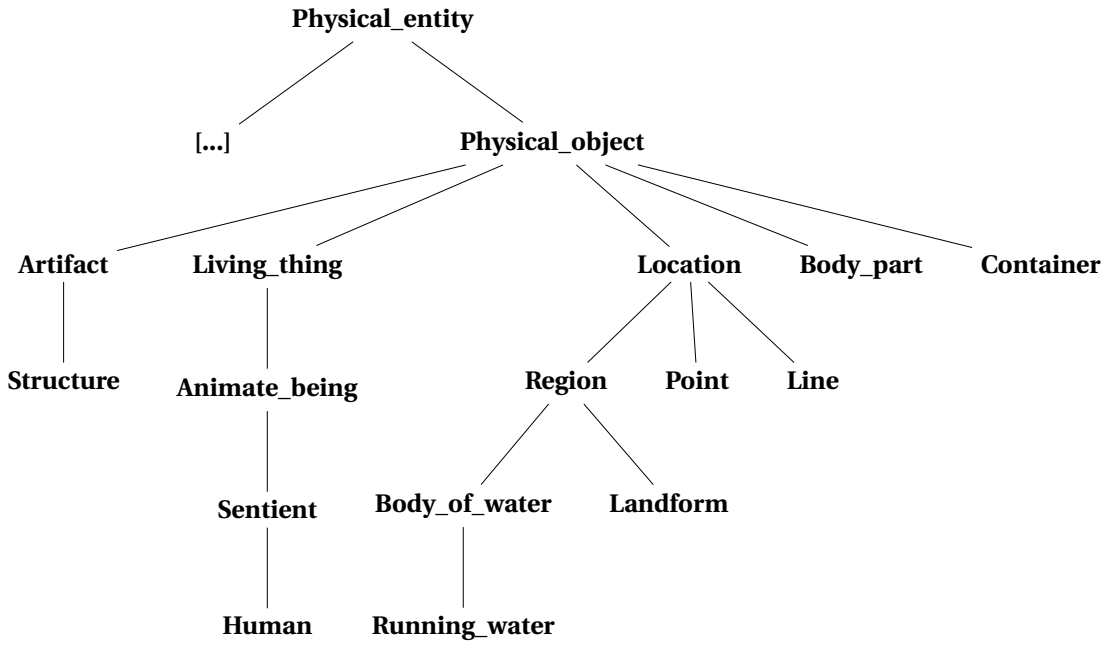


FIGURE 3.1: Extrait de la hiérarchie de types ontologiques FrameNet (figure empruntée à Lönneker-Rodman et Baker (2009, fig 2., p. 422))

Nous venons d'annoncer que chaque rôle était spécifique à un cadre, mais les noms de rôles ne sont pas nécessairement uniques : des cadres décrivant par exemple des situations liées sémantiquement pourront avoir des rôles de même nom. Il existe d'ailleurs des conventions, ainsi le rôle correspondant à l'agent de cadres décrivant des situations de type communication s'appelle généralement *Speaker*. Cependant, deux rôles portant le même nom n'ont pas nécessairement la même sémantique, et inversement deux rôles relativement similaires peuvent porter des noms différents.

Chaque rôle comporte une *définition*. Il s'agit d'une description informelle, en langue naturelle, du sens du rôle par rapport à son cadre. Cette définition utilise parfois d'autres rôles. Elle peut comporter des exemples, ainsi que potentiellement des patrons de réalisations syntaxiques canoniques, mais cela est très rare, certainement parce que le parti pris FrameNet est de fournir une définition avant tout sémantique.

Un rôle peut être assorti d'un *type sémantique*. Il s'agit d'un type ontologique, appartenant à la hiérarchie de types déjà vue figure 3.1, et qui correspond à une restriction de sélection sur le remplisseur du rôle. Pour le cadre KILLING, le type sémantique *Sentient* est associé au rôle Victim, c'est à dire qu'un constituant remplissant ce rôle doit référer à un être doué de sensation.

À chaque rôle est aussi assigné un statut de *nucléarité* (*coreness*), qui a trait à l'importance conceptuelle du rôle dans le cadre. Les distinctions présentées dans FrameNet ne nous paraissant pas toujours facile à appliquer, nous établirons au chapitre suivant (section 4.2.3.1) une définition de la nucléarité propre au French FrameNet, basée sur des critères plus précis.

Les *rôles noyaux* (*core frame elements*) sont les plus centraux. Dans les mots de Ruppenhofer *et al.* (2006), un rôle noyau « instancie un composant d'un cadre conceptuellement nécessaire, tout en rendant le cadre unique et différent des autres » [traduction]. Il s'agit là d'une définition sémantique, mais des critères formels sont parfois aussi considérés. Ruppenhofer *et al.* (2006) listent plusieurs conditions suffisantes de nucléarité. Ainsi, tout rôle qui doit nécessairement être exprimé, ou qui lorsqu'il n'est pas exprimé reçoit une interprétation définie, est noyau. Est aussi noyau tout rôle dont la sémantique ne peut être prédite à partir de la forme, ce qui signifie que les rôles exprimables sans marquage formel, ainsi que ceux dont le marquage formel est idiosyncratique, doivent être noyaux. Ces propriétés formelles sont suffisantes mais pas nécessaires pour déterminer la nucléarité d'un rôle : un rôle peut être noyau sans les posséder mais ne peut les posséder sans être noyau.

Lorsqu'ils ne sont pas noyaux, les rôles peuvent être *périphériques* ou *extra-thématiques*. Bien que nous présentions ici brièvement ces deux catégories, précisons par avance que la distinction entre elles n'est pas si claire. D'ailleurs, la visualisation des cadres proposée par le projet FrameNet sépare les rôles en deux groupes selon leur statut nucléaire ou non, mais n'indique pas lesquels des rôles non-noyaux sont périphériques et lesquels sont extra-thématiques [1].

Fillmore (2007) rapproche la distinction *noyau vs. périphérique* de celle opérée par Tesnière (1959) entre actants et circonstants. Les rôles noyaux relèvent du noyau de sens du cadre tandis que les rôles périphériques sont applicables à tout un ensemble de cadres relevant d'un même type sémantique (c'est-à-dire que le thème de tout évènement de type déplacement avance à une certaine vitesse, les cadres de type déplacement ont donc chacun un rôle périphérique *Speed*).

Contrairement aux rôles noyaux et périphériques, l'interprétation des rôles extra-thématiques ne dépend pas du cadre qui les contient, auquel ils n'appartiennent pas conceptuellement. Le rôle extra-thématique Beneficiary, présent dans 23 cadres, est décrit comme « le participant tirant un bénéfice de l'occurrence de l'évènement spécifié par le déclencheur du cadre » : cette définition est applicable à tout évènement. Ruppenhofer *et al.* (2006) indiquent que contrairement aux rôles périphériques — qui n'introduisent pas d'évènement autre que celui décrit par le cadre —, les rôles extra-thématiques placent l'évènement ou état caractérisé par leur cadre en relation avec un autre évènement ou état. Pour Ruppenhofer *et al.*, ces rôles seraient en fait ceux de cadres — existant ou non — plus abstraits, comme un cadre BENEFACTION, qui décrirait la relation entre une entité et

1. Le statut spécifique de chaque rôle non-noyau apparaît néanmoins dans la visualisation des annotations.

l'évènement décrit par un cadre dont cette entité bénéficierait.

Dans les faits, la distinction entre rôles périphériques et rôles extra-thématiques ne nous paraît pas très claire. Nous avons examiné une vingtaine de rôles non-noyaux nous paraissant relativement généraux, et seuls trois d'entre eux (Speed et Beneficiary, cités plus haut comme exemples, et Iteration, rôle extra-thématique de 31 cadres) partagent le même statut dans tous les cadres où ils apparaissent. Les autres sont parfois périphériques, parfois extra-thématiques (mais rarement en proportions égales). Le rôle Time est présenté par Ruppenhofer *et al.* (2006) comme exemple type du rôle périphérique (tout évènement a lieu à un moment donné, les cadres de type évènementiel auraient donc chacun un rôle périphérique Time) mais ce rôle apparaît dans 24 cadres comme extra-thématique. Pour nous, il est impossible d'établir une différence entre le rôle Time dans ces cadres et le rôle du même nom dans les 661 cadres où il apparaît comme périphérique. De même, le rôle Means — moyen employé par un agent pour effectuer l'action dénotée par le cadre — est périphérique dans 378 cadres et extra-thématique dans 2 : ACTIVITY_STOP, qui dénote un agent cessant une activité, et HALT, qui hérite de ce cadre. Il est en revanche périphérique dans ACTIVITY_START, où un agent se lance dans une activité, ACTIVITY_ONGOING, où l'agent continue à effectuer une activité, et tous les autres cadres décrivant la pratique d'une activité par un agent à un aspect différent. Il s'agit donc très manifestement d'une erreur pour les deux cas où ce rôle est marqué comme extra-thématique.

La table 3.1 présente la répartition des rôles de même nom apparaissant comme non-noyau dans plus de cent cadres. Il nous faut noter que comme nous l'avons mentionné plus haut, deux rôles portant le même nom n'ont pas nécessairement la même sémantique. Néanmoins, l'interprétation des rôles extra-thématiques ne dépendant pas du cadre qui les contient, nous considérons que ceux d'entre eux qui portent le même nom sont très probablement à interpréter de la même manière. De même pour les rôles périphériques, nous avons observé que les rôles portant le même nom au sein de cadres de même type semblaient être définis de la même manière. Nous considérons donc la table 3.1 comme significative. On y observe que certains rôles, comme Circumstances, sont très majoritairement extra-thématiques, tandis que d'autres, comme Manner, sont essentiellement périphériques. Cette asymétrie pointe qu'une distinction peut être effectuée entre rôles périphériques et rôles extra-thématiques. Cependant, le fait que parmi les rôles de la table, aucun n'apparaisse dans seulement un des deux statuts semble indiquer que cette distinction n'est pas aisée à effectuer.

3.2.1.3 Relations entre rôles (relations intra-cadres)

Au sein d'un cadre, les rôles peuvent être liés par trois types de relations, qui sont décrits ci-dessous.

Ensembles nucléaires

Il arrive que dans un cadre, on puisse identifier un groupe de rôles tel que la présence d'un des membres du groupe suffit à « satisfaire une valence sémantique du prédicat » (Ruppenhofer *et al.*, 2006). On dit alors que ces rôles sont un *ensemble nucléaire* (*Coreness set*, ou *CoreSet*).

Il peut arriver que, pour une même occurrence de cadre, plusieurs (voire l'intégralité) des rôles d'un ensemble nucléaire du cadre évoqué soient exprimés, mais cela n'est pas nécessaire, ni même fréquent. C'est ce qu'on peut voir en (15), où les rôles Instrument, Killer et Means, qui constituent un ensemble nucléaire du cadre KILLING, peuvent apparaître exprimés seuls ou en conjonction les uns avec les autres.

	nombre de rôles extra-thématiques	nombre de rôles périphériques	nombre total de rôles non noyaux
Time	24	661	685
Manner	6	568	574
Place	12	543	555
Degree	50	347	397
Means	2	378	380
Purpose	96	245	341
Depictive	231	3	234
Circumstances	220	5	225
Duration	57	165	222
Explanation	157	3	160
Reason	123	36	159
Frequency	145	12	157
Result	132	3	135
Instrument	8	102	110

TABLE 3.1: Répartition des rôles non-noyaux les plus nombreux

(15) a. Since someone tried to **KILL** him with a parcel bomb back in Lusaka, he's moved several times and today still goes in fear of his life.
« Depuis que quelqu'un a tenté de le TUER avec un colis piégé à Lusaka, il a déménagé plusieurs fois et a aujourd'hui encore peur pour sa vie. »

b. In Long (1830) 172 ER 756, a person who was not a doctor **KILLED** two patients by applying corrosive plasters to their chests.
« Dans l'affaire Long (1830) 172 ER 756, une personne qui n'était pas médecin a TUÉ deux patients en appliquant des cataplasmes corrosifs sur leurs poitrines. »

c. Agnes Rossnerin, who smothered and throttled her companion in order to steal some money, was **BEHEADED** by the sword "because she was a poor creature and had a wry neck". CNI
« Agnes Rossnerin, qui étouffa et étrangla son compagnon afin de voler de l'argent, fut DÉCAPITÉE à l'épée "parce qu'elle était une pauvre créature et avait un torticolis". »

KILLING: Cause, Instrument, Killer, Means, Victim

Exclusion mutuelle

Il arrive que dans un cadre, deux ou plusieurs rôles ne puissent apparaître ensemble pour une même occurrence du cadre auquel ils appartiennent. Cela est spécifié dans le cadre par une relation dite d'*exclusion*. On peut distinguer deux types de cas d'exclusion, selon qu'elle est liée à la syntaxe ou à la sémantique.

Dans le premier type de cas, l'exclusion signifie simplement qu'il est syntaxiquement impossible d'exprimer localement les rôles en relation d'exclusion. C'est le cas pour les cadres décrivant une relation symétrique entre deux actants : dans le cadre BE_IN_AGREEMENT_ON_ASSESSMENT par

exemple, les êtres partageant une même opinion peuvent être exprimés séparément, comme en (16a), ou conjointement, comme en (16b). Des rôles Cognizer_1 et Cognizer_2 sont donc utilisés pour le premier cas, alors que le deuxième s'annote avec le rôle Cognizers. Ces rôles décrivent un même type d'actant, et Cognizers est en relation d'exclusion avec Cognizer_1 et Cognizer_2.

(16) a. I very much **AGREE** with my hon. Friend.
« Je suis tout à fait D'ACCORD avec mon honorable collègue. »

b. The ministers **AGREED** on the urgency of the situation.
« Les ministres étaient D'ACCORD sur l'urgence de la situation. »

BE_IN_AGREEMENT_ON_ASSESSMENT : Cognizer_1, Cognizer_2, Cognizers

Dans le second type de cas, l'exclusion indique une incompatibilité sémantique entre deux types d'éléments jouant des rôles possédant des similarités. C'est le cas pour beaucoup de cadres décrivant un évènement pouvant être déclenché par un agent (être sensible et intentionnel) ou une cause (situation), comme le cadre KILLING. On voit en (17) que dans ce cadre, Cause et Killer, deux rôles en relation d'exclusion, permettent d'annoter ces deux cas de figure.

(17) a. The fiery blast **KILLED** everyone on deck instantly, with the single exception of the captain, who lived on for a short time before becoming unconscious and falling overboard.
« L'explosion fulgurante TUA instantanément toutes les personnes qui étaient sur le pont, à la seule exception du capitaine, qui survécut un moment avant de perdre connaissance et de tomber par dessus bord. »

b. She would **KILL** him later, she thought, when she was strong again.
« Elle le TUERAIT plus tard, pensa-t-elle, quand elle aurait repris des forces. »

KILLING: Cause, Instrument, Killer, Means, Victim

Nécessité

Il arrive que dans un cadre, la présence d'un rôle exprimé rende nécessaire la présence d'un autre rôle. Cette relation est spécifiée dans le cadre sous le nom de *nécessité* (*requires*). C'est le cas pour les rôles Cognizer_1 et Cognizer_2 du cadre BE_IN_AGREEMENT_ON_ASSESSMENT présenté en (16) qui se nécessitent l'un l'autre, ainsi que pour les rôles équivalents de tous les autres cadres pouvant être déclenchés par des verbes permettant une construction symétrique.

Comme le montre l'exemple des rôles Protagonist et Practice du cadre MENTAL_PROPERTY, cette relation n'est pas symétrique. Ce cadre décrit une propriété mentale d'un être sensible (le Protagonist) ou d'un comportement (le Behavior) de ce protagoniste. Dans certains cas, ladite propriété du protagoniste n'est valable qu'appliquée à un domaine ou capacité : le Practice. Lorsque le rôle Practice est réalisé (exemple (18a)), le rôle Protagonist doit être interprétable. On dit donc que Practice nécessite Protagonist. En revanche, comme on le voit en (18b), le Protagonist peut être exprimé sans Practice, et ne le nécessite donc pas.

(18) a. You knew he was **SMART** about cars; he was savvy about life too.
« Tu savais qu'il était MALIN pour les voitures ; il était astucieux dans la vie aussi. »

b. People are too **SMART** now.
« Les gens sont trop INTELLIGENTS de nos jours. »

MENTAL_PROPERTY : Protagonist, Practice

3.2.1.4 Relations inter-cadres

FrameNet définit plusieurs relations pouvant lier deux cadres entre eux. Ces relations sont dirigées (asymétriques) et présentent deux grands intérêts. Le premier est qu'elles permettent de modéliser des relations conceptuelles, telles que la subsomption, la précédence temporelle ou la causalité, entre les situations prototypiques décrites par les cadres. Le second est que comme elles impliquent d'établir des liens d'identité entre les rôles des deux cadres, ces relations permettent de généraliser sur les rôles. Cela rapproche les rôles FrameNet des autres conceptions des rôles sémantiques présentées au chapitre précédent section 2.1.

Baker (2009) regroupe les relations entre cadres définies par FrameNet en trois types : les relations de généralisation, les relations de structure évènementielle et les relations aspectuelles. Nous reprenons donc cette classification pour présenter rapidement les caractéristiques de chacune des relations.

Relations de généralisation

Baker (2009) regroupe sous ce nom les relations pour lesquelles tout ou partie des rôles du cadre enfant sont du même type que ceux du cadre parent. Baker explique qu'au plus haut du graphe orienté acyclique formé par les cadres et les relations de généralisation qui les lient les uns aux autres se trouvent des cadres très généraux. Nous changeons ici l'exemple utilisé par Baker car nous verrons un peu plus bas qu'il est problématique. Le cadre STATE est un de ces cadres très généraux, il correspond aux cas où une entité persiste dans une situation stable appelée un état. Ce cadre possède un rôle noyau, le rôle Entity, qui correspond à l'entité se trouvant dans l'état en question. Tout cadre héritant de STATE possède un équivalent de ce rôle, puisque tout état est subi par une entité. C'est ce lien entre rôles qui permet de rapprocher FrameNet des visions des rôles sémantiques possédant un inventaire de rôles plus réduit.

L'*héritage* (*inheritance*) est une relation de subsomption. Le cadre enfant modélise une version plus spécifique du concept modélisé par le cadre parent. Il s'agit de la relation la plus formellement contraignante : tout ce qui est strictement vrai de la sémantique du parent doit l'être aussi de celle de l'enfant. Ainsi, l'enfant hérite de tous les rôles (sauf extra-thématiques) du parent. Un rôle hérité doit être d'un type sémantique compatible avec son rôle parent, c'est-à-dire un type sémantique identique ou plus précis. L'héritage multiple est autorisé.

La définition FrameNet de la relation de *perspective* ne nous paraît pas très claire ni satisfaisante. Il est dit dans la littérature que cette relation lie un cadre neutre (le plus souvent non lexical) à un cadre référant à la même situation décrite d'un point de vue donné (celui d'un des participants). Prenons pour exemple le cadre RESPONSE_SCENARIO, dans lequel une réaction est fournie par une entité suite à un stimulus. Ce cadre est perspectivisé par MEET_WITH_RESPONSE dans lequel un stimulus provoque une réaction chez une entité, et RESPOND, dans lequel une entité réagit à un stimulus. Ces cadres décrivent la même situation en profilant (mettant en avant) respectivement le stimulus et l'entité réactive. Une situation pouvant être décrite par un de ces deux cadres perspectivisants pourrait aussi l'être par l'autre. Ruppenhofer *et al.* (2006) pointent que cette relation indique la présence d'au moins deux points de vue différents sur le cadre neutre, mais l'observation des données met en évidence 16 cadres liés à un unique cadre les perspectivisant (contre 38 perspectivisés par au moins deux cadres).

Cette relation pose pour nous le problème que la notion de *point de vue*, ou de *profilage* (mise en avant) d'un participant est à notre sens très floue. Nous n'avons rien trouvé dans la littérature FrameNet qui précise cette notion, qui semble implicitement faire référence à la syntaxe. En effet, les prédicats signalés comme introduisant une perspective différente sur une situation présentent généralement un linking canonique différent : *acheter* et *vendre* font tous deux référence à une transaction commerciale, mais le sujet canonique d'*acheter* est l'acheteur tandis que celui de *vendre* est le vendeur. Nous nous penchons plus avant sur les problèmes que posent selon nous cette

relation dans le chapitre 4, section 4.2.4.1.

On dit qu'un cadre en utilise un autre lorsque le cadre utilisateur présuppose le cadre utilisé comme arrière-plan, c'est-à-dire que la compréhension du cadre utilisateur demande la connaissance et compréhension du cadre utilisé. Bien que Baker (2009) cite l'*usage* (*using*) comme une relation de généralisation, le cadre utilisateur et le cadre utilisé ne réfèrent pas à la même situation. Ruppenhofer *et al.* (2006) expliquent qu'une partie de la scène évoquée par le cadre utilisateur réfère au cadre utilisé.

Cette relation couvre une variété de situations, entre autres : des cas où le cadre utilisé correspond à un participant du cadre qui l'utilise (exemple (19)) ou une méthode employée dans ce cadre (exemple (20)) ; des cas où le cadre utilisateur n'a de sens que parce que le concept décrit dans le cadre utilisé existe (exemple (21)) ; et des cas où le cadre utilisateur n'a de sens que si la situation décrite par le cadre utilisé a lieu, avec les mêmes participants (exemple (22)).

(19) a. INTOXICATION (une substance affecte une personne) utilise INTOXICANTS (substance utilisée pour s'enivrer)

b. HAIR_CONFIGURATION (coupe, coiffure ou nature de cheveux) utilise OBSERVABLE_BODY_PARTS (parties du corps)

(20) ATTEMPT_SUASION (un locuteur tente de convaincre son interlocuteur de faire quelque chose) utilise COMMUNICATION (un locuteur adresse un message à un interlocuteur)

(21) a. INSTITUTIONALIZATION (un patient est interné) utilise CURE (un soignant traite ou guérit un patient)

b. OBSCURITY (une entité est peu connue) utilise FAME (une entité est célèbre)

(22) a. LOCATING (un agent trouve un objet qu'il cherchait) utilise SEEKING (un agent cherche un objet)

b. PURPOSE (un agent cherche à atteindre un but) utilise DESIRING (un expérienceur souhaite l'occurrence d'un évènement)

Ces cas de figure correspondent tous à la définition de la relation énoncée supra. Néanmoins, il existe aussi des cas pour lesquels la raison du choix d'une relation d'usage plutôt que d'une autre relation n'est pas clair. C'est le cas par exemple pour les cadres KINSHIP (liens de parenté) et PERSONAL_RELATIONSHIP (relations amoureuses). Ces derniers utilisent tous les deux le cadre RELATION_BETWEEN_INDIVIDUALS (deux êtres sensibles ont une relation) alors que rien ne semble empêcher qu'ils n'en héritent. Ovchinnikova *et al.* (2010) pointent d'ailleurs que dans la plupart des cas où une relation d'usage a été définie entre deux cadres, une autre des relations proposées par FrameNet aurait pu convenir. On peut supposer qu'il s'agit là d'erreurs, ou des restes d'une pratique antérieure à la création d'autres relations.

La figure 3.2 montre un extrait du réseau de relations entre quelques cadres ayant trait aux transactions commerciales, et présente des exemples de chacune des trois relations de généralisation que nous venons d'évoquer.

Relations de structure évènementielle

Un cadre décrivant un évènement complexe peut être décomposé en sous-cadres, qui entretiennent avec lui une relation appelée *sous-cadre* (*subframe*). Ainsi, EMPLOYEE'S_SCENARIO, cadre non lexical décrivant l'ensemble d'une relation de travail entre un employeur et un employé du point de vue de cet employé, a pour sous-cadres GET_A_JOB, BEING_EMPLOYED et QUITTING. Ces derniers décrivent respectivement les situations où un employé décroche un travail, travaille pour un employeur et démissionne. Il s'agit d'un scénario typique, mais pas forcément toujours réalisé dans son intégralité : GET_A_JOB est nécessairement suivi de BEING_EMPLOYED, mais ce dernier n'implique pas nécessairement que QUITTING s'ensuive.

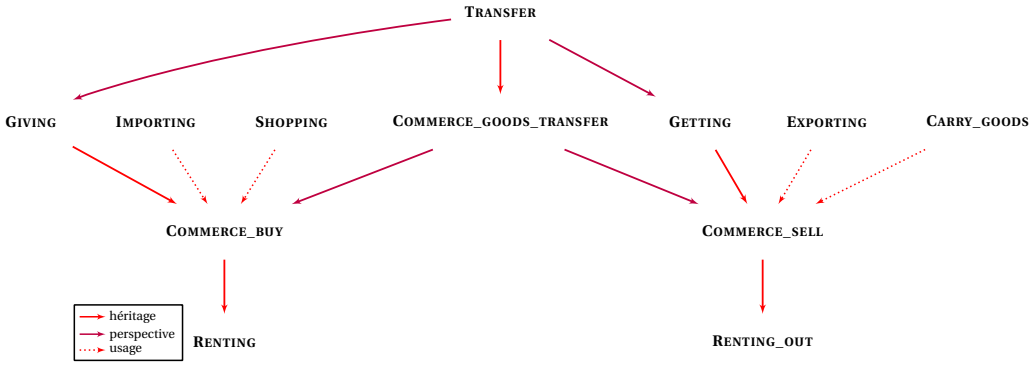


FIGURE 3.2: Exemple de relations de généralisation : les transactions commerciales

Les sous-cadres peuvent être liés par une relation de *précédence* (*precedes*), qui définit leur ordonnancement chronologique s'il existe. C'est le cas des susmentionnés GET_A_JOB, BEING_EMPLOYED et QUITTING, qui se succèdent dans cet ordre.

Cette relation n'accompagne pas systématiquement la relation de sous-cadre : certains évènements complexes donnent lieu à plusieurs sous-cadres pour lesquels on ne peut établir d'ordre chronologique. C'est le cas du cadre MOTION, qui décrit un déplacement d'une entité. Ses deux sous-cadres, GETTING_UNDERWAY (une entité quitte un lieu source) et HALT (une entité cesse son mouvement), peuvent se succéder dans n'importe quel ordre. De même, un cadre décrivant un évènement complexe peut donner lieu à des sous-cadres correspondant à plusieurs issues de l'évènement : INVASION_SCENARIO a pour sous-cadres INVADING, CONQUERING et REPEL. INVADING est en relation de précédence avec les deux autres cadres, puisqu'un envahisseur conquiert le terrain envahi ou en est repoussé, mais CONQUERING et REPEL ne sont liés par aucune relation puisqu'ils n'apparaissent pas dans la même réalité.

Il est aussi possible pour deux cadres d'être ordonnés par une relation de précédence sans être sous-cadres d'un même cadre. C'est par exemple le cas de DYING, qui décrit l'état d'un protagoniste approchant la mort et DEATH, évènement de la mort d'un protagoniste.

La figure 3.3 montre un extrait du réseau de relations entre quelques cadres, et présente des exemples des deux relations de structure évènementielle que nous venons d'évoquer.

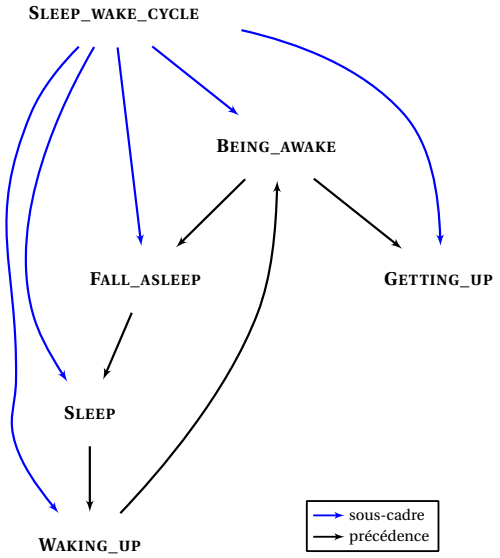


FIGURE 3.3: Exemples de relations de structure évènementielle : le sommeil et l'éveil

Relations aspectuelles

Deux cadres peuvent aussi entretenir une relation aspectuelle l'un avec l'autre. FrameNet spécifie ainsi la causativité et l'inchoativité. Le cadre vers lequel une de ces relations pointe est généralement statif.

La relation *causative* (*causative of*) lie deux cadres dont l'état décrit par l'un est la conséquence de la situation décrite par l'autre. Un exemple en est le cadre KILLING (une personne est tuée), causatif du cadre DEATH (une personne est morte), avec identité du rôle Victim de KILLING (la personne tuée) avec le rôle Protagonist de DEATH (la personne morte).

La relation *inchoative* (*inchoative of*) lie un cadre statif A à un cadre B décrivant le changement d'état amenant à l'état décrit par A. Ainsi, le cadre FORMING_RELATIONSHIP, dans lequel deux humains interagissent pour changer leur relation sociale, est inchoatif du cadre PERSONAL_RELATIONSHIP, qui décrit une variété de relations inter-personnelles.

3.2.2 Annotations

Nous l'avons brièvement évoqué dans l'introduction du chapitre, les annotations du projet FrameNet ont été effectuées en deux fois. FrameNet présente donc un ensemble d'annotations lexicographiques (plus de 86% du total des 173 018 annotations, la répartition exacte est représentée table 3.2) et un ensemble d'annotations de plein texte, dans lesquelles chaque mot prédicatif de chaque phrase de texte continu est censé avoir été annoté avec un cadre pertinent. Ces deux types d'annotation utilisent deux méthodologies bien différentes.

Dans la littérature du projet FrameNet, on retrouve souvent le terme d'*annotation sets*, ou *ensembles d'annotations*. Il s'agit en fait d'occurrences de cadre annotées, mais l'idée d'*ensemble* fait référence à l'existence de plusieurs « couches » d'annotation. Le coeur du projet est de proposer une annotation sémantique des déclencheurs (paires lemme-cadre déclenché) et de leurs remplisseurs de rôles (paires constituant-rôle rempli). De plus, comme FrameNet ne s'appuie pas sur un corpus préalablement annoté en morphologie et en syntaxe, les informations sur les catégories et fonctions syntaxiques des constituants sur lesquels porte cette annotation sémantique ont été annotées au même moment que ces annotations de cadres et rôles.

Pour plus de concision, nous parlerons simplement d'annotations pour référer à ces ensembles. Ainsi, il faut lire *une annotation* comme *une instance de cadre annoté*.

Les annotations lexicographiques (exemple figure 3.4) ont été réalisées au fur et à mesure de la création de cadres et de l'association de déclencheurs à ces cadres (voir section 3.3.1). Les annotations portent sur des phrases isolées comportant le déclencheur souhaité, choisies dans le BNC pour leur variété en termes de réalisations des rôles. Pour chaque déclencheur de chaque cadre, les annotateurs et annotatrices ont donc cherché à annoter en cadre et en rôles un petit nombre d'exemples de chaque réalisation syntaxique possible des différents rôles. Ces annotations sont accessibles depuis les descriptions des cadres, et les informations regroupées dans les entrées

	Nombre d'instances de cadres annotés
Annotation lexicographique	149 931
Annotation de plein texte	23 087
Total	173 018

TABLE 3.2: Répartition des instances de cadres annotés par type d'annotation

lexicales proviennent uniquement de ce type d'annotations. La version 1.5 de FrameNet comporte 149 931 annotations lexicographiques.

Les annotations dites full-text (exemple figure 3.5) sont venues s'ajouter aux annotations lexicographiques à partir de 2004 (d'après Lönneker-Rodman et Baker (2009, p. 419)). Pour ce type d'annotation, les annotateurs et annotatrices du projet FrameNet ont parcouru des textes continus en étiquetant de cadres tous les mots prédicatifs qu'elles et ils ont rencontrés. Cette annotation systématique en cadres et en rôles a nécessité l'ajout de nouveaux déclencheurs à des cadres préexistants, ainsi que, le cas échéant, la création de nouveaux cadres. Les annotations de plein texte présentent l'intérêt de rendre compte de la réelle distribution en corpus des déclencheurs, ainsi que celle des réalisations syntaxiques des remplisseurs de rôles. Cela les rend plus utiles pour l'apprentissage automatique que les annotations lexicographiques. Das *et al.* (2010) rapportent d'ailleurs qu'entraînés sur les annotations lexicographiques, leurs modèles ont des performances plus basses que lorsqu'ils sont entraînés sur les annotations de plein texte, et ce alors même que ces dernières sont en nombre bien plus réduit que les annotations lexicographiques. La version 1.5 de FrameNet comporte 23 087 annotations de plein texte.

3.2.3 Lexique

Chaque déclencheur possède une *entrée lexicale*, extraite de l'ensemble des annotations lexicographiques, comprenant en particulier des informations sur l'interface syntaxe/sémantique de l'unité lexicale. Ces entrées présentent une définition simple du déclencheur, souvent tirée du *Concise Oxford Dictionary*, ainsi que deux types d'informations : une liste de *réalisations des rôles* et une liste des *patrons de valence*.



FIGURE 3.4: Extrait des annotations lexicographiques du déclencheur *kill* (verbe) du cadre KILLING

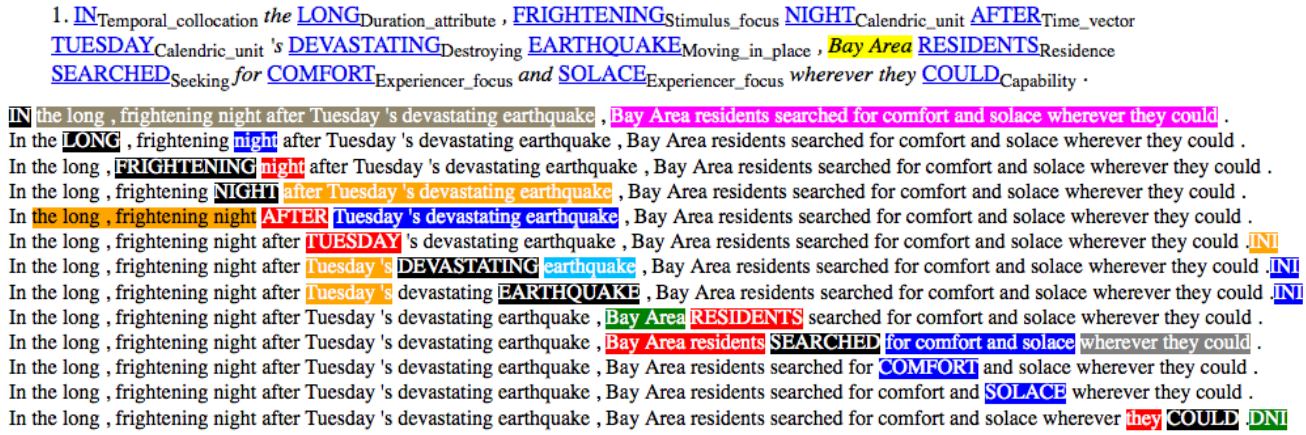


FIGURE 3.5: Extrait des annotations de plein texte

La liste des réalisations des rôles (représentée figure 3.6) énumère pour chaque rôle du cadre déclenché par l'unité lexicale :

— en colonne centrale : le nombre de phrases-exemples de ce déclencheur dans lesquelles le rôle en question est rempli. Cliquer sur ce nombre permet d'accéder à la visualisation de l'ensemble de ces phrases annotées ;

— en colonne de droite : les réalisations syntaxiques possibles, sous forme de concaténation `type de constituant.fonction grammaticale`, ainsi que le nombre d'exemples dans lesquels le rôle est réalisé de cette manière. Cliquer sur ce nombre permet aussi de visualiser l'ensemble de ces annotations.

La liste des patrons de valence observés (représentée figure 3.7) présente les agencements de rôles réalisés ensemble dans les exemples annotés, ainsi que pour chacun de ces agencements les combinaisons existantes de réalisations des rôles. Comme avec la liste des réalisations vue plus haut, les types de constituants et fonctions grammaticales sont regroupés, et la table permet d'accéder directement aux annotations.

3.3 Méthodologie du projet FrameNet

Il est avant tout nécessaire de souligner que la stratégie générale du projet est de travailler cadre à cadre plutôt que mot à mot comme le font bon nombre d'entreprises lexicographiques, par exemple.

Frame Element	Number annotated	Realization(s)
Cause	(8)	CNI.– (1) NP.Ext (2) PP[by].Dep (1) PP[in].Dep(4)
Killer	(33)	NP.Ext (24) CNI.– (3) PP[by].Dep (6)
Victim	(41)	NP.Obj (27) NP.Ext (14)

FIGURE 3.6: Entrée lexicale du déclencheur *kill* (verbe) du cadre KILLING : extrait des réalisations possibles des rôles

Number annotated	Patterns		
1 TOTAL	Beneficiary	Killer	Victim
(1)	PP[for] Dep	NP Ext	NP Obj
6 TOTAL	Cause	Victim	
(1)	NP Ext	NP Obj	
(1)	PP[by] Dep	NP Ext	
(4)	PP[in] Dep	NP Ext	
3 TOTAL	Instrument	Killer	Victim
(3)	PP[with] Dep	NP Ext	NP Obj

FIGURE 3.7: Entrée lexicale du déclencheur *kill* (verbe) du cadre KILLING : extrait de liste de patrons de valence

On définit des cadres pour des groupes de lexèmes sans se pencher sur chacun pour expliciter ses autres sens possibles et les associer à leurs propres cadres. Tant que sa couverture n'est pas totale, une ressource ainsi développée rend donc compte des relations de paraphrases entre les lexèmes mais, sauf coïncidence, pas de leur potentielle polysémie.

Pour Fillmore (2007), cette stratégie présente deux intérêts principaux. D'abord, le fait d'examiner les occurrences en corpus de tout un groupe de lexèmes en parallèle peut permettre de repérer plusieurs blocs au sein d'un ensemble de lexèmes que l'on jugeait sémantiquement proches. Ensuite, d'un point de vue pratique, comme les mots sémantiquement proches et de même catégorie grammaticale ont souvent un comportement syntaxique similaire, pour de nombreux déclencheurs d'un même cadre, les mêmes règles peuvent être utilisées pour rassembler des exemples en corpus afin de les annoter (cf. 3.3.1).

L'auteur explique la nécessité de choisir entre cette approche et une approche mot à mot plus traditionnelle par l'inopérabilité d'une méthode qui combinerait les deux buts. Partant d'un lexème, il faudrait créer un cadre pour chacun de ses sens, puis peupler chacun de ces cadres d'autres déclencheurs, déclencheurs dont il faudrait ensuite examiner tous les sens pour les intégrer dans des cadres qui auraient à leur tour de nouveaux déclencheurs, et ainsi de suite. Il souligne que l'important pour le projet FrameNet est d'explorer les déclencheurs d'un même cadre et de pouvoir marquer les relations de paraphrase. Pour lui, tenter d'explorer chacun des sens des mots étudiés ne serait pas une manière rationnelle de progresser.

3.3.1 Organisation du travail

La littérature FrameNet s'accorde à définir plusieurs étapes au processus de création de la ressource. La première est un travail préparatoire, consistant à définir grossièrement des cadres sous forme d'une définition, une liste de participants et accessoires, et une liste de déclencheurs. Des exemples en corpus sont observés pour repérer des patrons syntaxiques. Ces derniers permettent d'extraire les sous-corpus d'exemples, qui sont ensuite annotés. Ces étapes sont ici décrites de manière un peu plus détaillée.

Lönneker-Rodman et Baker (2009) indiquent que la *délimitation* des cadres s'appuie à la fois sur l'introspection et sur les données du corpus. Les travaux de Johnson *et al.* (2001) et Ruppenhofer *et al.* (2006) permettent de préciser que le travail consiste à établir des listes de mots de sens similaire et d'observer leurs attestations en corpus pour tenter de les regrouper en cadres.

Lönneker-Rodman et Baker (2009) évoquent l'existence de domaines sémantiques dans lesquels s'inscriraient les premiers cadres créés, mais ne citent pas ces domaines, et dans la littérature, aucune autre description du processus de travail n'y fait référence. L'examen de la littérature laisse apparaître que des domaines ont en effet été utilisés pendant la première phase du projet, sans réellement expliquer comment cela s'intègre au travail de création de cadres. Baker *et al.* (1998) indiquent simplement leur intention de couvrir les domaines sémantiques de la santé, le hasard, la perception, la communication, les transactions, le temps, l'espace, le corps, les mouvements, les étapes de la vie, la société, les émotions et la cognition. On retrouve tous ces domaines sauf le hasard dans la description des cadres de chaque domaine de Johnson *et al.* (2001), qui indiquent que c'est pour des raisons pratiques que les cadres sont organisés par domaines (sans préciser ces raisons). Ils expliquent que ces domaines correspondent à des catégories très larges d'expérience et de connaissance humaine, et que bien qu'ils constituent un degré de généralisation sur un ensemble de cadres, leur importance théorique pour caractériser les unités lexicales est faible et indirecte. Baker *et al.* (2003) expliquent que la base théorique des domaines proposés n'a jamais été explicitée, et que les relations entre cadres, plus précises, sont venues remplacer ce type de généralisation large. Johnson *et al.* (2003) précisent que c'est la relation *using* qui capture le concept de domaines de FrameNet I.

D'après ces différents travaux, il semblerait que ces domaines aient au départ été utilisés comme base pour créer des cadres, dans le but de voir si la sémantique des cadres permettrait de rendre compte de leur diversité. Aux cadres ainsi créés ont ensuite été ajoutés des cadres en lien sémantique avec eux, ou permettant de couvrir d'autres sens des lexèmes de ces cadres. Lönneker-Rodman et Baker (2009) indiquent en effet que l'appariement d'un déclencheur avec un cadre nécessite de repérer les autres sens du lexème-déclencheur et de voir s'ils appartiennent à des cadres qui existent déjà ou s'il faut créer de nouveaux cadres, mais ne précisent pas si ces nouveaux cadres sont systématiquement créés, et nous n'avons trouvé nulle part ailleurs de mention de cet écart à la méthodologie cadre à cadre décrite en 3.3. De plus, Lönneker-Rodman et Baker étant seuls à faire état de l'utilisation des domaines pour la création de cadres, et la littérature plus récente ne faisant plus aucune mention desdits domaines, nous supposons que la manière de choisir les groupes de lexèmes sur lesquels travailler n'a pas été constante ni systématique.

Une fois le groupe de déclencheurs identifié intervient la *définition* du cadre et de ses rôles. Pour la définition des rôles, les membres du projet FrameNet ont observé des exemples en corpus afin de repérer les éléments qui semblaient apparaître avec le déclencheur. Des cadres plus abstraits (parfois non-lexicaux), ainsi que des relations entre cadres, ont été définis au fur à mesure des créations de cadres pour permettre de capturer des généralisations linguistiques et conceptuelles.

Pour chaque déclencheur, les membres du projet FrameNet ont ensuite à nouveau observé des occurrences en corpus et fait preuve d'introspection afin de trouver les principaux *patrons syntaxiques* associés au déclencheur. Ils se sont servi de ces patrons pour écrire des règles avec lesquelles ils ont extrait du BNC un échantillon de phrases exemples illustrant les constructions repérées.

Cette extraction terminée, les annotateurs et annotatrices du projet FrameNet ont pu *annoter* assez d'exemples en cadre et en rôles pour illustrer les réalisations syntaxiques « lexicographiquement pertinentes » de chaque rôle [2]. Les pratiques employées pour ce travail d'annotation sont décrites section 3.3.3.

Enfin, les annotations ont été utilisées pour *générer les données visibles*. Il s'agit ici d'obtenir les informations sur les entrées lexicales décrites section 3.2.3, ainsi que d'ajouter à ces entrées lexicales et aux déclencheurs listés dans les cadres une visualisation des annotations qui leur sont pertinentes.

2. Lönneker-Rodman et Baker (2009) précisent que les alternances valables pour une très grande proportion du lexique, comme le passif pour les verbes transitifs, ne sont pas considérées comme pertinentes.

Le processus de création décrit ci-dessus concerne uniquement la partie *annotation lexicographique* du projet. Un mode de travail différent a été utilisé pour la production des annotations en plein texte. Pour cette partie du projet, les annotateurs et annotatrices ont parcouru des textes continus en étiquetant de cadres toutes les expressions les évoquant qu'elles et ils ont rencontrées. Contrairement à ce qui a été fait pour l'annotation lexicographique, c'est donc au fur et à mesure de l'annotation et en fonction des besoins que de nouveaux déclencheurs ont été ajoutés à des cadres existants et que de nouveaux cadres ont été créés.

Les textes continus annotés avec ce type de fonctionnement sont divers : des extraits de PropBank [3], des textes de questions-réponses à thème militaire du projet AQUAINT [4] (Advanced Question & Answering for Intelligence project), et plus tard, des textes de l'American National Corpus MASC project [5]. D'autres documents très variés (dépêches, e-mails, traductions de l'arabe et transcriptions de conversations téléphoniques) ont aussi été annotés dans le cadre de l'annotation de plein texte.

Lönneker-Rodman et Baker (2009) pointent que l'annotation en plein texte oblige à annoter des phrases contenant des difficultés syntaxiques, chose que l'annotation lexicographique permettait d'éviter, ce que Ruppenhofer *et al.* (2006) explicitent :

> « Dans l'annotation de plein texte, nous sommes obligés d'annoter chaque instance de déclencheur, qu'elle apporte ou non des informations lexicographiques utiles sur ce déclencheur. Par exemple, dans un texte donné le déclencheur peut apparaître dans des constructions syntaxiques qui n'ajoutent rien à notre conception du sens du déclencheur. Un emploi particulier du lexème pourrait être dans un style pauvre ou quasi agrammatical. [...] Cela contraste avec notre travail d'annotation lexicographique, où éviter de telles phrases fait intentionnellement partie de nos principes d'annotation. » [6] [traduction] (Ruppenhofer *et al.*, 2006, p. 17)

De notre point de vue, ce parti pris est problématique. Sélectionner les phrases à annoter pour leur simplicité et exemplarité nous semble une simplification injustifiée de la réalité. De plus, d'un point de vue pratique, un analyseur sémantique en cadres et rôles devra nécessairement se confronter à ces difficultés, il semble donc illusoire de vouloir les éviter.

3.3.2 Création et délimitation des cadres

Nous avons vu section 3.3.1 que les cadres étaient délimités et définis à partir de l'observation de données en corpus. La procédure consiste en l'examen d'exemples en corpus de mots qui paraissent avoir un noyau de sens commun, puis la constitution de groupes de ces mots qui formeront le lexique d'un cadre.

Nous avons aussi vu que Fillmore (2007) opposait cette méthode de travail cadre à cadre (au départ groupe de mots à groupe de mots) avec une approche différente qui consisterait à examiner chaque mot pour en définir tous les sens. Pour Fillmore, il est impossible de partir dans les deux directions à la fois car cela ferait croître trop rapidement le nombre de cadres, sur lesquels un travail simultané deviendrait extrêmement difficile. FrameNet a donc choisi de se concentrer dans un premier temps sur une méthode cadre à cadre.

3. `http://verbs.colorado.edu/~mpalmer/projects/ace.html`
4. `http://www-nlpir.nist.gov/projects/aquaint`
5. `http://www.anc.org/data/masc/`
6. « In full-text annotation, we have no choice but to annotate each instance of a target word whether or not it contributes useful lexicographic information about the target. For instance, in a given text the LU may occur in complicated syntactic constructions that add nothing to our understanding of the target's meaning. A particular use of the LU may be very poor style or borderline ungrammatical. Some instances only provide more exemplification of an already amply documented argument realization pattern for the LU rather than exhibiting a rarer pattern. This contrasts with our sentence-annotation (lexicographic) work, where it is intentionally part of our annotation principles to avoid such sentences. » (Ruppenhofer *et al.*, 2006, p. 17)

La première partie de la section qui suit sera consacrée à la description des critères pertinents au regroupement de lexèmes dans des cadres recensés par Ruppenhofer *et al.* (2006). Une deuxième partie se penchera ensuite brièvement sur le traitement accordé à la polysémie dans le projet FrameNet, c'est-à-dire plus précisément sur les critères utilisés pour séparer les mots polysémiques en plusieurs cadres.

3.3.2.1 Critères de regroupement des lexèmes en cadres

Les contraintes sur la définition des cadres sont à la fois très fortes et assez peu détaillées dans la littérature FrameNet. Comme c'est le cas pour les notions de cadre, rôle, etc. (cf. section 3.2.1), les critères avancés sont avant tout sémantiques — bien que la syntaxe semble pour certains implicitement sous-jacente — et semblent parfois difficiles à mettre en œuvre. Pour le French FrameNet, nous verrons section 4.2.4 que avons tâché de clarifier la présence de la syntaxe dans les critères de délimitation pour rendre ces derniers plus faciles à appliquer.

Tout d'abord, les déclencheurs d'un cadre doivent avoir le même *nombre et type de rôles*, à la fois dans les concepts explicites et implicites. Pour ce qui est du nombre, cela signifie concrètement que les déclencheurs doivent avoir le même nombre de ce que Ruppenhofer *et al.* (2006, p.9) appellent rôles « essentiels et syntaxiquement proéminents » [traduction], et que nous interprétons comme rôles noyaux. En ce qui concerne le typage des rôles, Ruppenhofer *et al.* (2006) considèrent qu'un même rôle devrait n'être rempli que par des constituants d'un même type sémantique, et qu'une variation de type entre les réalisations est un critère pour créer deux rôles plutôt qu'un seul. Ces rôles seront alors en relation d'ensemble nucléaire ou d'exclusion (cf. section 3.2.1.3) selon s'ils peuvent ou non être réalisés conjointement.

Les déclencheurs d'un cadre aspectuellement complexe doivent *partager leur ensemble d'implications* pour tout ce qui est des *étapes et des transitions* qui les constituent. Ruppenhofer *et al.* (2006) donnent l'exemple des verbes *work on* et *develop*, qui évoquent tous les deux une idée d'agent qui dépense temps et énergie dans le but d'accomplir un but, mais dont seul *develop* implique que le but a été atteint. Ces verbes doivent donc évoquer deux cadres différents.

Pour Ruppenhofer *et al.* (2006, p. 10), les déclencheurs d'un cadre doivent « mettre l'accent sur le point de vue du même participant » de la situation décrite par le cadre. C'est ce que les auteurs appellent *profilage du même rôle* par tous les déclencheurs d'un cadre. Ils donnent l'exemple des verbes *acheter* et *vendre*, qui décrivent tous deux une transaction commerciale : la perspective du premier est celle de l'acheteur tandis que celle du second est celle du vendeur. Si deux déclencheurs potentiels examinés décrivent exactement la même situation mais en focalisant sur un participant différent, comme c'est le cas pour *acheter* et *vendre*, cela doit donner lieu à la création de deux cadres, tous deux en relation de perspective avec un même super-cadre. Nous avons mentionné en 3.2.1.4 que la notion de « profilage du point de vue d'un participant » nous paraissait vague, et nous semblait implicitement s'appuyer sur la syntaxe. Nous développerons ce point plus avant au chapitre suivant, section 4.2.4.1.

Les déclencheurs d'un cadre doivent partager *les mêmes relations entre rôles*. Il ne semble pas s'agir là uniquement des relations définies en 3.2.1.4 mais aussi de relations moins formellement définies. En effet, Ruppenhofer *et al.* (2006) donnent pour exemple de cette règle le fait qu'un rôle But associé à *acheter* ou à *vendre* n'aurait pas le même sens : il exprimerait respectivement le but de l'Acheteur et celui du Vendeur.

Les déclencheurs d'un cadre doivent partager *les mêmes présuppositions, attentes et concomitants*. Ruppenhofer *et al.* (2006) donnent l'exemple de *receive* (*recevoir*) et *take* (*prendre*), qui déclenchent des cadres différents puisque *receive* présuppose un Donneur consentant alors que ce n'est pas le cas de *take*. Cette règle ne semble pas avoir toujours été suivie : on retrouve dans le lexique de certains cadres des lexèmes déclenchant une présupposition que d'autres ne déclenchent pas. C'est par exemple le cas des verbes *know* (*savoir*) et *think* (*penser*), déclencheurs du cadre

AWARENESS : *know* présuppose la véridicité du contenu mental connu du cogniteur, tandis que ce n'est pas le cas pour *think*. La définition du cadre précise néanmoins que cet état de fait devrait être corrigé.

Les déclencheurs d'un cadre doivent avoir une *dénotation grossièrement similaire*. L'intérêt de ce critère n'est pas très clair : Ruppenhofer *et al.* (2006) donnent l'exemple des adjectifs *broken* et *blue*, qui bien que décrivant tous deux des états ne devraient pas évoquer le même cadre puisqu'il s'agit d'états très différents l'un de l'autre. Il nous semble néanmoins que de tels adjectifs seraient de toute façon séparés en cadres différents par le critère précédent. En effet, dire qu'un objet est cassé présuppose qu'il a été entier, tandis que dire qu'un objet est bleu ne présuppose rien de tel. Par ailleurs, on peut constater des disparités de niveau de granularité sémantique entre les cadres, ne serait-ce qu'entre les cadres relativement hauts vs. bas dans la hiérarchie des relations cadre à cadre, ce qui tend à rendre ce critère peu évident à appliquer.

Ruppenhofer *et al.* (2006) ajoutent à ces critères celui selon lequel les pré-spécifications que les déclencheurs imposent à leurs rôles doivent être partagées au sein d'un cadre. Néanmoins, ils précisent que ce critère est en pratique souvent relâché car il conduirait à la création de cadres à la granularité trop fine pour être gérés facilement. Il nous semble par ailleurs que ce critère est proche de celui sur les présuppositions portées par les déclencheurs : à notre sens, les restrictions de sélection qu'ils imposent à leurs rôles peuvent être incluses dans leurs « présuppositions et attentes ».

Ruppenhofer *et al.* (2006) précisent aussi que par choix, certains éléments qui pourraient être pris en compte pour le regroupement des lexèmes en ensembles de déclencheurs d'un même cadre ne le sont en fait jamais. Il s'agit des phénomènes de changements de diathèse n'impliquant pas de changement du nombre d'actants (passif, se-moyen…), de l'antonymie, et des différences relevant du contexte discursif (deixis, registre, dialecte…).

L'existence de tous ces critères d'appartenance à un cadre fait que comme le pointent Lönneker-Rodman et Baker (2009, section 3.4.2), chaque cadre décrit le comportement de tous ses déclencheurs. Les auteurs admettent que les décisions sur quels critères ont été retenus et sur la manière de les appliquer peuvent sembler arbitraires, et soulignent que les choix effectués découlent du but de créer une ressource telle que FrameNet : « documenter le comportement des déclencheurs et faire sur eux des généralisations utiles ». Il faut donc être le plus précis possible sur la sémantique assignée aux cadres tout en rendant compte des spécificités syntaxiques de chaque déclencheur. D'après Lönneker-Rodman et Baker (2009), cela est possible en utilisant judicieusement à la fois les relations inter-cadres pour créer des divisions et les relations intra-cadres (entre rôles) pour regrouper des déclencheurs aux comportements syntaxiques un peu différents.

3.3.2.2 Attribution de cadres aux mots polysémiques

Bien que nous ayons vu supra que FrameNet utilise une approche cadre à cadre plutôt que mot-à-mot pour le travail de délimitation des cadres, l'étude de la polysémie des mots a aussi une part dans ce travail. Fillmore (2007) pointe que même s'il n'est pas possible de décrire chaque sens de chaque mot examiné, il faut tout de même établir une limite entre le sens du mot qu'on souhaite garder comme déclencheur du cadre considéré et ses autres sens.

Le problème de délimitation des sens est connu pour être complexe. Kleiber (2008) indique par exemple que « tout le monde (ou presque) accepte qu'une expression linguistique puisse présenter une variation d'interprétations selon le contexte » mais que les analyses de ce qui constitue des sens distincts d'un même lexème varient fortement. Pour ce qui nous concerne, Fillmore (2007) indique que la philosophie du projet FrameNet est plutôt de multiplier les sens : elle porte plus à considérer plusieurs sens sur la base de très petites distinctions qu'à amalgamer des sens potentiellement différents en cadres plus génériques. Nous avons vu précédemment que les critères nécessaires pour pouvoir considérer que plusieurs mots déclenchaient un même cadre étaient nombreux. Dans

le même esprit, Fillmore (2007) inventorie les critères utilisés dans le cadre du projet FrameNet pour déterminer si un mot proposé comme déclencheur potentiel d'un cadre, et qu'on soupçonne d'être polysémique, appartient au lexique d'un seul ou plusieurs cadres (et donc a un ou plusieurs sens) :

— relations à d'autres lexèmes (antonymie, synonymie…) ;
— relation déclencheurs/patrons de valence ;
— réalisations syntaxiques particulières ;
— possibilités d'inférence différentes ;
— ambiguités non associées au type de complément ;
— relations morphologiques à des mots de la même famille.

Ces critères signalent l'existence de deux sens lorsqu'ils ne s'appliquent pas de la même manière aux deux sens proposés. Pour donner un exemple d'utilisation, le premier critère mentionné permet d'établir deux sens pour le verbe *dwell*. En effet, un sens de ce verbe, qui signifie habiter à un endroit, est synonyme du verbe *inhabit*, tandis que l'autre, celui de cogitation, est lui synonyme du verbe *ponder*. Le dernier test évoqué permet de confirmer ce diagnostic, puisque seul le premier sens donne lieu à la nominalisation *dwelling*, qui signifie une habitation.

Nous avons dit plus haut qu'il n'était pas toujours évident d'établir une claire distinction entre deux sens potentiels d'un mot, et la manière d'utiliser les critères utilisés par FrameNet que nous venons d'énoncer n'est pas décrite dans la littérature. Nous reviendrons sur ce problème au chapitre 5, section 5.1, où nous examinerons des tests classiques de polysémie pour tenter de les appliquer à notre cadre.

3.3.3 Pratiques d'annotation

Cette section nous permettra de passer brièvement en revue les pratiques décrites dans le guide d'annotation FrameNet (Ruppenhofer *et al.*, 2006).

Tout d'abord, sur quoi porte l'annotation? Comme nous l'avons mentionné en 3.2.2, pour chaque déclencheur examiné, les annotateurs FrameNet ont annoté le déclencheur lui-même, mais aussi les remplisseurs des rôles du cadre évoqué. Nous avons vu que les phrases annotées n'étaient dotées au départ d'aucune annotation morphologique ou syntaxique. Ces informations sont fournies manuellement au moment de l'annotation en cadres et en remplisseurs de rôles, et portent sur les déclencheurs et remplisseurs uniquement. Sont annotés le type de constituants des remplisseurs de rôles, leur fonction grammaticale par rapport au déclencheur, et parfois des informations spécifiques au type de construction dans lesquels ils s'intègrent. Suivent en vrac quelques règles d'annotation qui nous paraissent dignes d'être mentionnées.

Notons dès à présent que certaines catégories de déclencheurs n'ont pas du tout été considérées : c'est le cas des connecteurs de discours comme *while* et *although*. Par ailleurs, Ruppenhofer *et al.* indiquent qu'« en général [ils sélectionnent] pour l'annotation des phrases où, à l'exception des sujets, tous les rôles sont réalisés localement par des constituants appartenant à la projection maximale du déclencheur » [traduction] (Ruppenhofer *et al.*, 2006, p. 18). Cela englobe les dépendants directs et indirects du déclencheur, mais aussi le déclencheur lui-même. Les auteurs listent cependant des cas précis et contraints syntaxiquement pour lesquels l'annotation non-locale est permise. C'est le cas par exemple des phénomènes de contrôle et de montée, ou encore des constructions à verbe support. Bien que le guide d'annotation ne précise pas si la même politique est appliquée pour l'annotation de plein texte — mentionnant seulement que cette dernière oblige à annoter des phrases où le déclencheur s'insère dans des constructions quasi-agrammaticales —, l'observation des données révèle que c'est le cas et que seuls les remplisseurs de rôles réalisés localement (et phénomènes syntaxiques recensés dans le guide) ont été annotés.

Certains déclencheurs incorporent un de leurs rôles. C'est le cas par exemple du verbe *butter*, déclencheur de FILLING, dans lequel un Agent remplit ou couvre un Goal d'un Theme. Le déclen-

cheur *butter* incorpore le rôle Theme, puisque l'idée du beurre comme substance étalée fait partie de sa sémantique. Les incorporations de ce type sont précisées pour chaque déclencheur concerné au moment de la création du cadre.

Dans certains cas, un remplisseur de rôle est discontinu. C'est le cas par exemple dans les constructions à montée dans lesquelles le sujet réel de l'infinitive est en fait le sujet du verbe à montée. Chacun des types de cas recensés dans le guide a des consignes précises sur la manière d'assigner rôles et fonctions grammaticales à ces éléments discontinus.

Il arrive que l'information sur deux rôles d'un même cadre soit exprimée par un même constituant. Ce phénomène s'appelle la *conflation de rôles*. Fillmore (2007) en donne pour exemple la phrase (23a), dans laquelle *cured* évoque le cadre CURE (un soignant (Healer) ou un médicament (Medication) traite et guérit l'Affliction d'un Patient). Dans cette phrase, *my asthma* remplit le rôle d'Affliction et, par la présence du possessif *my*, donne aussi une information sur le Patient. Par contraste, en (23b), les rôles Patient et Affliction sont respectivement remplis par les constituants indépendants *me* et *of my asthma*. Lorsqu'il y a conflation, chaque remplisseur est annoté : un même mot ou constituant peut donc faire partie des remplisseurs de plusieurs rôles. Ainsi en (23a), *my* fait partie du remplisseur du rôle Affliction *my asthma* mais remplit aussi le rôle Patient.

(23) a. The medecine cured my asthma.
« Le médicament a GUÉRI mon asthme. »

b. The medecine cured me of my asthma.
« Le médicament m'a GUÉRI de mon asthme. »

CURE : Affliction, Medication, Patient

Il arrive aussi qu'un rôle ne soit rempli par aucun élément d'une phrase annotée. Les types de cas dans lesquels cela arrive, ainsi que le traitement FrameNet de chacun d'entre eux, sont détaillés dans la sous-section qui suit.

3.3.3.1 Zoom sur l'instanciation nulle

Lorsqu'un des rôles noyaux d'un cadre n'est pas rempli dans une phrase où un déclencheur évoque ce cadre, on parle d'*instanciation nulle* (*null instantiation*). Ruppenhofer *et al.* (2006) justifient leur volonté d'annoter ces cas par l'utilité de telles annotations pour rassembler des informations sur les conditions d'omissibilité des arguments sémantiques. Des procédures d'annotation spécifiques ont donc été définies pour ces cas. FrameNet distingue trois types d'instanciation nulle, que nous présentons ci-dessous.

Instantiation nulle définie (DNI)

Dans les cas d'*instanciation nulle définie* (*Definite Null Instantiation*, DNI), ou anaphorique, l'élément non réalisé doit être déjà présent dans le contexte linguistique ou discursif. Il s'agit d'un type d'instanciation nulle dépendant du lexique, c'est à dire qu'il ne s'applique pas à n'importe quel argument de n'importe quel lexème : ce sont les lexèmes eux-mêmes qui imposent lesquels de leurs arguments peuvent être omis.

(24) Sigibert **RETALIATED** by calling in his allies from across the Rhine. DNI DNI
« Sigibert a RIPOSTÉ en faisant venir ses alliés depuis l'autre côté du Rhin. »

REVENGE : Avenger, Injury, Offender, Punishment

(24) comporte un exemple de DNI. Dans cette phrase, le verbe *retaliate* déclenche le cadre REVENGE, qui décrit la situation où un protagoniste (l'Avenger) inflige une punition (Punishment) à un deuxième protagoniste (Offender) en réponse à une action (Injury) de ce dernier jugée mauvaise par le premier protagoniste. Les rôles Offender et Injury ne sont ici pas réalisés mais sont interprétables

si on a connaissance du contexte de la phrase. Si ce n'est pas le cas, cette dernière sera jugée non acceptable.

Instanciation nulle indéfinie (INI)

L'*instanciation nulle indéfinie* (*Indefinite Null Instatiation*, INI), ou existentielle, est, comme la DNI, un type d'instantiation nulle dépendant du lexique. Elle s'applique principalement aux verbes qui, comme *eat* et *drink*, s'insèrent aussi bien dans les constructions transitives (par exemple *Alex ate two avocados*; *Sam drank the whole bottle of orange juice*) qu'intransitives (par exemple *Alex always eats at night*; *Sam has been drinking*). Ruppenhofer *et al.* (2006) pointent que ces verbes ont été très décrits et qu'on considère qu'en emploi intransitif, l'objet omis a une interprétation stéréotypique : un repas pour *eat* et des boissons alcoolisées pour *drink* par exemple. La connaissance précise de la nature de l'objet omis n'est néanmoins pas nécessaire pour que la phrase soit acceptable et interprétable.

Lorsqu'un verbe de ce type évoque un cadre dans une phrase où il est en emploi intransitif, le rôle qui aurait été rempli par l'objet omis est considéré INI. Dans la phrase (25), le verbe *shave* évoque le cadre GROOMING, dans lequel un Agent fait sa toilette. Comme beaucoup de ses déclencheurs ont trait à la toilette d'une partie du corps en particulier, le cadre inclut aussi un rôle noyau Body_part. Ce rôle n'est pas rempli dans la phrase (25). On considère stéréotypiquement qu'il doit être question de la barbe de l'Agent, mais il pourrait tout aussi bien s'agir de ses cheveux, ou d'autres de ses poils. La connaissance de cette information, ou de quelque contexte que ce soit, n'est pour autant pas nécessaire pour que la phrase soit acceptable. Le rôle Body_part est donc INI.

(25) In a particularly bold move , I decided to **SHAVE** less often .INI
« Dans un geste particulièrement audacieux, j'ai décidé de me RASER moins souvent. »
GROOMING : Agent, Frequency, Body_part

(26) He **TAKES** and never gives back. INI
« Il PREND et ne donne jamais en retour. »
TAKING : Agent, Source, Theme

Bien qu'elle soit spécifique à certains verbes pour les cas que nous venons de voir, l'instanciation nulle indéfinie peut aussi avoir lieu avec des verbes qui n'autorisent normalement pas de construction intransitive. C'est le cas dans les constructions génériques, comme on peut le voir en (26), l'exemple proposé par Ruppenhofer *et al.* (2006).

Instantiation nulle constructionnelle (CNI)

Des constituants peuvent être omis constructionnellement, c'est-à-dire structurellement. On parle alors d'instanciation nulle constructionnelle (*Constructional Null Instantiation*, CNI). Il s'agit par exemple du sujet canonique d'un verbe transitif lorsque ce dernier est employé au passif, ou encore des sujets des verbes à l'impératif ou des propositions infinitives ou gérondives hors cas de contrôle ou montée (pour lesquels le contrôleur ou sujet de l'infinitive est accessible comme remplisseur de rôle, bien que non local). Le rôle qui aurait été rempli par ces constituants omis est donc considéré CNI. Les phrases en (27) et (28) contiennent toutes des exemples de CNI.

(27) a. According to Genesis the apple was **CREATED** three days before man, and it is the world's most cultivated tree fruit. CNI
« D'après la Genèse, la pomme fut CRÉÉE trois jours avant l'homme, et c'est l'arbre fruitier le plus cultivé au monde. »

b. The best way to solve pollution problems is not to **CREATE** them in the first place. CNI
« Le meilleur moyen de résoudre les problèmes de pollution est de commencer par ne pas les CRÉER. »

CREATING : Creator, Means, Created_entity

(28) "Please come and **SIT DOWN**." CNI
« "S'il te plaît viens et ASSIEDS-toi" »

CHANGE_POSTURE – Protagonist

Ruppenhofer *et al.* (2006) font remarquer qu'un rôle CNI peut recevoir une interprétation définie (anaphorique) ou indéfinie (existentielle), selon ce qu'impose ou non la construction impliquée. Contrairement à la DNI et à l'INI, la seule présence d'une CNI ne donne donc pas d'information sur l'interprétation à donner à l'élément non réalisé.

3.4 Projets d'adaptation de FrameNet dans d'autres langues que l'anglais

NB Dorénavant, nous utiliserons l'acronyme *BFN* (Berkeley FrameNet) pour référer au FrameNet de Berkeley, le projet FrameNet originel.

En s'appuyant entre autres sur l'expérience des différents projets de FrameNets non-anglophones que nous décrirons ci-après, Boas (2009b), puis Tonelli (2010), ont établi que la structure FrameNet était largement transposable à d'autres langues. L'orientation plus sémantique que celle d'autres projets de généralisation sur des prédicats, comme VerbNet (Kipper Schuler, 2005), peut laisser penser à une plus grande transposabilité. De fait, la plupart des projets d'adaptation de BFN à d'autres langues que nous décrirons un peu plus bas n'ont dû modifier qu'une petite partie des cadres inclus dans leur ressource. Cette transposabilité et l'observation de Boas (2009b) selon laquelle l'usage de cadres permet de pallier un certain nombre de problèmes standards du travail cross-linguistique semble indiquer que le réseau de cadres BFN peut être réutilisé tel quel ou presque pour créer un FrameNet dans une nouvelle langue : il « suffirait » d'identifier les bons déclencheurs et d'annoter des phrases les contenant.

Nous examinons dans cette section une partie des projets FrameNet qui ont émergé à la suite du Berkeley FrameNet : ceux qui nous semblent les plus aboutis, et illustrer le mieux la variété d'approches possibles. Une comparaison de ces projets est proposée table 3.3. La section 3.4.5 présente d'autres types de projets s'appuyant sur le modèle FrameNet : l'application du modèle FrameNet à des domaines spécialisés, puis la création automatique de ressources à partir du FrameNet de Berkeley.

3.4.1 Espagnol : Spanish FrameNet (SFN)

Spanish FrameNet (Subirats et Petruck, 2003; Subirats, 2009), ou SFN, est un projet de type FrameNet pour l'espagnol, développé en coopération avec le projet BFN à l'*Universitat Autònoma de Barcelona* et l'*International Computer Science Institute* à Berkeley. À l'heure actuelle, SFN propose 325 cadres comportant en tout 1 124 déclencheurs exemplifiés par 10 334 phrases annotées. Les annotations sont faites sur des phrases extraites d'un corpus électronique de 937 millions de mots.

Ce corpus inclut de l'espagnol d'Espagne comme de l'espagnol américain, et appartient à plusieurs genres différents : textes journalistiques, dépêches, essais dans le domaine des humanités, articles Wikipédia, critiques littéraires, résolutions de l'ONU, textes littéraires, et transcriptions de langue orale (sessions de Parlement espagnol et européen). Les phrases sélectionnées pour faire partie du corpus SFN sont étiquetées automatiquement en morphologie avant leur annotation en cadres et en rôles.

Subirats (2009) semble indiquer que la philosophie de SFN est de faire primer la sémantique sur la syntaxe en considérant comme accessoire que les remplisseurs de rôles d'un cadre évoqué par un prédicat espagnol soient ou non des arguments syntaxiques de ce prédicat ; d'autant que, dit-il, en espagnol, l'instanciation nulle des arguments est très souvent acceptable. Le projet SFN s'appuie donc sur les cadres déjà définis par BFN et utilise ensuite une méthodologie assez proche de celle du projet originel : les mots identifiés comme déclencheurs possibles sont examinés en corpus pour répertorier leurs réalisations syntaxiques possibles. Un sous-corpus de phrases illustrant les réalisations possibles de chaque déclencheur est ensuite automatiquement extrait, lemmatisé et étiqueté en morphologie, avant d'être annoté manuellement en cadres et en rôles.

Au début du projet, Subirats et Petruck (2003) indiquaient que SFN se penchait pour le moment sur des zones du lexique proches de ce qui existait déjà dans BFN, et que la plupart des cadres BFN restaient valables d'une langue à l'autre. Lönneker-Rodman et Baker (2009) évoquent des cas où un cadre BFN ne semblait pas adaptable à la langue espagnole, cas dans lesquels de nouveaux cadres spécifiques à l'espagnol auraient été créés pour les déclencheurs ne rentrant pas dans les cadres existants. Néanmoins, ni Subirats et Petruck (2003) ni Subirats (2009) n'en font mention, et nulle part dans les données SFN ne trouve-t-on d'indication que certains cadres SFN ont été créés spécialement pour le projet. Cela dit, 26 noms de cadres SFN actuels n'apparaissent pas dans les données BFN 1.5. On peut donc supposer que SFN comporte 26 cadres spécifiques à l'espagnol sur ses 325 cadres au total. De plus, Lönneker-Rodman et Baker (2009) font état de changements du statut de nucléarité de certains rôles dans 26 cadres sur les 103 adaptés de BFN qu'en comptait SFN en octobre 2008.

3.4.2 Japonais : Japanese FrameNet (JFN)

Japanese FrameNet (Ohara *et al.*, 2004; Ohara, 2009; Kabbach et Ohara, 2015), ou JFN, est un projet de type FrameNet pour le japonais développé entre Keio University et Tokyo University depuis 2002. Une version officielle de JFN est prévue pour mars 2017. Toutefois, un échantillon de données est d'ores et déjà consultable en ligne [7]. Kabbach et Ohara (2015) indiquent qu'en mars 2014, JFN comportait 73 372 occurrences de déclencheurs annotés pour 3 401 déclencheurs distincts répartis dans 990 cadres. Le corpus duquel sont extraites les phrases annotées comprend 8 millions de phrases journalistiques et des textes de romans et d'essais.

Comme c'est le cas pour SFN, la méthodologie de JFN suit le modèle décrit par Lönneker-Rodman et Baker (2009) sous le nom d'approche par expansion (*expand*). Le projet JFN prend donc les cadres déjà définis par BFN comme point de départ et évalue leur pertinence pour le japonais. Si un cadre ne semble pas convenir pour les mots japonais qui sembleraient pouvoir l'évoquer, les annotateurs ou annotatrices peuvent décider d'effectuer des modifications spécifiques au japonais sur le cadre originel, ou de créer un nouveau cadre spécifique au japonais qui sera évoqué par le mot considéré. Nous n'avons pour l'instant trouvé aucune information sur la proportion de cadres modifiés ou créés par rapport aux cadres BFN utilisés tels quels.

7. `http://sato.fm.senshu-u.ac.jp/frameSQL/jfn23/notes/index2.html`

3.4.3 Allemand : le projet SALSA

Le projet SALSA (Burchardt *et al.*, 2006b, 2009b; Rehbein *et al.*, 2012) est un projet de type FrameNet pour l'allemand développé à l'Universität des Saarlandes de 2002 à 2010. La dernière version des données (*SALSA Release 2.0*) comporte 36 251 occurrences de déclencheurs annotés pour 1 826 déclencheurs distincts définis sur 1 023 cadres. Les annotations sont effectuées sur la première version du TIGER corpus (Brants *et al.*, 2002), qui comporte environ 40 000 phrases (700 000 tokens) manuellement analysées en syntaxe de textes journalistiques allemands.

À la différence de BFN, SALSA procède à une annotation lemme à lemme et non cadre à cadre. L'ensemble des lemmes à annoter a d'abord été choisi sur divers critères, dont le principal est leur fréquence d'apparition dans le corpus. L'étape suivante est l'annotation de chaque occurrence de chaque lemme en cadre et en rôles, annotation effectuée en double-aveugle, puis adjudiquée. Rehbein *et al.* (2012) pointent que la méthodologie d'annotation lemme à lemme combine les intérêts des annotations lexicographique et de plein texte BFN, puisqu'elle illustre à la fois tous les sens possibles des lemmes annotés et pour chacun de ces sens, toutes ses constructions existantes en corpus.

Pour l'annotation, lorsqu'aucun cadre BFN existant ne permettait de rendre compte du sens et des rôles du lemme examiné, un proto-cadre spécifique à ce sens était créé pour l'annoter. (Rehbein *et al.*, 2012) indiquent aussi que dans certains cas, les cadres BFN ont dû être modifiés, que ce soit pour mieux s'adapter à des spécificités de l'allemand ou pour rendre la granularité de l'annotation plus consistante. Sur les 1 023 cadres annotés dans le cadre du projet, on compte 730 proto-cadres et 293 cadres BFN, dont 37 (soit environ 8%) modifiés.

Il existe aussi un autre projet de FrameNet allemand, le German FrameNet at Austin [8], basé à l'University of Texas at Austin et dirigé par Hans Boas. Le site du projet indique que ce projet, en collaboration avec BFN et SALSA, vise à créer un autre FrameNet allemand. Le site du projet ne présente néanmoins pas de données et les publications récentes recensées ne parlent pas de l'avancement du projet.

3.4.4 Suédois : Swedish FrameNet++ (SweFN++)

Swedish FrameNet++ (Borin *et al.*, 2009; Friberg Heppin et Toporowska Gronostaj, 2014), ou SweFN++, est un projet de type FrameNet pour le suédois développé à la Göteborgs universitet depuis 2009. SweFN++, qui a pour objectif final d'être partie intégrante d'une ressource lexicale multi-façettes, est en développement, mais une version de développement est actuellement consultable en ligne [9]. Ahlberg *et al.* (2014) indiquent qu'en octobre 2014, SweFN++ comportait 34 000 déclencheurs distincts répartis dans un peu moins de 1 200 cadres. Environ 8 500 occurrences de ces déclencheurs auraient été annotées. Les phrases-exemples sont extraites des corpus de Språkbanken (the Swedish Language Bank).

Friberg Heppin et Toporowska Gronostaj (2014) décrivent les étapes ayant présidé à la création de SweFN++. Comme pour SFN et JFN, on part d'un cadre BFN qui parait compatible avec le suédois. Des déclencheurs sont ensuite identifiés à l'aide d'un lexique associatif du suédois, ce qui permet d'obtenir un grand nombre de propositions. Une fois les déclencheurs trouvés, l'annotatrice ou annotateur choisit les plus représentatifs du cadre et leur cherche des phrases-exemples dans les corpus. Une procédure complémentaire consistant à partir de mots polysémiques pour répartir leurs sens dans des cadres est aussi utilisée.

Pour Friberg Heppin et Toporowska Gronostaj (2014), la plupart des cadres SweFN++ restent identiques aux cadres BFN originels. Lorsque des modifications mineures sont effectuées, les cadres sont identifiés par un changement de nom. Ces modifications consistent en général en la division

8. http://www.laits.utexas.edu/gframenet/index.html
9. https://svn.spraakdata.gu.se/sb-arkiv/pub/html/swefn_eng.html

d'un cadre en plusieurs à la granularité plus adaptée à diverses applications TAL. Friberg Heppin et Toporowska Gronostaj (2014) considèrent que la plupart des nouveaux cadres auraient leur place dans BFN, et n'identifient qu'un seul cadre (sur une cinquantaine de modifiés) comme ayant été créé pour raisons d'incompatibilité anglais-suédois.

3.4.5 Une diversité d'autres projets

Il existe de nombreux autres projets de type FrameNet pour des langues autres que l'anglais. On peut citer par exemple le chinois (You et Liu, 2005), l'italien (Lenci *et al.*, 2010), le polonais (Derwojedowa *et al.*, 2008), le portugais brésilien (Salomão, 2009), l'arabe (Ghneim *et al.*, 2009), le coréen (Nam *et al.*, 2014), l'hébreu (Hayoun et Elhadad, 2015)... Il faut noter que ces projets sont dans des états d'avancement très divers, et utilisent des méthodologies très diverses aussi, en particulier ils ne partagent pas tous l'idée de réutiliser la structure BFN.

Le modèle FrameNet a aussi donné lieu à d'autres types de ressources. Nous en présentons des exemples dans les deux sous-sections qui suivent. La première traite de l'application du modèle FrameNet à des domaines spécialisés, tandis que la seconde aborde la création automatique de ressources à partir du FrameNet de Berkeley.

3.4.5.1 Application du modèle FrameNet à des domaines spécialisés

Le modèle FrameNet a parfois été utilisé pour créer des ressources adaptées à des domaines spécialisés. Certains travaux proposent des extensions de FrameNet à ces domaines. C'est par exemple le cas de Dolbey (2009), qui étend le BFN au domaine de la biologie moléculaire en anglais. BioFrameNet, la ressource résultante, contient des cadres spécifiques à son domaine que Dolbey lie par des relations cadre-à-cadre aux cadres du BFN. Les annotations des déclencheurs suivent le modèle et la méthodologie de l'annotation lexicographique du BFN.

D'autres travaux s'appuient sur le modèle FrameNet pour créer des ressources terminologiques qui en diffèrent par certains aspects. Un exemple en est le Kicktionary de Schmidt (2009). Il s'agit d'un dictionnaire multilingue (allemand, anglais et français) du football, appuyé sur les principes FrameNet et pouvant à ce titre être vu comme un FrameNet multilingue pour domaine spécialisé.

Ressource	**BFN**	**SFN**	**JFN**	**SweFN++**	**SALSA**
Approche	CàC	CàC	CàC	CàC	LàL
Cadres	1 215	325	990	< 1 200	1 023
Déclencheurs distincts	13 311	1 124	3 401	> 34 000	1 826
Occurrences annotées	172 609 (lex) 28 157 (FT)	10 334	73 372	8 500	36 251
Tokens dans corpus utilisé	100 M	937 M	8 M	–	700 000

CàC : cadre à cadre
FT : plein texte
LàL : lemme à lemme

TABLE 3.3: Comparaison des différents projets de type FrameNet présentés

Comme FrameNet, il comporte des cadres peuplés de déclencheurs (dans les trois langues), ainsi que d'annotations pour ces déclencheurs, mais la ressource présente aussi des particularités qui l'éloignent du modèle FrameNet. Les cadres s'insèrent dans une hiérarchie avec ce que Schmidt appelle des *scènes*, des évènements prototypiques donnant lieu à différents cadres : la scène *Goal* (but) rassemble entre autres les cadres SCORE GOAL (marquer un but), AWARDING GOAL (accorder un but) et CELEBRATE GOAL (fêter un but). Les rôles ne sont pas dotés d'un statut de nucléarité. Par ailleurs, les unités lexicales sont aussi organisées en synsets à la manière de WordNet. Ces synsets sont mis en relation (par exemple hyperonymie/hyponymie) les uns avec les autres dans ce que l'auteur appelle une hiérarchie de concepts. Le Kicktionary compte 1 926 unités lexicales pour l'ensemble des trois langues, réparties dans 104 cadres et donnant lieu à 8 164 exemples annotés.

Le DiCoEnviro de L'Homme (2016) est un exemple plus récent de ressource terminologique inspirée de la sémantique des cadres et de la méthodologie du projet FrameNet. Il regroupe les termes prédicatifs du lexique de l'environnement et décrit leurs propriétés lexico-syntaxiques grâce à trois niveaux de description. Les cadres, inspirés des cadres FrameNet, proposent une définition générale, un ensemble de participants, et rassemblent des termes français, anglais et espagnols. Chacun de ces termes possède une entrée lexicale qui décrit sa structure actancielle et fournit des liens lexicaux à d'autres termes ainsi que des équivalents dans les deux autres langues. Enfin, chaque terme est lié à un ensemble d'annotations (en cadres et rôles mais aussi syntaxiques) sur des phrases choisies pour illustrer une variété de provenances, d'emplois et de réalisations des rôles. Les cadres peuvent être ajoutés à la ressource de deux manières. Certains proviennent de l'examen des cadres BFN pour repérer ceux qui auraient trait au domaine de l'environnement. Les autres ont été créés à partir des termes : ces derniers sont automatiquement extraits de corpus créés pour le domaine de spécialité, puis validés par un terminologue. On crée ensuite leur entrée lexicale et on annote des exemples tirés du corpus. Les termes peuvent alors être regroupés en cadres en fonction de leurs participants, selon les critères de délimitation en cadres du BFN. L'Homme (2016) indique que DiCoEnviro compte 125 cadres, dont plus de la moitié ont été créés plutôt qu'inspirés de cadres BFN, et 395 termes français.

Il existe encore d'autres travaux utilisant la sémantique des cadres dans un contexte terminologique, comme la « terminologie basée sur les cadres » de Faber *et al.* (2005). L'Homme (2016) souligne que « l'intérêt de la Sémantique des cadres et de son adaptation à un contexte terminologique résident dans le fait que la représentation conceptuelle ne fait plus l'impasse sur les propriétés lexico-sémantiques des termes. Au contraire, ce sont ces propriétés lexico-sémantiques qui guident sa définition » (L'Homme, 2016, p. 12).

3.4.5.2 Création automatique de ressources à partir du FrameNet de Berkeley

La production manuelle de ressources et d'annotations de type FrameNet est très coûteuse en temps et en travail. Obtenir des annotations FrameNet, ou même un lexique de déclencheurs FrameNet à peu de frais peut à la fois permettre d'accélérer ce processus, ou dans le cas des annotations, d'entraîner des analyseurs sémantiques en cadres et en rôles pour les langues ne possédant pas de corpus d'annotations FrameNet manuelles. Nous présentons brièvement dans la présente section des approches de création automatique de ressources de type FrameNet qui ont été développées pour le français ou appliquées au français. Notons cependant ici que notre travail de création d'un French FrameNet doit être comparé aux ressources développées manuellement décrites sections 3.4.1 à 3.4.4 et non à ce qui suit.

Padó et Lapata (2005, 2006), Padó (2007) et Padó et Pitel (2007) proposent de créer automatiquement un lexique de déclencheurs et un corpus d'annotations FrameNet. Ils utilisent pour cela la projection d'annotations en cadres, puis en rôles à travers des corpus parallèles dont une des langues possède un corpus d'annotations FrameNet. Leur méthode se veut indépendante de la langue à partir du moment où les langues source et cible possèdent un parallélisme sémantique assez haut (cela est mesuré en évaluant la correspondance d'annotations manuelles en cadres et

en rôles dans les deux langues sur un corpus parallèle). Elle utilise pour l'annotation des rôles des modèles d'alignement de constituants et des filtres pour réduire l'impact des erreurs d'alignement. La méthode a été testée sur l'allemand comme sur le français, en partant d'annotations soit manuelles, soit automatiques, sur la partie en anglais des corpus parallèles, et a été évaluée contre des annotations manuelles en langue cible. Les meilleurs modèles et filtres amènent pour les rôles à un F-score de, pour l'allemand 69% à partir d'annotations manuelles et 56,6% à partir d'annotations automatiques, et pour le français 63,1% en partant d'annotations manuelles et 57,2% à partir d'annotations automatiques.

Pitel (2009) présente une approche de classification de cadres et de rôles (c'est-à-dire qui sélectionne parmi les rôles d'un cadre celui qui est rempli par un fragment de texte donné) utilisant des annotations FrameNet anglaises et un corpus aligné (en phrases) anglais-langue cible, ainsi qu'optionnellement un étiqueteur morpho-syntaxique et un lemmatiseur pour les deux langues. Sa méthode utilise des calculs de similarité entre le fragment à catégoriser et des vecteurs créés à partir des annotations BFN représentant des rôles. Cela s'appuie sur l'idée que la plupart des rôles FrameNet présentent une forte cohésion sémantique, c'est-à-dire qu'on retrouve souvent des termes similaires dans leurs remplisseurs. Testée contre des annotations manuelles du français, cette approche donne, avec les meilleurs paramètres, un F-score de 65,8%.

Pitel propose aussi une méthode de construction d'un lexique FrameNet pour le français à partir du BFN et de ressources de traduction français-anglais mot à mot. Mouton *et al.* (2010) proposent pour leur part une autre méthode pour obtenir un lexique FrameNet français : une traduction de l'anglais vers le français des déclencheurs des cadres du lexique à partir de dictionnaires bilingues accompagnée d'heuristiques de filtrage et d'enrichissement. Pitel indique que de tels lexiques peuvent être utilisés pour guider un analyseur sémantique en cadres, mais aussi faciliter la tâche de potentiels annotateurs qui chercheraient à annoter directement du texte d'une langue dépourvue de FrameNet spécifique avec les cadres BFN.

L'idée de développer de telles ressources s'insère, comme les ressources créées manuellement présentées plus haut, dans le postulat formulé par Tonelli (2010) que la structure BFN est — en gros — indépendante de la langue, et que de nouveaux FrameNets peuvent être créés en leur liant des déclencheurs (lexique) et annotations dans la langue cible.

3.5 Motivation de notre choix d'utiliser FrameNet

L'objectif général du projet ASFALDA, dans lequel s'insère cette thèse, était de créer un corpus d'annotations sémantiques et un analyseur sémantique entraîné sur ces annotations. Nous ne souhaitions pas repartir de zéro, et avons donc pris la décision de partir d'une ressource existante. Nous avions donc le choix entre plusieurs types de modélisation. PropBank (Palmer *et al.*, 2005) aurait par exemple pu convenir, ou encore VerbNet (Kipper Schuler, 2005), puisqu'à l'époque du lancement du projet, le VerbɜNet du français de Pradet *et al.* (2014) n'existait pas encore. La présente section est consacrée à reprendre les caractéristiques de FrameNet qui font selon nous son intérêt et ont fait porter notre choix sur cette ressource.

Tout d'abord, le modèle FrameNet est orienté vers la sémantique bien plus que vers la syntaxe ; nous avons d'ailleurs vu plus haut que les notions spécifiques à FrameNet étaient généralement définies principalement en termes sémantiques. Cette orientation sémantique fait que comme nous l'avons mentionné, FrameNet est transcatégorie. Il s'agit d'une des raisons qui nous l'a fait préférer à VerbNet.

Cette orientation sémantique fait aussi que cette ressource est théoriquement plus portable à travers les langues qu'une ressource plus orientée syntaxe. Cela est d'ailleurs confirmé par le grand nombre de projets de type FrameNet s'appuyant sur le BFN dans une grande variété de langues. Dans beaucoup de ces projets, relativement peu de cadres ont été modifiés par rapport aux cadres BFN originels.

Il est par ailleurs intéressant de noter que malgré l'orientation très sémantique de FrameNet, ses annotations collent à l'organisation lexicale des phrases annotées : deux phrases en relation de paraphrase peuvent tout à fait être annotées avec des cadres différents. Ruppenhofer *et al.* (2006) donnent l'exemple des phrases en (29) qui seront annotées avec le cadre CHANGE_OF_CONSISTENCY pour (29a) et CAUSE_CHANGE_OF_CONSISTENCY pour (29b).

(29) a. The paste **HARDENED** due to hydration of the cement.
« La pâte a durci à cause de l'hydratation du ciment. »
Change_of_consistency : Reason, Undergoer

b. The hydration of the cement **HARDENED** the paste.
« L'hydratation du ciment a fait durcir la pâte. »
Cause_change_of_consistency : Cause, Undergoer

La raison principale qui nous a poussés à choisir FrameNet plutôt que PropBank est que son orientation plus sémantique le rend plus intéressant pour l'étude de l'interface syntaxe-sémantique. FrameNet propose une solution intéressante au problème de la difficulté à trouver une liste opérationnelle de rôles sémantiques : des rôles spécifiques à chaque cadre, mais sur lesquels il est aisé de généraliser grâce aux relations entre cadres. PropBank propose à l'inverse des rôles spécifiques à chaque verbe, qui semblent — à l'exception de l'arg0 censé correspondre au proto-agent de Dowty (1991) — correspondre à une simple numérotation des arguments, utilisant un ordre d'oblicité des fonctions grammaticales canoniques.

Verbnet était aussi une sérieuse possibilité, mais en plus du désavantage de ne comporter que des verbes, la ressource présente l'inconvénient de n'être associée à aucun corpus d'annotations. Il n'existe donc pas de retour d'expérience de l'annotation en corpus de classes et rôles VerbNet. L'existence de plusieurs corpus annotés en cadres et en rôles était plus rassurante quant à l'applicabilité du format FrameNet au corpus.

Enfin, FrameNet présente de nombreuses possibilités applicatives. La ressource a par exemple été utilisée pour améliorer des systèmes de questions-réponses (Shen et Lapata, 2007), d'extraction d'information (Barzdins, 2014), ou encore travailler sur la traduction automatique (Boas, 2002) ou la résolution d'implications textuelles (RTE) (Burchardt *et al.*, 2009c).

DEUXIÈME PARTIE

CRÉATION D'UN FRAMENET DU FRANÇAIS

CHAPITRE 4

LE PROJET ASFALDA

Sommaire

4.1 Un projet de création de ressource **63**
4.1.1 Un objectif différent de celui du FrameNet de Berkeley 64
4.1.2 Une stratégie en domaines guidée par le corpus 65
4.1.2.1 Les domaines : une brève description 67
4.2 Précision de notions pour le French FrameNet **68**
4.2.1 De la nécessité d'introduire explicitement un ancrage syntaxique des remplisseurs de rôles 68
4.2.2 Notions nécessaires à nos éclaircissements 71
4.2.2.1 Obligatoire sémantique 71
4.2.2.2 Sous-catégorisation 71
4.2.3 Nucléarité et rôles à annoter au niveau du déclencheur 75
4.2.3.1 Nucléarité au niveau du déclencheur 75
4.2.3.2 Rôles et remplisseurs à annoter pour un déclencheur 76
4.2.4 Délimitation d'un cadre 78
4.2.4.1 Profilage/perspective 79
4.2.4.2 Critères de délimitation retenus 81
4.2.4.3 Rôles et remplisseurs à annoter au niveau du cadre 82
4.2.5 Point sur les relations entre cadres dans le French FrameNet 83

4.1 Un projet de création de ressource

ASFALDA est un projet financé par l'ANR (ANR-12-CORD-023) lancé en septembre 2012 et terminé en juin 2016. L'objectif général de ce projet était de créer un corpus d'annotations sémantiques et un analyseur sémantique entraîné sur ces annotations. Le projet se divise en plusieurs tâches :

1. l'étude et enrichissement du modèle et de la structure de FrameNet ;
2. la création d'un lexique FrameNet du français ;
3. l'annotation d'un corpus en cadres et en rôles à partir de ce lexique ;
4. le développement d'un système d'analyse sémantique de surface à même d'utiliser les données annotées ;
5. et enfin, l'intégration de ce système au moteur de recherche d'un outil de gestion de contenu.

Cette thèse traite uniquement des trois premières tâches, c'est-à-dire de la création de la ressource annotée — le French FrameNet. Cette annotation s'est déroulée en plusieurs phases. La première, détaillée au chapitre 5, comprend la sélection et l'adaptation des cadres pour le French

FrameNet, ainsi que la construction de leur lexique. Ce travail a été effectué sur les sept domaines notionnels que nous présenterons en 4.1.2.1. Il a été effectué par onze personnes :

— pour le domaine Causalité : Philippe Muller et Laure Vieu

— pour le domaine Communication langagière : Marie Candito et Benoît Sagot

— pour le domaine Jugement/évaluation : Farah Benamara et Yannick Mathieu

— pour le domaine Positions cognitives : Pascal Amsili, Marie Candito et Marianne Djemaa

— pour le domaine Relations spatiales : Lucie Barque, Pauline Haas et Richard Huyghe

— pour le domaine Relations temporelles : Lucie Barque et Philippe Muller

— pour le domaine Transactions commerciales : Marie Candito et Marianne Djemaa

Le chapitre 7 traite de la phase suivante : l'annotation du corpus en cadres et en rôles. Sur les sept domaines concernés par la première phase, ce travail n'a porté que sur quatre : les transactions commerciales, les positions cognitives, la causalité et la communication langagière. Il a été coordonné par Marie Candito et moi-même, ainsi que Philippe Muller et Laure Vieu qui se sont concentrés plus particulièrement sur le domaine de la causalité. L'annotation double et une partie des adjudications ont été réalisées par six annotatrices : Vanessa Combet, Noémie Faivre, Virginie Mouilleron, Marjorie Raufast, Anny Soubeille et Émilia Verzeni.

Sur la totalité du travail effectué par l'ensemble des acteurs pour la création de la ressource, nous avons pour notre part travaillé plus particulièrement sur les points suivants :

— les choix généraux de stratégie de construction du French FrameNet

— les précisions de notions FrameNet retenues pour le French FrameNet

— le guide d'annotation et l'organisation de la campagne d'annotation sur corpus, avec entre autres des tâches d'organisation pratique et d'encadrement des annotatrices.

Nous avons vu en 2.2.7 et en 3.5 que parmi les différentes ressources de généralisation sur les lexèmes que nous avions présentées, FrameNet permettait différents niveaux de généralisation sur les rôles sémantiques ainsi qu'une structuration du lexique en classes, pouvant chacune regrouper des lexèmes de diverses catégories. La définition de ces classes de lexèmes est plus orientée vers la sémantique que vers la syntaxe ; à la différence de VerbNet (Kipper Schuler, 2005), le regroupement de divers verbes ne se focalise pas sur la similarité de leurs comportements syntaxiques. Cela présente l'intérêt de laisser apparaître des parallèles sémantiques divergeant par la syntaxe, comme par exemple les verbes converses. L'orientation sémantique du modèle FrameNet le rend aussi théoriquement plus portable à travers les langues qu'une ressource plus orientée syntaxe ; le modèle a d'ailleurs été appliqué à de nombreuses langues. Il présente aussi l'intérêt d'avoir déjà été annoté à grande échelle et utilisé pour l'analyse sémantique, puis pour de nombreux types d'applications. Ce sont tous ces éléments qui font que la ressource créée dans le cadre du projet ASFALDA s'appuie sur le modèle FrameNet.

Le French FrameNet présente néanmoins un certain nombre de différences avec le Berkeley FrameNet. Pour commencer, les deux ressources n'ont pas été créées avec le même objectif.

4.1.1 Un objectif différent de celui du FrameNet de Berkeley

Rappelons d'abord que les annotations du BFN ont été effectuées en deux phases distinctes, suivant deux méthodologies différentes. Nous l'avons vu en 3.2.2 : dans l'annotation lexicographique, des phrases isolées ont été annotées pour exemplifier les réalisations possibles des rôles pour chaque déclencheur d'un cadre. En revanche, dans l'annotation full-text —survenue dans un second temps— les cadres évoqués par tous les mots prédicatifs de textes continus ont été annotés, ainsi que les remplisseurs des rôles de ces cadres.

Un objectif central du projet ASFALDA est de produire un corpus annoté en cadres et rôles respectant la distribution (au sens probabiliste) des phénomènes linguistiques annotés. Il s'agit ici en particulier de respecter la fréquence d'apparition des cadres, mais aussi des sens des lemmes et de leurs possibles sous-catégorisations. Notre volonté de respecter cette distribution a deux raisons principales. D'une part, la quantification de ces phénomènes est en elle-même une information linguistique, qui mérite d'être étudiée. D'autre part, cette information est cruciale pour l'exploitabilité du corpus comme base à des expériences guidées par les données. En effet, la distribution statistique —ou fréquence en corpus— de ces différents phénomènes permet d'apprendre un modèle probabiliste valable pour la langue en général, ou du moins la langue utilisée dans notre corpus. Le respect de la distribution naturelle des cadres et rôles annotés permet donc, en plus de l'extraction d'un lexique quantifié, que les données servent de corpus d'entraînement pour l'apprentissage d'un analyseur en cadres et rôles. Das *et al.* (2010) ont d'ailleurs montré que les annotations lexicographiques BFN, qui sont pourtant beaucoup plus nombreuses, étaient bien moins utiles que leurs annotations full-text pour l'apprentissage automatique.

Cet objectif est à distinguer de la mission du FrameNet de Berkeley que présente Fillmore (2007) : décrire les possibles patrons de valence des déclencheurs, c'est-à-dire les façons dont les arguments sémantiques sont réalisés syntaxiquement, et accompagner ces descriptions d'attestations en corpus. En effet, le fait que les deux types d'annotation (lexicographique et full-text) aient été appliqués en phases successives et que l'annotation full-text n'ait pas fait partie des descriptions initiales du projet montre que BFN cherchait au premier chef à décrire l'exhaustivité des possibilités d'interface syntaxe-sémantique et ne s'intéressait donc pas au départ à la notion de fréquence en corpus.

À l'inverse, notre objectif présente le désavantage de ne pas pouvoir prétendre à l'exhaustivité des patrons de valence, ni même des sens couverts pour un même déclencheur potentiel : certains sens, certaines constructions peuvent être absents du corpus. Il s'agit d'un inconvénient bien connu de ce type d'approche, de même que le fait que nos données quantitatives sont nécessairement liées au corpus sur lequel le travail a été effectué. C'est néanmoins l'approche que nous avons choisie. L'annotation en cadres et en rôles a été effectuée sur deux corpus concaténés, le French Treebank (Abeillé *et al.*, 2003) et le Sequoia Treebank (Candito et Seddah, 2012). Nous verrons en 7.2 que notre stratégie a été d'annoter systématiquement les cent premières occurrences dans le corpus de chaque lemme pouvant déclencher un de nos cadres. Ainsi, la distribution des sens et des constructions au sein du corpus est reflétée.

4.1.2 Une stratégie en domaines guidée par le corpus

Pour des raisons pratiques, il nous semblait peu envisageable de travailler sur le modèle de l'annotation *full-text* pratiquée par l'équipe du Berkeley FrameNet après l'annotation lexicographique. Ce type d'annotation consiste à parcourir des textes continus en étiquetant de cadres tous les mots prédicatifs qu'on rencontre. Des essais préliminaires d'annotation full-text ont été faits, et il s'est avéré extrêmement difficile de maîtriser la sémantique de cadres très divers : pour une occurrence en corpus, il faut chercher parmi plus de mille cadres lequel pourrait convenir, ainsi que prévoir une stratégie pour les cas où aucun cadre existant ne conviendrait. Ce travail nous est apparu infaisable de but en blanc, sans grande connaissance préalable des différents cadres à notre disposition et des relations qu'ils entretiennent entre eux, ni expérience de la manière dont ils pourraient s'appliquer au français. De plus, on peut supposer que ce type d'annotation ne nous aurait fourni que des données très dispersées, avec peu d'exemples de chaque déclencheur et probablement aussi peu d'exemples de chaque cadre. Les deux autres principaux modes d'annotation utilisés par les différents projets FrameNet sont l'annotation lexicographique cadre à cadre et l'annotation lemme à lemme.

L'annotation lexicographique cadre à cadre est le premier mode de travail utilisé par le BFN. Nous l'avons vu en 3.2.2, l'idée est de concevoir des cadres, identifier leurs déclencheurs, et fournir

des exemples annotés des différentes réalisations syntaxiques des différents rôles de ces cadres. Cette méthodologie a plusieurs avantages. D'abord, elle permet une couverture lexicale totale à l'intérieur des cadres. Ensuite, on obtient aussi une couverture totale des possibles variations syntaxiques, à la fois pour un couple déclencheur-cadre et pour l'ensemble des déclencheurs d'un cadre. Il est par ailleurs plus facile d'effectuer des annotations pour un cadre que l'on connaît bien, ce que le travail cadre à cadre favorise. En revanche, cette méthode conduit à un lexique biaisé, intégrant seulement les sens couverts par les cadres existants. Le travail en isolation sur un cadre conduit par ailleurs à ne pas repérer ses similarités avec d'autres cadres, et donc potentiellement à créer une polysémie artificielle avec des frontières vagues entre les sens, et donc les cadres. Enfin, le travail lexicographique plutôt que guidé par les données amène à une mauvaise représentativité à la fois de la distribution des sens en corpus et à celle des constructions syntaxiques pour un sens donné.

Le travail lemme à lemme est le mode d'annotation utilisé pour le projet SALSA (Burchardt *et al.*, 2006b). Il s'agit comme nous l'avons vu en 3.4.3 d'annoter en cadres et en rôles toutes les occurrences en corpus de chaque lemme donné. À la différence de l'approche lexicographique cadre à cadre, celle-ci permet une représentativité parfaite (modulo les sens absents du corpus) pour les lemmes annotés, à la fois pour la distribution des sens et pour celle des constructions syntaxiques pour un sens donné. Le lexique est lui aussi représentatif puisque les lemmes y apparaissant sont appariés à tous leurs sens. De plus, pour un lemme donné, la couverture syntaxique est totale pour le corpus puisque toutes les constructions présentes dans le corpus sont annotées. En revanche, tant que tous les mots pleins du lexique d'une langue n'ont pas été considérés —c'est-à-dire tant que le travail de constitution d'un FrameNet lemme à lemme n'est pas terminé— la couverture lexicale d'aucun cadre ne peut être considérée comme totale. En cela cette méthode contraste avec l'approche cadre à cadre du travail lexicographique du BFN. Il est par ailleurs difficile de bien maîtriser les cadres qu'on annote, puisqu'ils changent a priori pour chaque nouveau lemme annoté. Enfin, dans la mesure où FrameNet ne dispose pas d'une couverture totale, l'approche lemme à lemme nécessite souvent la création de nouveaux cadres. Ces cadres seront à chaque fois créés à partir d'un lemme, ce qui rend difficile l'identification des liens et des frontières avec d'autres cadres. Dans le cas du projet SALSA, les lemmes n'évoquant aucun cadre existant donnaient lieu à la création d'un proto-cadre qui lui était spécifique. Cela implique une perte de la possibilité de généraliser sur les cadres.

Nous avons donc tenté de concevoir une approche conciliant le meilleur des deux mondes, une approche cadre à cadre guidée par les données, avec un ajout : les domaines notionnels. Nous avons sélectionné un certain nombre de concepts, comme la transaction commerciale, la communication langagière et la causalité, autour desquels créer nos domaines notionnels : des ensembles de cadres se rattachant à ces concepts. Nous avons ensuite rassemblé les cadres pré-existants appartenant à chacun de ces domaines, et défini pour chacun son lexique de déclencheurs. Au cours de ce travail, on examine les similarités entre cadres, en en créant de nouveaux ou en en fusionnant le cas échéant, ce qui permet d'éviter les redondances et d'obtenir des frontières claires entre les cadres. Nous avons ensuite eu pour objectif de décrire intégralement le lexique d'un domaine et d'en annoter les occurrences. En pratique, comme nous le verrons dans le flux de travail décrit chapitre 7 (section 7.2), nous avons adopté une stratégie lemme-par-lemme guidée par les domaines : on commence par les lemmes ne pouvant évoquer qu'un seul cadre (qui n'appartient généralement qu'à un seul domaine), puis les lemmes ambigus sur plusieurs cadres, mais au sein d'un même domaine, et enfin, les lemmes dont l'ambiguïté concerne des cadres de plusieurs domaines. Ce dernier cas demande plus de maîtrise des cadres que les autres et a à ce titre été traité en fin de projet, une fois les annotatrices rôdées. Cette méthode présente plusieurs avantages. D'abord, une couverture lexicale exhaustive pour les domaines visés : une fois le travail terminé pour un domaine donné, on a un cadre pour chaque lexème pouvant appartenir au domaine, le lexique de chacun des cadres de ce domaine est complet, et les cent premières occurrences en corpus de chacun

des lexèmes du domaine sont annotées. Le deuxième avantage est le respect de la distribution naturelle des phénomènes au sein du corpus ; cela concerne à la fois la distribution des sens et celle des constructions syntaxiques. Les données peuvent donc être utilisées pour entraîner des systèmes automatiques de désambiguïsation lexicale partielle (car restreinte aux domaines annotés) et d'étiquetage en rôles sémantiques. Par ailleurs, le fait que l'annotation soit effectuée par groupes de lemmes évoquant potentiellement les mêmes cadres permet une certaine expertise au moment de l'annotation, puisqu'on a pour chaque groupe un nombre limité de cadres à maîtriser et qu'on est bien au fait des frontières de sens. Notre approche présente néanmoins certains inconvénients. À l'inverse de ce que permet l'annotation lexicographique du BFN mais comme pour toute approche guidée par les données, notre mode d'annotation fait que si certaines constructions sont absentes du corpus pour une paire déclencheur-cadre, elles ne figureront pas parmi les annotations. Le deuxième désavantage est inhérent à la restriction de notre approche à certains domaines : pour un lemme donné, on ne peut couvrir que les sens appartenant à nos domaines. Nous verrons au chapitre 8 qu'environ un tiers des occurrences annotées ont été marquées comme relevant d'un sens hors des domaines visés. Ce pourcentage est très variable d'un lemme à l'autre, avec des lemmes dont toutes les occurrences sont annotées et d'autres pour lesquels la majorité des occurrences est hors-domaines.

4.1.2.1 Les domaines : une brève description

Les domaines initialement retenus dans le projet étaient au nombre de sept, et peuvent être regroupés en deux types : les domaines notionnels ciblés et les domaines transversaux. Les domaines ne sont pas nécessairement disjoints : des relations peuvent lier leurs cadres respectifs ; certains cadres chevauchent même plusieurs domaines.

Les *domaines notionnels ciblés* réunissent des cadres se rapportant à une idée spécifique, et présentent un intérêt particulier pour les applications telles que l'extraction d'information. Il s'agit des domaines *Transactions commerciales*, *Communication verbale*, *Positions cognitives* et *Jugement/évaluation*.

Le domaine *Transactions commerciales* regroupe les cadres décrivant une situation contenant un échange d'argent contre des biens ou services (propriété), l'usage temporaire de biens (location), ou du travail (paiement de salaire). La notion est relativement aisée à modéliser et le domaine bien circonscrit. L'ensemble de cadres associé est bien étudié dans le BFN ; ces cadres servent d'ailleurs souvent d'exemples dans la littérature FrameNet. Le domaine a la particularité d'inclure des verbes converses (comme *acheter* et *vendre*). Le modèle FrameNet est particulièrement adapté pour l'annotation de ce type de verbes : contrairement à des ressources plus orientées syntaxe, il permet d'utiliser le même rôle — par le biais de généralisation sur les relations entre rôles — pour le sujet canonique de *vendre* et l'objet indirect canonique d'*acheter*.

Le domaine *Positions cognitives* rassemble des prédicats exprimant la position d'un être doué de conscience envers la valeur de vérité d'un contenu propositionnel. Il s'agit donc principalement de croyances, pouvant varier sur plusieurs axes comme le degré de certitude (*penser* vs. *savoir*), l'aspect (*penser* vs. *réaliser* vs. *convaincre*), ou encore le moment de réalisation de la propositon (*penser* vs. *prévoir* vs. *se souvenir*). Nous avons considéré que l'annotation de ce domaine pourrait par exemple être utile pour l'extraction de faits. Les nombreux prédicats factifs pourraient en particulier être intéressants dans la perspective d'exploiter un corpus annoté en cadres pour la reconnaissance d'implications textuelles *Recognizing Textual Entailments* présentée dans (Burchardt et Pennacchiotti, 2008). Ce domaine sera décrit de manière bien plus approfondie au chapitre 6.

Le domaine *Communication langagière* regroupe les cadres décrivant une situation dans laquelle quelqu'un exprime, au moyen du langage, un message ou parle d'un sujet, en s'adressant éventuellement à un ou plusieurs destinataires. Les prédicats de ce domaine sont omniprésents dans le corpus sur lequel nous avons travaillé, du fait de sa nature journalistique. Des annotations des

cadres de ce domaine pourraient être utiles en Extraction d'information, puisqu'elles permettraient de repérer *qui a dit quoi à qui*.

Le domaine *Jugement/évaluation* regroupe les situations dans lesquelles une entité (personne, objet ou évènement) est évaluée ou jugée, positivement ou négativement, par rapport à une norme. Les déclencheurs de ce domaine doivent intrinsèquement exprimer un jugement ou une évaluation, comme c'est le cas pour *bon*, *mauvais* ou *beau*. Les éléments lexicaux qui, bien qu'autorisant une lecture évaluative, peuvent être objectifs dans certains contextes ne sont pas intégrés au domaine. C'est le cas par exemple de l'adjectif *grand*, objectif dans *la table est grande* mais évaluatif dans *il a de grandes dents*.

Les *domaines transversaux* ont quant à eux trait à des phénomènes sémantiques plus généraux. Ces domaines se situent à l'interface entre sémantique lexicale et discours, et les termes évoquant leurs cadres incluent des prépositions, conjonctions et adverbes, ce qui n'est généralement pas le cas pour les domaines notionnels ciblés.

Le domaine *Cause* incorpore des prédicats dénotant des relations de causalité entre événements, processus ou états qui sont réalisés (ou envisagés comme réalisés), ainsi que des relations de conséquence logique entre faits ou propositions. Ces deux types de relations sont considérés mais pas nécessairement distingués par les cadres du domaine. Le travail effectué sur ce domaine est détaillé dans Vieu *et al.* (2016).

Le domaine du *Temps* s'intéresse aux cadres décrivant des relations temporelles (ordre ou inclusion) et aux relations de durée. Les cadres temporaux-aspectuels comme PROCESS_START et PROCESS_STOP, ainsi que les cadres temporels impliquant une causation, comme CHANGE_EVENT_DURATION et CHANGE_EVENT_TIME ont été laissés de côté. Il est intéressant de noter que la sémantique de la relation temporelle (antériorité/postériorité/simultanéité/etc. de la cible par rapport au site) est précisée au niveau lexical.

Enfin, le domaine des *Relations spatiales* regroupe les cadres de localisation statique et de déplacement dans l'espace (et exclut les cadres de placement et de mouvement sans déplacement). Sa difficulté, mais aussi un de ses intérêts, sera la différenciation des cas métaphoriques de ceux où le déclencheur dénote bien une relation spatiale. L'annotation des cas métaphoriques sera peut-être conduite dans un second temps, avec une annotation en deux cadres (un pour le sens littéral et un pour le sens figuré) sur le modèle du projet SALSA (Burchardt *et al.*, 2006b).

Notons dès à présent que faute de temps, le travail n'est complet que pour quatre des sept domaines cités : les transactions commerciales, positions cognitives, communication verbale et causalité. Pour les trois autres, les étapes de sélection/adaptation de cadres et de constitution du lexique (décrits section 5.2) ont été réalisées, mais pas la phase d'annotation en corpus (cf. chapitre 7).

4.2 Précision de notions pour le French FrameNet

4.2.1 De la nécessité d'introduire explicitement un ancrage syntaxique des remplisseurs de rôles

Nous avons mentionné plusieurs fois au chapitre précédent que certains concepts définis par la littérature FrameNet, présentés principalement en termes sémantiques, nous paraissaient pour le moins flous. Les rôles sont par exemple simplement décrits comme nommant les « participants, accessoires, phases et parties des sortes de situations nommées par le cadre » (Fillmore, 2007, p. 130). De même, les notions de statut d'un rôle (noyau, périphérique ou extra-thématique, cf. section 3.2.1.2) sont décrites en termes principalement sémantiques, avec très peu de références à des concepts syntaxiques. Nous avons vu section 3.3.2 que c'était aussi le cas des critères de délimitation

d'un cadre.

Par ailleurs, la ressource présente certaines incohérences qui nous semblent aussi liées à ce flou définitionnel. Dans certains cadres, un rôle est prévu (et annoté) pour un élément qui sera pour tout déclencheur du cadre réalisé comme un modifieur. C'est le cas par exemple du rôle extra-thématique Depictive du cadre MANIPULATION, où un agent manipule physiquement une entité. Ce rôle, qu'on voit utilisé en (30), correspond à une description de l'agent au moment où il ou elle manipule l'entité. Tout verbe d'activité permet un modifieur de type Depictive, mais le rôle n'est pas prévu dans tous les cadres évoqués par des verbes d'activité. Ainsi, dans le cadre INGESTION, où un protagoniste ingère une substance, aucun rôle n'est prévu pour un modifieur qui décrirait l'état du protagoniste, alors que *Avec une grimace, elle avala l'huile de foie de morue* est aussi acceptable que (30). La présence d'incohérences de ce type pourrait être limitée par l'utilisation de critères de définition des rôles plus clairs, plus bornés par la syntaxe.

(30) With a grimace she **GRABBED** her hair and twisted it into a knot on top of her head.
« Avec une grimace, elle saisit ses cheveux et les tordit en chignon sur le sommet de sa tête. »

MANIPULATION : Agent, Entity, Depictive

De même, comme nous l'avons mentionné en 3.2.1.2, de nombreux rôles apparaissent dans certains cadres comme extra-thématiques et dans d'autres comme périphériques, alors que leur utilisation semble être la même dans les deux cas. Ainsi le rôle Time apparaît comme extra-thématique dans 24 cadres et périphérique dans 661, et dénote dans les deux cas le moment auquel se déroule l'évènement nommé par le cadre. Pour nous, il s'agit là d'un signe de ce que la distinction entre rôle périphérique et rôle extra-thématique est trop floue.

L'existence et l'utilisation de ces rôles extra-thématiques nous paraît elle-même peu cohérente. Ruppenhofer *et al.* (2006) disent explicitement que l'annotation de rôles extra-thématiques est alternative à l'annotation de leurs remplisseurs comme déclenchant eux-mêmes des cadres. Ainsi, dans la phrase (31), *7h30* pourrait être annoté comme rôle extra-thématique du cadre COMMERCE_BUY évoqué par *a acheté*, mais aussi comme remplisseur du rôle Trajector_period du cadre TEMPORAL_COLLOCATION, qui mettrait alors en relation l'évènement d'achat et *7h30*, le moment auquel il a eu lieu.

(31) Sam a acheté des croissants à 7h30.

Dans l'optique d'une annotation totale (comme l'annotation de plein texte du BFN), les rôles extra-thématiques sont donc redondants avec les annotations en cadres. L'existence de tels rôles nous semble découler de deux éléments : d'abord, le parti pris de FrameNet de partir de la sémantique et pas de la syntaxe, et donc de ne pas partir des cadres de sous-catégorisation syntaxique pour créer les cadres. Les modifieurs peuvent donc être traités de la même manière que les arguments sous-catégorisés par un verbe du moment qu'ils peuvent apparaître dans la phrase. Ensuite, la méthode lexicographique fait que l'on doit avoir tendance à annoter tout élément réalisé localement au déclencheur, afin de capturer le plus d'informations possible. Nous avons vu que notre méthode de travail visait à obtenir des annotations en corpus, non sélectionnées pour leur exemplarité. Il nous a semblé préférable d'éviter la redondance d'avoir par exemple une préposition temporelle (comme *à* dans l'exemple (31)) annotée à la fois comme tête d'un remplisseur de rôle extra-thématique et comme déclencheur d'un cadre temporel. Vu notre choix de ne couvrir que certains domaines et de n'annoter que les cent premières occurrences de chaque déclencheur potentiel, il n'est bien sûr pas garanti que dans une phrase donnée, tous les mots qui devraient déclencher un cadre seront annotés comme tels. Néanmoins, l'important est pour nous qu'une telle annotation est possible et souhaitable, et de toutes façons théoriquement existante. Nous préférons donc pour ces cas

une annotation en cadres qui sera interprétable de manière compositionnelle. Une définition plus formelle de ce qui peut ou non constituer un rôle permettra d'éviter ce type de redondance.

Nous avons dit souhaiter être cohérents, entre autres :

— proposer le même traitement pour les modifieurs remplissant le même rôle auprès des déclencheurs de cadres différents ;
— proposer le même statut pour les rôles occupant exactement la même fonction dans des cadres différents ;
— et éviter les annotations redondantes pour les modifieurs d'un déclencheur dont la tête déclenche elle-même un cadre.

Les définitions très sémantiques de ce qu'est un rôle et des statuts de ces rôles qui sont proposées par la littérature FrameNet ne nous permettent pas de garantir que ces volontés soient respectées. Il nous semble donc important d'avoir des définitions plus syntaxiques à la fois du statut des rôles et des rôles que nous souhaitons annoter.

Les sections qui suivent visent à repréciser les définitions proposées par la littérature FrameNet. Les propositions que nous y formulons portent uniquement sur le cas des déclencheurs verbaux. Nous expliquons ici pourquoi les autres catégories de déclencheurs n'ont pas été traitées.

Le fait que les définitions des concepts spécifiques à FrameNet soient entièrement sémantiques ne nous a posé problème que pour certaines catégories de déclencheurs. Pour d'autres, comme les conjonctions et adverbes dénotant des relations de discours, se reposer sur une modélisation sémantique fonctionne parfaitement. C'est le cas par exemple avec *donc* (adverbe) et *parce que* (conjonction) : les deux arguments sémantiques traditionnellement représentés en modélisation discursive (cf. par exemple les annotations du Penn Discourse Treebank (Miltsakaki *et al.*, 2004)) correspondent aux rôles noyaux des cadres évoqués. De même pour les déclencheurs prépositionnels, il semble clair que leurs arguments sémantiques, et donc les rôles noyaux de leurs cadres, sont nécessairement au nombre de deux : en général l'objet de la préposition et la tête modifiée par la préposition.

Pour les autres catégories de déclencheurs, à savoir les verbes, les noms et les adjectifs, les annotations sur la base de cadres calqués de l'anglais ont rapidement posé des problèmes de définition des rôles, et de choix du statut de ces derniers. Dans les sections qui suivent, nous tenterons d'utiliser la syntaxe pour borner la définition des rôles et de leurs statuts (ce qui nous permettra aussi de spécifier plus précisément comment se délimite un cadre), ainsi que ce que nous souhaitons annoter précisément dans le French FrameNet. Nous nous pencherons sur ces problèmes uniquement dans le cadre des déclencheurs verbaux, puisque ce sont eux qui apparaissent majoritairement dans le domaine des positions cognitives, domaine sur lequel nous avons passé le plus de temps.

Nous ne nous sommes pas réellement penchés sur des critères particuliers pour les noms. Les noms évènementiels apparaissent généralement dans des cadres ayant également des déclencheurs verbaux, qui permettent de définir les rôles grâce aux critères que nous précisons plus bas. Seront annotés comme remplisseurs des rôles du cadre les éléments qui peuvent être réalisés comme des syntagmes prépositionnels dont la préposition est régie par le nom. Le cas des noms relationnels est traité en 7.3.2, où nous expliquons avoir prévu des modélisations où le nom, s'il est en emploi référentiel, réfère au remplisseur d'un des rôles du cadre. C'est le cas par exemple des noms *conséquence* et *opinion*.

Pour ce qui est des adjectifs, qui ne sont pas très nombreux, nous avons simplement considéré comme arguments sémantiques de l'adjectif (annotés comme remplisseurs de rôles) l'élément modifié lorsque l'adjectif est employé comme épithète, et les éléments réalisables comme syntagme prépositionnel régi par l'adjectif — s'ils existent.

4.2.2 Notions nécessaires à nos éclaircissements

Pour définir à la fois ce qu'est un rôle nucléaire et précisément quels rôles seront annotés dans le French FrameNet, nous utiliserons les notions d'obligatoire sémantique et de sous-catégorisation d'un dépendant. Nous décrivons donc dans les deux sections qui suivent ce que couvre chacune de ces deux notions et dans quel sens nous les entendons.

4.2.2.1 Obligatoire sémantique

Bonami (1999) décrit la notion d'*obligatoire sémantique*. Pour lui, un argument d'un prédicat est sémantiquement obligatoire s'il est impossible de l'omettre, ou s'il y a systématiquement une relation d'implication entre la forme avec omission et une forme où l'argument est réalisé (même indéfini, comme *quelque chose* ou *quelque part*). C'est le cas pour l'exemple (32), où (32a) implique nécessairement (32b), et pour (33), où (33a) implique nécessairement (33b).

(32) a. Alex a informé les autorités.
b. Alex a informé les autorités de quelque chose.

(33) a. Alex a mangé une pomme.
b. Alex a mangé une pomme quelque part.

Cette notion d'obligatoire sémantique permet de faire la différence entre les cas comme (32) ou (33) et ceux comme (34), où (34a) n'implique pas (34b) et donc un complément référant à un interlocuteur n'est pas sémantiquement obligatoire.

(34) a. Alex a murmuré quelques mots.
b. Alex a murmuré quelques mots à quelqu'un.

Le test de Bonami nous a parfois paru difficile à juger, mais nous n'avons pas trouvé d'alternative parfaite. Citons tout de même le travail de Colinet *et al.* (2014). Dans le cadre d'une étude sur les infinitives en *pour*, les auteurs citent le critère selon lequel un complément sous-catégorisé est sémantiquement obligatoire, contrairement à un ajout. Ils traduisent ce critère en un test spécifique aux infinitives en *pour* mais, « le caractère sémantiquement obligatoire d'un argument pouvant être sujet à caution », choisissent de ne pas le retenir dans leur étude.

Dans notre travail, nous souhaitons conserver la distinction entre obligatoire sémantique et sous-catégorisation, en particulier à cause de l'existence de cas comme (34) où le dépendant correspondant à l'interlocuteur est clairement sous-catégorisé mais dont nous venons d'établir qu'il n'était pas sémantiquement obligatoire. Nous suivrons donc le découpage de Bonami, et c'est à sa définition que nous réfèrerons ci-après quand nous parlerons d'obligatoire sémantique.

4.2.2.2 Sous-catégorisation

On reconnaît généralement deux types de fonctionnement pour les unités se rattachant à un prédicat. Les *arguments* (ou compléments essentiels, régis, sous-catégorisés…) sont sélectionnés (sémantiquement et syntaxiquement) par leur prédicat, et font partie de son schéma de construction. Les *ajouts* (ou modifieurs, compléments adjoints, compléments non-essentiels, circonstants…) modifient leur prédicat de rattachement mais ne sont pas sélectionnés par lui. On retrouve généralement deux axes d'analyse de cette distinction : la syntaxe (les prédicats sous-catégorisent des arguments syntaxiques) et la sémantique (les prédicats dénotent des relations, leurs dépendants syntaxiques sont potentiellement les arguments sémantiques de ces relations). Comme nous le montre la section précédente, arguments syntaxiques et arguments sémantiques ne concordent néanmoins pas nécessairement. La présente section inventorie les tests rencontrés dans la littérature en s'intéressant uniquement aux critères formels (non sémantiques) qui permettent de distinguer

les arguments syntaxiques des ajouts, c'est-à-dire les dépendants sous-catégorisés par leur prédicat de ceux qui ne le sont pas.

Le critère d'obligatoire syntaxique revient régulièrement dans la littérature pour déterminer si on a affaire à un argument syntaxique ou à un ajout : un dépendant syntaxiquement obligatoire est nécessairement sous-catégorisé. Ce critère est suffisant mais pas nécessaire : Miller (1997) pointe que bien que l'objet direct constitue l'exemple le plus typique d'argument syntaxique, il est facultatif pour de nombreux verbes (comme *manger*). Le fait de pouvoir supprimer un dépendant n'indique donc rien sur son statut sous-catégorisé ou non.

Bonami (1999) présente la propriété d'itérabilité, censée permettre de distinguer les arguments des ajouts. Il cite Pollard et Sag (1987) :

> « En général, deux instances ou plus du même type d'ajout peuvent être combinées avec la même tête [...] ce qui est impossible avec les compléments. » (Pollard et Sag, 1987, p. 136), [traduction de Bonami (1999, p. 32)]

Pour Bonami, certains types d'arguments syntaxiques sont itérables, et si on considère comme il le fait que deux ajouts sont du même type s'ils peuvent être coordonnés, alors de nombreux ajouts ne sont pas itérables. L'auteur ne considère donc pas ce test comme opératoire, et à sa suite nous ne l'utiliserons pas.

On considère généralement qu'à la différence des arguments, les ajouts peuvent être déplacés en tête de phrase, mais Bonami montre que ce n'est pas le cas. En examinant les propriétés de placement des dépendants, il établit cependant trois tests, chacun fonctionnant comme condition suffisante du statut d'ajout :

— Si un dépendant peut être réalisé en position pré-verbe fini (entre le SN sujet et le verbe fini), c'est un ajout. Le fait que (35a) soit acceptable mais pas (35b) montre qu'*aimablement* est ajout du verbe *accepter* dans la première mais argument syntaxique de *se comporter* dans la deuxième.
— Si un dépendant non-adverbial peut être réalisé dans la position post-verbe fini (l'environnement immédiat du verbe fini à sa droite), c'est un ajout. Le fait que (36a) soit acceptable mais pas (36b) montre qu'*avec rudesse* est ajout pour le verbe *bousculer* dans la première phrase mais argument syntaxique de *se comporter* dans la seconde.
— Si un dépendant non-topicalisable peut être réalisé en tête de phrase, c'est un ajout. Le clivage est un procédé classique d'extraction, et Bonami (p. 56) indique que l'extraction d'un dépendant rend possible l'inversion du sujet. L'inacceptabilité de (37a) et de (37b) montre donc que *malheureusement* ne peut pas être extrait. Or, on voit en (37c), que cet adverbe peut être placé en tête de phrase. On a donc affaire à un ajout.

(35) a. Alex, **aimablement**, a accepté mon invitation.
b. *Alex, **aimablement**, s'est comportée.

(36) a. Sam a, **avec rudesse**, bousculé Alex.
b. *Sam s'est, **avec rudesse**, comportée.

(37) a. *C'est **malheureusement** que Sam a refusé de venir.
b. ***Malheureusement**, Sam croit que se trompe Alex.
c. **Malheureusement**, Sam croit qu'Alex se trompe.

Pour Bonami, la seule condition nécessaire et suffisante du statut d'ajout parmi les tests qu'il inventorie est ce qu'il appelle le test de *le faire*. Il semble que *le faire* puisse être suivi d'un ajout (*Alex est allé à Paris lundi soir, alors que Sam l'a fait mardi matin*) mais pas d'un argument (**Alex a rencontré son frère alors que Sam l'a fait sa sœur*). L'auteur montre qu'on a là un test adéquat, condition nécessaire et suffisante à la prétention au statut d'ajout pour tous les dépendants de verbes d'action. En effet, ce test fonctionne avec tous les types de dépendants. Avec les exemples en

(38) et (39), Bonami montre que *le faire* peut être combiné avec les ajouts de tout type — cf. (38) — mais n'est compatible avec les arguments syntaxiques d'aucun type — cf. (39).

(38) (Repris des exemples (5) de Bonami (1999, p. 70))

a. Jean a posé sa candidature au poste de directeur. Il l'a fait **à la demande de Marie**.

b. J'ai demandé à Paul de relire ce texte. Il le fera **s'il a le temps**.

c. Paul a traversé le désert les mains dans les poches, alors que Marie l'a fait **chargée d'un sac de 30 kilos**.

d. J'ai demandé à Paul de nous expliquer la situation. Il le fera **lundi matin**.

e. – Où Jean a-t-il pris cette photo ?
– Il l'a fait **place du marché**.

(39) (Repris des exemples (6) de Bonami (1999, p. 71))

a. Jean a expliqué le problème en détail. *Il l'a fait **à Marie**.

b. Je voulais que Jean convainque Marie. *Il l'a fait **qu'il avait raison**.

c. *Jean a exigé d'être entendu par la commission, et Paul l'a fait **d'assister à la réunion**.

d. Je voulais que Jean rende son texte plus cohérent. *Il l'a fait **illisible**.

e. *Jean s'est comporté admirablement, alors que Marie l'a fait **lamentablement**.

Miller (1997) évoque un autre critère, le *degré de sélection du verbe*, qu'il définit ainsi :

> *« Par degré de sélection il faut entendre le degré auquel le verbe détermine son complément à la fois du point de vue formel et du point de vue sémantique (cf. Allerton 1982 : 61-64 ; Radford 1988 : 235). »* (Miller, 1997, p. 3)

Ce critère se manifeste par le fait que lorsque la forme que prend un dépendant est contrainte par le verbe, et que sa contribution sémantique au sens de la phrase dépend aussi de ce dernier, on considère avoir affaire à un argument. Ainsi, en (40), la présence (et type) ou absence de préposition sont déterminées par le verbe, ainsi que la contribution sémantique du dépendant au sens de la phrase, ce qui permet de dire qu'on a affaire à un argument sémantique et pas à un ajout. À l'inverse, comme on peut l'observer en (41), le choix de préposition des ajouts est relativement libre, et leur contribution sémantique plus indépendante du choix du verbe.

(40) (Repris des exemples (5) de Miller (1997, p. 3))

a. Marie a répondu **{à/*∅/*vers} la question**.

b. Marie a écouté **{∅/*à/*vers} la question**.

(41) a. Le hamster mangeait une carotte **{dans/sur/sous/près de/vers} la boîte**. [1]

b. Le hamster faisait ses besoins **{dans/sur/sous/près de/vers} la boîte**.

L'auteur pointe cependant vers deux types de cas problématiques. D'abord, il arrive que la forme d'un dépendant sous-catégorisé ne soit pas contrainte ; c'est le cas par exemple des locatifs utilisés comme compléments de destination des verbes sous-catégorisant ces derniers, comme en (42). Dans ces cas-là, une partie de la contribution sémantique du dépendant vient de la préposition, mais cette contribution reste néanmoins déterminée par le verbe, puisque c'est lui qui permet de savoir qu'on a affaire à une destination et pas à un locatif. Ce type de cas indique que bien que le critère de contrainte de forme du dépendant soit suffisant pour indiquer qu'un dépendant est sous-catégorisé, il n'est pas nécessaire.

1. Traduction de l'exemple (6a) de Miller (1997, p. 4).

(42) Marie a mis le livre **{dans/sur/sous/près de/vers} la boîte**. [2]

Le deuxième problème est posé par le fait qu'il existe des exemples d'ajouts pour lesquels la contribution sémantique dépend du verbe. Ainsi en (43), on peut voir que l'interprétation d'*avec la clé* comme indiquant un instrument ou un accompagnant dépend du verbe employé. Si le critère d'indépendance sémantique du verbe utilisé est suffisant pour indiquer le statut d'ajout, il n'est en revanche pas nécessaire. Nous verrons cependant en 4.2.4.3 que pour le French FrameNet, nous avons cherché à annoter tous les dépendants dont le sens ne pouvait être interprété qu'à l'aide du verbe auquel ils se rattachaient.

(43) (Traduction des exemples (7) de Miller (1997, p. 4))

a. Marie a ouvert la porte **avec la clé**.

b. Marie est venue **avec la clé**.

Bonami (1999) comme Miller (1997) pointent des incohérences dans les tests généralement utilisés, ainsi que des cas dans lesquels ils se contredisent. Sur la base de l'examen de ces contradictions, Miller fait l'hypothèse, plutôt que d'une rupture nette entre arguments et ajouts, d'un continuum entre arguments prototypiques et ajouts prototypiques. Fabre et Bourigault (2008) confirment cette hypothèse en identifiant des degrés d'autonomie verbe-dépendant, y compris des degrés médians entre arguments prototypiques et ajouts prototypiques. On voit donc qu'en TAL comme en linguistique, la distinction argument-ajout est problématique. Nous verrons section 4.2.3 que dans notre travail, nous avons souvent eu besoin de trancher. Pour ce faire, quand nous nous sommes trouvés devant un cas de dépendant de statut inconnu, nous avons tenté d'appliquer les différents tests identifiés ci-dessus. En cas de contradiction ou d'incohérence, nous avons généralement choisi de considérer que nous avions affaire à un complément sous-catégorisé. Cette vision « large » du complément sous-catégorisé nous fournit en effet plus de matériel annotable, puisque nous verrons en 4.2.3 que les arguments sous-catégorisés par un déclencheur sont systématiquement annotés.

Nous avons utilisé les tests décrits supra lorsque nous avions besoin de déterminer le statut sous-catégorisé ou non d'un élément, mais nous l'avons dit, cela ne s'avère pas toujours facile. Il existe parfois des tests spécialement conçus pour déterminer le statut d'un type de dépendant. C'est le cas pour les propositions de type *pour* suivie d'une infinitive, pour lesquelles Colinet *et al.* (2014) ont proposé plusieurs tests spécifiques permettant de distinguer entre *pour* introducteur d'argument sous-catégorisé et *pour* connecteur de discours (introduisant un ajout). Ces propositions en *pour* (que nous noterons *pourVinf* comme l'ont fait les auteurs) nous ayant souvent posé problème, ce sont leurs critères que nous avons utilisés [3]. Il s'agit de :

— Si *pour* peut commuter avec une préposition listée comme introduisant un argument dans une ressource connue, on considère que le *pourVinf* est sous-catégorisé.

— Si l'objet de la phrase matrice contrôle le sujet du *pourVinf*, on considère que le *pourVinf* est sous-catégorisé.

— Si le *pourVinf* peut-être précédé de *et ce*, idéalement précédé d'une virgule, on a affaire à un ajout.

— Si l'interprétation la plus naturelle de la réfutation d'une phrase avec *pourVinf* est qu'elle porte sur le contenu de la phrase matrice, le *pourVinf* est sous-catégorisé. À l'inverse, si la réfutation est ambigüe (peut porter sur le contenu de la phrase matrice ou sur la relation entre cette dernière et le contenu du *pourVinf*), il s'agit d'un ajout.

— Si une phrase comportant un *pourVinf* peut être prolongée par un *pourVinf* de but sans effet d'itération, le *pourVinf* originel est sous-catégorisé.

2. Traduction de l'exemple (6b) de Miller (1997, p. 4).

3. Nous avons en fait repris leurs annotations (mais ces dernières ne concernent que le cas des pourVinf dépendant d'un verbe).

Les sections suivantes sont consacrées à la reprécision de concepts FrameNet nécessaires à notre travail de création du French FrameNet.

4.2.3 Nucléarité et rôles à annoter au niveau du déclencheur

Nous venons de nous pencher sur les notions d'obligatoire sémantique et de sous-catégorisation, dont nous avons vu qu'elles ne se recoupaient pas toujours. Il est possible pour un dépendant d'un prédicat d'être sémantiquement obligatoire, sous-catégorisé, les deux à la fois, ou aucun des deux. C'est ce qu'illustre la table 4.1 qui présente un exemple de chacune de ces situations.

Dans cette section, nous utilisons ces concepts d'obligatoire sémantique et de sous-catégorisation pour définir au niveau du prédicat d'abord des critères formels de nucléarité pour les rôles, puis une délimitation précise des rôles et remplisseurs qui seront annotés dans le French FrameNet. Les définitions obtenues seront plus tard utilisées au niveau du cadre : en 4.2.4.2, le statut nucléaire des rôles sera utilisé pour la délimitation de groupes de déclencheurs en cadres, et en 4.2.4.3, les remplisseurs à annoter seront considérés par cadre plutôt que par déclencheur.

4.2.3.1 Nucléarité au niveau du déclencheur

BFN définit le concept de nucléarité (*coreness*) de manière très sémantique : un rôle noyau est un rôle « conceptuellement nécessaire tout en rendant le cadre unique et différent des autres cadres » (Ruppenhofer *et al.*, 2006, p. 19) [4]. Cette définition se place au niveau du cadre, mais vu les critères de délimitation d'un cadre que nous rappellerons section 4.2.4, il faut identifier les rôles noyaux de chaque déclencheur potentiel pour pouvoir déterminer lesquels doivent appartenir à un même cadre. Ainsi nous examinons ici le concept de nucléarité d'un rôle pour un prédicat, et utiliserons plus tard la définition obtenue pour délimiter le périmètre de nos cadres.

Nous l'avons vu en 3.2.1.2 : en plus de la définition sémantique, qui nous paraît un peu floue, la littérature FrameNet propose plusieurs propriétés suffisantes pour déterminer la nucléarité d'un rôle :

- si un rôle doit forcément être exprimé, il est noyau
- si un rôle reçoit forcément une interprétation définie lorsqu'il n'est pas exprimé, il est noyau
- si la sémantique d'un rôle ne peut être prédite à partir de la forme de sa réalisation, il est noyau, ce qui entraîne que :
  - les rôles exprimables sans marquage formel (c'est-à-dire les sujets et objets directs) sont noyaux
  - les rôles dont le marquage formel est idiosyncratique sont noyaux

Rappelons que nous n'avons ici que des propriétés suffisantes, qui ne nous permettent pas de dire de tout rôle s'il est ou non noyau. Par ailleurs, ces critères (en partie syntaxiques) contredisent parfois la définition (sémantique) énoncée supra : un dépendant ayant un « marquage formel idiosyncratique » peut ne pas etre sémantiquement obligatoire. C'est le cas dans une phrase comme

	Sémantiquement obligatoire	**Non sémantiquement obligatoire**
Sous-catégorisé	*Cette question concerne **Sam et Alex**.*	*Alex a murmuré quelque chose **à Sam**.*
Non sous-catégorisé	*Sam a acheté des fleurs **à Paris**.*	*Sam est allée à la plage **avec Alex**.*

TABLE 4.1: Exemples de dépendants sémantiquement obligatoires, sous-catégorisés, les deux à la fois, ou aucun des deux.

4. Dans le texte : *A core frame element is one that instantiates a conceptually necessary component of a frame, while making the frame unique and different from other frames.*

Alex murmura quelques mots (à Sam), et le rôle rempli par *Sam* ne devrait donc pas être noyau. Nous cherchons donc à remplacer ces propriétés par des conditions nécessaires et suffisantes, qui ne contrediraient pas la définition sémantique annoncée supra.

Pour ce faire, nous tentons de traduire l'idée de « conceptuellement nécessaire », qui nous paraît difficile à définir et à tester, par la notion d'*obligatoire sémantique* telle que décrite par Bonami (1999), que nous avons présentée en 4.2.2.1. Nous avons vu qu'un argument d'un prédicat était sémantiquement obligatoire si on ne pouvait pas syntaxiquement l'omettre, ou si une forme avec omission impliquait nécessairement une forme avec l'argument réalisé.

C'est cette notion d'obligatoire sémantique qui va nous permettre de faire la différence entre les cas comme (32) ou (33) et ceux comme (34), où (34a) n'implique pas (34b) et donc un complément référant à un interlocuteur n'est pas sémantiquement obligatoire. On ne souhaite donc pas qu'un rôle désignant cet interlocuteur soit noyau pour *murmurer*, ni pour le cadre qu'il évoque.

Néanmoins, l'obligatoire sémantique seul ne nous paraît pas suffisant pour permettre de déterminer le statut de nucléarité des différents rôles. En effet, nous ne souhaitons pas considérer comme rôles noyaux et à annoter ceux qui définissent par exemple la temporalité ou la localisation d'un évènement, et qui seront interprétés de la même manière pour de larges classes de prédicats (pour notre exemple, tous les évènements). Nous voyons à cela deux raisons. D'abord, comme nous l'avons mentionné en 4.2.1, la relation qu'entretiennent ces rôles plus génériques et leur prédicat est compositionnelle, et leur annotation en tant que rôles d'un cadre serait redondante avec leur annotation compositionnelle en cadres. Ensuite, le fait que ces rôles soient interprétés de la même manière pour de larges classes de prédicats va à l'encontre de ce que semble indiquer la formulation BFN de la définition de la nucléarité : un rôle noyau est conceptuellement nécessaire *tout en rendant le cadre unique et différent des autres cadres.*

Nous avons vu en 4.2.2.2 qu'on postulait généralement que les prédicats sélectionnaient la forme de leurs arguments sous-catégorisés, et que l'interprétation de ces derniers était dépendante du verbe, ce qui n'était pas le cas pour les ajouts. Tout dépendant non sous-catégorisé par un déclencheur serait donc interprétable compositionnellement, ce qui, vu nos remarques du paragraphe précédent, nous mène à dire qu'il ne nous faudrait pas considérer de tels éléments comme des rôles noyaux. Nous avons aussi vu en 4.2.2.2 que Miller (1997) indiquait que ce critère de sélection du verbe donnait parfois des résultats contradictoires avec celui d'obligatoire sémantique. Les deux critères apportent donc une information complémentaire. De ce fait, nous proposons de définir la nucléarité comme suit :

> *Un rôle est noyau pour un déclencheur si et seulement si il est sémantiquement obligatoire et sous-catégorisé par ce déclencheur.*

L'ajout de ce critère de sous-catégorisation nous permet bien de considérer que même si comme on l'a vu en (33), l'argument exprimant la localisation est sémantiquement obligatoire pour le verbe *manger*, ce dernier ne sous-catégorise pas cet argument, et le rôle qu'il remplit n'est donc pas noyau. Nous verrons section 4.2.4.3 que les rôles sous-catégorisés mais non sémantiquement obligatoires seront aussi annotés.

4.2.3.2 Rôles et remplisseurs à annoter pour un déclencheur

Nous tentons ici de spécifier quels remplisseurs de rôles nous souhaitons voir annotés pour un déclencheur donné, en particulier s'il faut ou pas s'en tenir uniquement aux rôles noyaux.

Tout d'abord, nous ne souhaitons considérer que les éléments pouvant être exprimés *localement* au déclencheur. Nous entendons comme locaux à un prédicat ses dépendants, direct ou via préposition, ainsi que son gouverneur syntaxique dans le cas d'un prédicat employé comme ajout : le gouverneur syntaxique du prédicat est en effet dans ces cas-là son dépendant sémantique. C'est ce qu'on peut voir en (44), où la participiale dont *observés* est la tête est syntaxiquement gouvernée par *tous les effets indésirables*, qui est aussi son argument sémantique.

(44) Pour avoir le détail de tous les effets indésirables OBSERVÉS lors de l'utilisation de Angiox, voir la notice.

Nous excluons de la notion de localité les cas de périphrase avec ajout de prédicat (verbal par exemple), comme c'est le cas en (45) où le contenu de l'insulte ne peut être exprimé localement. Il s'agit là d'un ancrage syntaxique important, qui nous permet de borner l'information capturée par l'annotation en rôles à ce qui est exprimable dans le domaine syntaxique d'un prédicat. On maximise ainsi la quantité de données qui devraient être utilisables pour étudier l'interface syntaxe-sémantique. Rappelons ici que pour sa partie d'annotations lexicographiques, le BFN se restreint de la même manière aux remplisseurs de rôles locaux, exception faite de quelques phénomènes syntaxiques bien circonscrits comme le contrôle ou les constructions à verbe support.

(45) a. *Alex insulte Sam que sa coiffure est hideuse.
b. Alex insulte Sam, disant que sa coiffure est hideuse.

Dans le cas d'élément exprimable localement au prédicat, mais qui en contexte serait exprimé non localement (de manière syntaxiquement régulière comme dans les cas de contrôle et de montée, mais aussi de manière ad hoc), nous faisons le choix de l'uniformité sémantique. L'élément sera annoté comme remplisseur du rôle, et signalé d'un marquage spécial pour signifier sa non-localité. Ainsi on voit en (46a), la situation dont Sam se plaint peut être exprimée localement au déclencheur se plaindre. Elle sera donc annotée de la même manière en (46b), avec un marquage signalant sa non localité.

(46) a. Sam se plaint qu'Alex est désagréable.
b. Sam se plaint, disant qu'Alex est désagréable.

Nous l'avons mentionné plus haut, Ruppenhofer *et al.* (2006) font remarquer que les rôles extra-thématiques BFN n'appartiennent pas conceptuellement au cadre *F* qui les englobe mais correspondent en fait à des cadres plus abstraits qui décriraient la relation entre l'évènement ou état décrit par le cadre *F* et un autre évènement ou état décrit par le remplisseur du rôle extra-thématique. Selon nous, cela est vrai aussi des rôles périphériques. Si ces derniers — à la différence des rôles extra-thématiques — n'introduisent pas d'évènement autre que celui décrit par le cadre *F* qui les englobe, ils sont — à la différence des rôles noyaux — applicables à tout un ensemble de cadres relevant d'un même type sémantique, et interprétés au moins grossièrement de la même manière pour tous ces cadres. Des cadres spécifiques peuvent donc être utilisés pour dénoter la relation entre ces rôles et tout cadre qui les inclut, par exemple la relation entre tout évènement et le moment auquel il se produit. Si on souhaite éviter toute redondance, il faut donc choisir d'annoter cette information soit sous forme de cadres, soit par le biais de rôles. La méthode par cadres permet de manière plus évidente de rendre compte de la généricité des rôles non-noyaux, ce qui la rend à notre sens plus cohérente. Il nous faudrait donc à première vue n'annoter que les rôles noyaux, dont nous avons vu supra qu'ils couvraient les dépendants à la fois sous-catégorisés et sémantiquement obligatoires.

La question est cependant un peu plus compliquée. Certes, nous souhaitons éviter d'annoter comme rôles d'un cadre les dépendants dont l'interprétation n'est pas liée au déclencheur, mais nous souhaitons aussi annoter tous ceux dont l'interprétation dépend de leur déclencheur, même partiellement. En effet, si l'interprétation ne peut pas se faire compositionnellement, une annotation au niveau du cadre du déclencheur est nécessaire. Ce n'est qu'ainsi qu'on peut envisager de capturer toute l'information de la phrase à l'aide d'annotations de type FrameNet. D'après ce que nous avons rapporté plus haut de ce que Miller (1997) appelle critère de degré de sélection du verbe, les dépendants en question sont les dépendants sous-catégorisés (et certains ajouts, si marginaux que nous nous permettons de les ignorer). Nous ajoutons à cela, dans le cas d'un prédicat employé comme ajout, le gouverneur syntaxique dudit prédicat. En effet, dans ces cas-là, l'ajout-prédicat a une interprétation dépendante du prédicat qu'il modifie.

Vu tous ces éléments, pour un déclencheur, on annote et assigne un rôle aux éléments qui peuvent être réalisés comme un complément sous-catégorisé (ou comme gouverneur syntaxique dans les cas de déclencheur-ajout). Parmi ceux-ci, les rôles noyaux sont ceux qui sont sémantiquement obligatoires. Les rôles réalisables uniquement comme ajout ne sont pas considérés dans les rôles du cadres, car leur sémantique sera systématiquement capturable compositionnellement. Les cadres French FrameNet ne comporteront plus dans leur description que les rôles dits à annoter, séparés selon leur statut de nucléarité. Notons que pour le French FrameNet, nous ne considérons donc plus que deux statuts : *noyau* et *non-noyau*.

Nous profitons de cette section pour mentionner que nous n'annotons que les éléments sous-catégorisés qui correspondent à des arguments sémantiques, sélectionnés par le déclencheur. Nous ne considérons par exemple pas de rôle pour le *il* explétif, comme dans *il neige*, ni pour le sujet des verbes à montée. Cela nous amène aussi à ne pas considérer de rôle pour l'objet direct d'un déclencheur verbal dans les constructions à attribut de l'objet, puisque l'objet est sélectionné par l'attribut mais pas par le verbe. Nous verrons en 7.3.1 que dans ces cas précis, l'objet direct est intégré au rôle rempli par l'attribut puisqu'il en est le sujet, et que nous avons pour politique de saturer les remplisseurs de rôles (cf. 7.1.2.4).

Nous amendons donc notre résumé supra :

> *Pour prévoir un rôle d'un prédicat, il faut non seulement qu'il puisse être réalisé comme un élément sous-catégorisé par le déclencheur (ou comme l'élément qu'il modifie pour les déclencheurs-ajouts), mais aussi que cet élément soit un argument sémantiquement plein, sélectionné par le déclencheur.*

Nous obtenons là une définition syntaxique de ce qu'est un rôle, ce qui n'est pas dans la tradition FrameNet mais est apparu nécessaire pour obtenir des critères opératoires.

4.2.4 Délimitation d'un cadre

Le workflow BFN est décrit comme passant par la délimitation de cadres grossiers (on suppose qu'il s'agit de délimiter des groupes de déclencheurs potentiels vraisemblablement similaires), puis la définition plus précise des cadres (définition, rôles et relations entre rôles, liste des déclencheurs). Les rôles sont définis à partir des exemples en corpus des déclencheurs, c'est-à-dire, peut-on supposer, en regardant l'ensemble des déclencheurs (par exemple en dégageant leurs points communs et en en dérivant des rôles).

Ruppenhofer *et al.* (2006) décrivent des critères d'appartenance à un cadre pour les déclencheurs. Un cadre est donc défini à partir d'un ensemble de déclencheurs présentant un certain nombre de points communs :

— même nombre et type de rôles noyaux ;
— profilage du même rôle ;
— mêmes relations entre rôles ;
— même ensemble d'implications sur les étapes et transitions ;
— mêmes présuppositions, attentes et concomitants ;
— dénotation grossièrement similaire.

Il s'agit en théorie de conditions nécessaires à l'appartenance à un même cadre, mais il nous semble qu'en pratique, ces critères ne sont pas toujours respectés. Le cadre AWARENESS par exemple, dans lequel un être doué de conscience considère la véridicité d'un contenu mental, contient à la fois des déclencheurs pour lesquels le contenu est présupposé vrai (comme *know* ou *aware*) et d'autres pour lesquels ce n'est pas le cas (comme *think* ou *believe*). Dans la version 1.5, la volonté de faire basculer les déclencheurs pour lesquels le contenu n'est pas présupposé dans un autre cadre était explicitée, mais plusieurs années après, dans la version 1.6, cette modification n'a toujours pas été effectuée.

Les trois premières des propriétés énoncées supra ont trait aux rôles définis pour les déclencheurs d'un même cadre, et en particulier aux rôles noyaux, c'est aussi pourquoi nous avons en 4.2.3.1 tenté de préciser la notion de nucléarité. La section 4.2.4.1 s'intéressera à la notion de profilage d'un rôle, qu'il nous paraît aussi utile d'étudier. Nous serons alors mieux outillés pour décider, en 4.2.4.2, des critères que nous souhaitons utiliser.

4.2.4.1 Profilage/perspective

Nous avons dans le chapitre précédent présenté rapidement le concept de perspective dans la section 3.2.1.4 qui décrit les relations cadre à cadre. Nous y avons aussi mentionné que nous n'avons pas pour le French FrameNet utilisé la notion de perspective. Nous revenons ici sur cette notion pour expliquer les raisons de son abandon.

Ruppenhofer *et al.* (2006) expliquent l'idée que tous les déclencheurs d'un cadre doivent profiler le même rôle comme le fait que l'accent devrait être mis sur le point de vue du même participant. Il précisent que les cadres ne différant que par le ou les rôles profilés sont typiquement liés à un même cadre neutre par la relation de Perspective. Cette dernière est d'ailleurs similairement décrite dans le même guide d'annotation. L'exemple habituellement donné de ce type de relation est celui des verbes *acheter* et *vendre*, qui décrivent objectivement la même situation mais mettent en avant l'acheteur (ou l'acheteur et les biens, suivant les versions) pour *acheter* et le *vendeur* (ou le vendeur et les biens suivant les versions) pour *vendre*. La description du profilage et de la perspective proposée par Ruppenhofer *et al.* (2006) nous semble très vague. S'il nous faut utiliser le fait de partager la même perspective comme critère d'évocation d'un même cadre pour deux déclencheurs potentiels, il nous faut comme pour la nucléarité avoir une vision plus claire de cette notion.

L'observation de la littérature du projet BFN n'apporte pas beaucoup plus de précisions. Le travail plus ancien de Fillmore sur les cas profonds comporte un premier indice de ce qui peut être formellement entendu par ces « points de vue » :

> « Dans l'utilisation que je suggère, n'importe quel verbe identifiant n'importe quel aspect particulier d'un événement commercial va nous contraindre à mettre EN PERSPECTIVE au moins une des entités de cet événement ; ce choix se manifeste en Anglais par la sélection de fonctions grammaticales correspondant aux notions de sujet et d'objet direct sous-jacents. Par exemple, si je souhaite prendre le point de vue du vendeur et des biens, j'utiliserai le verbe vendre. Si je souhaitais prendre le point de vue de l'acheteur et de l'argent, j'utiliserais le verbe dépenser [...] Et ainsi de suite[5]. » [traduction] Fillmore (1977a)

Gawron (2011) décrit une fonction de *perspectivation* qui semble correspondre à la notion de perspective. Il explique que le rôle de la fonction de perspectivation est de sélectionner une réalisation syntaxique pour les rôles de l'évènement mis en perspective. Nous retombons donc sur la même idée mais toujours pas assez précisée pour nous permettre de justifier la séparation d'un groupe de déclencheurs potentiels en deux cadres différents.

Continuant sur la piste de la syntaxe, nous nous appuyons sur les données BFN et sur Gawron (2011) pour formuler une hypothèse de définition syntaxico-sémantique du profilage et de la relation de perspective. Nous supposons que le rôle qu'on dit « profilé » par un prédicat est celui qui est assigné à son sujet canonique. Dès lors, il s'agit, pour un noyau de sens de cadre (avec un ensemble de rôles donnés), de regrouper dans un même cadre les déclencheurs qui assignent le même rôle (ou ensemble nucléaire de rôles) au sujet canonique. Les différents cadres résultant de cette opération

5. Dans le texte : « In the usage that I suggest, any verb identifying any particular aspect of the commercial event will constrain us to bring one or more of the entities in the event INTO PERSPECTIVE, the manifestation of this choice for English being the selection of grammatical functions corresponding to the notions of underlying subject and direct object. For example, if I wish to take the perspective of the seller and the goods, I will use the verb sell. Should I wish to take the perspective of the buyer and the money, I will use the verb spend [...] And so on. » Fillmore (1977a)

sont en relation de perspective avec un cadre neutre non-lexical, et chacun d'eux profile un rôle différent. Cela fonctionne pour les verbes. *Acheter* et *acquérir* ont le même linking canonique (l'acheteur est sujet, les biens achetés objet et le vendeur oblique), qui diffère de celui de *vendre* et *céder* (le vendeur est sujet, les biens vendus objet et l'acheteur oblique). *Acheter* et *acquérir* profilent donc l'acheteur tandis que *vendre* et *céder* profilent le vendeur.

Gawron (2011) présente deux intérêts à descendre au niveau de la perspective pour décrire les prédicats. Tout d'abord, cela permet de préserver le principe souvent considéré comme valable qu'à travers les langues, les agents ont tendance à se trouver en position sujet. Cela permet aussi d'interpréter les ajouts qui ont une relation particulière avec les agents, comme ceux dénotant un instrument, un bénéficiaire ou un but. Il est donc intéressant de regrouper en cadres uniquement les déclencheurs présentant une même perspective. Mais nous allons voir que les notions de profilage et de perspective posent plusieurs problèmes qui selon nous pèsent plus lourd dans la balance que ces avantages.

D'abord, un point pratique : la notion n'est pas toujours facile à appliquer aux catégories de déclencheurs autres que les verbes. Dans le cas des conjonctions, comme *car* et *donc*, on peut imaginer spécifier un peu arbitrairement que le rôle profilé est celui qui sera réalisé comme premier argument de la conjonction : cela permet de grouper *donc* avec *causer* et *car* avec *résulter*. Mais dans le cas des déclencheurs nominaux, qu'est-ce qui justifie qu'*achat* (les biens sont `de_obj`, le vendeur `a_obj` et l'acheteur `par_obj`) évoque le même cadre qu'*acheter* et *acquérir*? Le `par_obj` est très oblique, pourquoi serait-il la fonction qui correspond au sujet pour les verbes? On peut imaginer des contournements pour les déclencheurs nominaux. On peut s'ils réfèrent à un participant, considérer que ce dernier est profilé : cela permet par exemple de grouper *vente* et *vendeur*. Ou s'ils réfèrent à un évènement, considérer quel participant est impliqué si on emploie un ajout exprimant le but, la manière ou un bénéficiaire, ce qui permettrait de rassembler *achat* et *acheter*... Mais ces stratégies ne s'appliquent pas à tous les types de déclencheurs : quid des noms dénotant un évènement ne faisant pas intervenir d'agent?

De plus, le critère de profilage des mêmes rôles n'est pas uniformément appliqué dans le BFN. Cela est parfois fait en connaissance de cause : Ruppenhofer *et al.* (2006) précisent que pour gagner du temps, certains cadres ont été signalés comme « *non-perspectivés* ». Il peut s'agir de cadres « neutres » perspectivés par d'autres cadres, mais aussi — et c'est ce qui nous intéresse — de cadres dans lesquels on a mélangé des déclencheurs présentant des perspectives différentes sur le cadre. C'est par exemple le cas du cadre EDUCATION_TEACHING, qui mélange les déclencheurs comme *étudier* et *réviser* dont le sujet canonique est l'élève, et les déclencheurs comme *enseigner* et *éduquer*, dont le sujet canonique est l'enseignant. On rencontre aussi des cadres mélangeant des déclencheurs profilant différents rôles sans que cela soit signalé par un marquage de type « non-perspectivé ». C'est le cas du cadre CAUSATION, qui comporte à la fois des déclencheurs nominaux comme *cause* et *reason*, référant au remplisseur du rôle Cause, et d'autres comme *result*, qui réfère au remplisseur du rôle Effect.

Un autre point est que si l'on a besoin de cette information, une fois le corpus annoté, il sera possible d'utiliser les données pour retrouver les linkings canoniques des déclencheurs (et donc, si on le souhaite, de repérer au sein des cadres les groupes profilant le même rôle).

Enfin, il nous semble que l'idée de faire des cadres différents dès lors que les linkings sont différents ne correspond pas à l'esprit FrameNet : ce dernier se veut plus sémantique que syntaxique, allant même jusqu'à mélanger les catégories au sein de ses cadres. Pour nous, le profilage d'un rôle relève du même type de mécanisme de mise en avant que les constructions syntaxiques comme la passivation, or ces dernières ne donnent pas lieu à des changement de cadres.

Cette raison théorique s'ajoute aux trois points plus pratiques évoqués supra — à savoir que la notion de perspective est difficile à appliquer formellement à une bonne partie de nos déclencheurs, qu'elle n'est pas nécessaire pour capturer l'information sur les linkings, et que le critère n'a de toutes façons pas toujours été respecté dans le BFN — pour orienter notre décision. Nous avons choisi

de ne pas créer de nouvelles distinctions basées sur la perspective dans les cadres existants, et de ne pas non plus utiliser ce critère pour la délimitation des nouveaux cadres. En revanche, nous ne neutraliserons pas la distinction dans les cadres pour lesquels elle a été utilisée, maximisant ainsi notre compatibilité avec les données BFN. Cela implique que dans les données French FrameNet cohabiteront des cadres distingués sur la base du linking canonique de leurs déclencheurs (comme COMMERCE_BUY et COMMERCE_SELL), et des cadres comme CAUSATION pour lesquels ce n'est pas le cas.

4.2.4.2 Critères de délimitation retenus

À présent que nous avons reprécisé les notions qui nous paraissaient peu claires, il est temps de présenter les critères d'appartenance à un même cadre retenus pour le French FrameNet.

Notre premier critère est celui selon lequel les déclencheurs d'un cadre doivent partager un noyau de sens important. C'est d'ailleurs lui qui nous poussera à tenter de rassembler des déclencheurs potentiels en un même cadre. Ce critère reste délibérément assez vague, ce qui laisse une certaine liberté lors de la conception concernant le degré de granularité sémantique des cadres.

Maintenant que les critères de nucléarité d'un rôle ont été redéfinis, nous pouvons établir que les déclencheurs d'un même cadre devront avoir le même nombre et type de rôles noyaux, et que ces rôles devront entretenir les mêmes relations entre eux. Nous modulons un peu ces deux règles : nous imposons seulement que chaque déclencheur du cadre sous-catégorise au moins un des rôles d'un ensemble nucléaire ou ensemble en exclusion mutuelle. Cela crée de la souplesse, et évite de multiplier les cadres, puisqu'on permet formellement d'avoir dans le même cadre des prédicats qui n'ont pas exactement le même jeu de rôles. Ainsi, *s'apercevoir* et *détecter* évoquent tous deux le cadre FR_BECOMING_AWARE, bien que comme on peut le voir dans les exemples fabriqués en (47), seul *détecter* permette l'expression du rôle Instrument (qui est en ensemble nucléaire avec le rôle Cognizer).

(47) a. Alex a **DÉTECTÉ** plusieurs erreurs.
b. Le programme a **DÉTECTÉ** plusieurs erreurs.
c. Alex **S'**est **APERÇU** de plusieurs erreurs.
d. *Le programme **S'**est **APERÇU** de plusieurs erreurs.
FR_BECOMING_AWARE: Cognizer, Evidence, Instrument, Phenomenon, Topic

Comme nous l'avons annoncé en 4.2.4.1, nous ne retenons pas le critère selon lequel tous les déclencheurs d'un même cadre doivent « profiler » le même rôle. Cela est principalement dû au fait que nous ne parvenons pas à définir cette notion de manière à la fois satisfaisante et assez formelle pour permettre d'effectuer des regroupements valides.

Le critère concernant le partage des présuppositions, attentes et concomitants est de même relaxé. Nous avons déjà mentionné qu'il nous semblait que ce critère n'était pas toujours respecté dans le BFN, et il ne nous a pas semblé souhaitable ni pratique de rétablir cet état de fait dans le French FrameNet. Nous avons évoqué l'idée d'utiliser des traits sur les déclencheurs pour signaler une présupposition notable mais nous laissons cela pour de futurs travaux.

Enfin, nous avons conservé l'exigence que les déclencheurs d'un cadre aspectuellement complexe doivent partager un ensemble d'implications pour tout ce qui est des étapes et des transitions qui les constituent. Ainsi, *arguer* et *prouver* évoquent tous deux un locuteur parlant à un interlocuteur afin de le convaincre de la véracité de quelque chose, mais seul *prouver* implique que ce but a été atteint. Ces verbes évoquent donc des cadres différents (respectivement FR_ATTEMPT_SUASION.VERACITY et FR_PROVING).

4.2.4.3 Rôles et remplisseurs à annoter au niveau du cadre

Nous avons dit en 4.2.3.2 que pour un déclencheur donné, les rôles considérés étaient ceux pouvant être réalisés par des arguments sous-catégorisés par le déclencheur — ceux de ces arguments qui sont sémantiquement obligatoires ont le statut de rôles noyaux. Cela dit, pour une occurrence de déclencheur, nous annotons les rôles même lorsqu'ils sont réalisés par un dépendant non sous-catégorisé ou autre élément non local. Nous signalons ces cas de non-localité à l'aide d'un marquage spécial. Ainsi, on peut voir en (48a) que le rôle Speaker du cadre FR_COMMITMENT.TO_DO peut-être exprimé comme sujet (sous-catégorisé et sémantiquement obligatoire, donc rôle noyau) du verbe *promettre*. Lorsque ce même rôle est rempli par un élément non local au déclencheur, comme c'est le cas de *Volkswagen* en (48b), on l'annote quand même. De même pour les rôles non noyaux, bien qu'on puisse voir en (49a) que le rôle Addressee du cadre FR_STATEMENT-MANNER-NOISE peut être réalisé sous forme d'objet indirect (donc sous-catégorisé) du verbe *dire*, ce rôle n'est pas sémantiquement obligatoire (*dire quelque chose* n'implique pas nécessairement de *dire quelque chose à quelqu'un*). Il ne s'agit donc pas d'un rôle noyau, mais on peut voir en (49b) que lorsqu'il est réalisé non localement au déclencheur, ce rôle est tout de même annoté. Notons ici que ces cas se sont révélés rares.

(48) a. En contrepartie, la Banque centrale **PROMET** de baisser immédiatement ses taux.

b. Tout en **PROMETTANT** de faire revivre la firme sans débaucher, Volkswagen a dû licencier en masse et Skoda est aujourd'hui au bord de la faillite.

FR_COMMITMENT.TO_DO: Addressee, Entity, Eventuality, Medium, Speaker, Topic

(49) a. Vous me **DIREZ** que si l'un téléphone, l'autre peut écrire... et vice versa.

b. Je me suis même laissé **DIRE** qu'ils constituaient facilement les trois quarts de la clientèle... UNI

FR_STATEMENT-MANNER-NOISE: Addressee, Medium, Message, Speaker, Topic

Passons maintenant au niveau du cadre pour la description de ce qui sera ou non annoté. Nous venons de le voir en 4.2.4, pour un cadre donné, si un rôle est noyau, chacun des déclencheurs du cadre le sous-catégorise, ou sous-catégorise un rôle en ensemble nucléaire ou en exclusion avec lui. Cela laisse la possibilité qu'un rôle noyau, faisant partie d'un ensemble nucléaire, ne soit pas réalisable localement pour tous les déclencheurs d'un cadre mais seulement pour certains. Pour une instance de cadre évoquée par un déclencheur, si un rôle noyau du cadre est rempli par un dépendant qui ne peut jamais être sous-catégorisé par le déclencheur, nous annotons tout de même cet élément comme remplisseur du rôle (avec un marqueur de non localité). Pour rendre ce choix plus clair, présentons ici l'exemple du cadre FR_ATTRIBUTING_CAUSE, dans lequel un protagoniste (Attributor) explique une situation (Known_state_of_affairs) par une cause (Cause), ou présente une entité (Responsible_party) comme responsable de cette situation. Les rôles Cause et Responsible_party, tous deux noyaux, forment un ensemble nucléaire. On peut voir en (50a) et (50b) que le verbe *attribuer*, qui déclenche ce cadre, permet de réaliser chacun des deux rôles sous forme d'objet indirect. Les exemples fabriqués (51a) et (51b) nous montrent que pour le verbe *expliquer*, le rôle Cause est sous-catégorisé alors que le rôle Responsible_party n'est pas exprimable localement. Le choix transcrit dans le paragraphe précédent nous amène à, dans la phrase (51c), annoter le rôle Responsible_party comme rempli par le pronom *la* (qui réfère à *l'obésité*), alors même que ce rôle n'est pas exprimable localement au déclencheur *expliquer*.

(50) a. Enfin, une majorité des sondés **ATTRIBUENT** aux médias la montée de M. Jean-Marie Le Pen (69 %), l'impopularité de Mme Edith Cresson (56 %) ou la dégradation de l' image du monde politique (57 %).

b. On ne se trompe sans doute guère en **ATTRIBUANT** l'essentiel de cet accroissement à l'octroi systématique de subventions.

FR_ATTRIBUTING_CAUSE: Attributor, Cause, Known_state_of_affairs, Responsible_party

(51) a. Les sociologues **EXPLIQUENT** l'accroissement de la délinquance par une forte augmentation de l'obésité.

b. *Les sociologues **EXPLIQUENT** l'accroissement de la délinquance par l'obésité.

c. D'autres ont craint des effets de panique et d'anxiété, tandis que les sociologues la pointent du doigt pour **EXPLIQUER** l'accroissement de la délinquance.

FR_ATTRIBUTING_CAUSE: Attributor, Cause, Known_state_of_affairs, Responsible_party

L'annotation d'une réalisation non locale est nécessairement plus sujette à subjectivité que celle d'un remplisseur sous-catégorisé puisque la syntaxe n'y est d'aucune aide. Néanmoins, le choix d'inclure ces réalisations permet une meilleure cohérence au sein des annotations d'un cadre, et comme les annotations FrameNet peuvent entre autres être utiles pour observer les restrictions de sélection lexicales, il nous semble intéressant d'obtenir le plus possible d'information annotée. Les réalisations non locales pourront de toutes façon être filtrées, puisque nous verrons au chapitre 7 (section 7.1.2.3) qu'elles sont signalées par un marquage spécial.

4.2.5 Point sur les relations entre cadres dans le French FrameNet

Nous avons mentionné en 4.2.4.1 que pour le French FrameNet, nous ne créerions pas de nouvelles relations de perspective entre cadres. Nous profitons de ce chapitre pour préciser que nous avons choisi de ne pas utiliser deux autres types de relations entre cadres proposées par BFN : les relations *using* et *see also*.

La relation BFN *using* ne nous a pas semblée suffisamment bien définie pour être utilisée de manière cohérente. Pour Ruppenhofer *et al.* (2006), dans la plupart des cas, si un cadre F1 *utilise* un cadre F2, une partie de la scène évoquée par F1 réfère à F2. Cette description peut s'appliquer à une variété de situations, dont nous avons décrit une partie en 3.2.1.4. Cette variété fait que la relation *using* nous semble plus vague, moins contrainte que les autres relations. De plus, Ovchinnikova *et al.* (2010) précisent que dans les données BFN, elle est souvent utilisée pour lier deux cadres dans des cas où une autre relation aurait pu convenir. Nous avons donc décidé de ne pas utiliser cette relation dans le French FrameNet.

Une relation appelée *see also* permet de repérer les cadres qui présentent des similarités pour pouvoir les comparer. Ruppenhofer *et al.* (2006) indiquent que cette relation n'implique aucune relation particulière entre les cadres concernés, c'est d'ailleurs pourquoi nous ne l'avons pas évoquée en 3.2.1.4. Elle est d'un autre ordre que les « véritables » relations entre cadres : il s'agit simplement de faciliter la lecture des cadres par un utilisateur en pointant vers des cadres présentant une différence clé avec le cadre courant. Cette relation est asymétrique : pour un groupe de cadres similaires, la définition d'un des cadres contraste entre eux les différents cadres du groupe, tandis que chacun des autres pointe vers ce cadre à l'aide de la relation *see also*. Pour le French FrameNet, nous avons choisi de préciser dans la définition de chaque cadre l'existence de cadres similaires et ce qui les différencie du cadre courant. Ainsi, il n'y a pas besoin d'aller consulter la définition de chacun des cadres pour être en mesure de comprendre leurs sens distincts et la frontière qui les sépare.

Le French FrameNet utilise toutes les autres relations FrameNet. Nous avons néanmoins choisi de préciser les contraintes sur l'héritage qui apparaissent dans la littérature FrameNet. Pour des

cadres F1 et F2 unis par une relation d'héritage telle que *F1 hérite de F2* : plutôt que de devoir hériter de tous les rôles sauf extra-thématiques de F2, dans le French FrameNet, F1 ne devra hériter que de ses rôles noyaux — qui devront aussi être noyaux pour F1. Nous conservons la règle selon laquelle il est autorisé pour F1 de n'hériter que d'un des rôles en relation d'exclusion mutuelle ou en ensemble nucléaire dans F2. Si plusieurs de ces rôles sont hérités, ils devront partager les mêmes relations dans F1 et F2.

Nous introduisons par ailleurs deux nouvelles relations permettant de lier deux cadres. Le fait de travailler à la fois sur le domaine des positions cognitives et celui de la communication langagière nous a permis de repérer que certains cadres semblaient correspondre à une verbalisation de certains autres. Ainsi PREDICTING (un locuteur annonce qu'un évènement va se produire dans le futur — *prédire, pronostiquer*...) consiste en une communication d'EXPECTATION (un être doué de conscience croit qu'un évènement va se produire dans le futur — *prévoir, pressentir*...) : *Alex prédit qu'il va faire beau* a grossièrement pour sens *Alex dit : « je prévois qu'il va faire beau »*. Nous avons repéré ce même type de situation avec plusieurs paires de cadres, et avons donc décidé de les marquer à l'aide d'une nouvelle relation : *communiqué par/communicatif de*. EXPECTATION est communiqué par PREDICTING.

La deuxième relation que nous introduisons permet de lier les cadres comme JUSTIFYING et DESERVING, qui ne sont dans le BFN liés par aucune relation. Dans le premier de ces cadres, un agent fournit une raison visant à légitimer un acte qu'il a commis ou omis, ou une situation dont il est responsable (*Alex justifie son départ par l'heure tardive*), tandis que dans le deuxième, l'existence d'une situation est une raison suffisante pour qu'une action soit considérée comme justifiée (*L'heure tardive justifie le départ d'Alex*). Il nous semble que JUSTIFYING correspond à l'attribution d'un évènement DESERVING à un locuteur : *Alex justifie son départ par l'heure tardive* correspond à *Alex considère que l'heure tardive justifie son départ*. Nous avons donc pour ce type de cas défini la relation d'*attribution* : DESERVING est attribué par JUSTIFYING. Nous avons utilisé cette relation pour une seule autre paire de cadres : FR_ATTRIBUTING_CAUSE, où un locuteur explique une situation en indiquant sa cause, est attributif de EXPLAINING_THE_FACTS, dans lequel une situation est présentée comme étant causée par un certain fait.

Enfin, pour le French FrameNet, nous avons décidé de ne conserver que les relations entre cadres de nos domaines, et de supprimer celles avec les cadres BFN n'appartenant pas au French FrameNet. Si ces cadres avaient été adaptés pour le français, ils auraient possiblement été modifiés. Sans connaissance de ces potentielles modifications, il nous est impossible de vérifier que les contraintes sur les relations entre cadres sont respectées. Il nous semble donc préférable de ne pas les intégrer au réseau de relations. Notons néanmoins que ces relations sont récupérables : pour chaque cadre spécifique au French FrameNet, une relation spéciale permet de repérer le cadre BFN dont il est issu lorsque c'est pertinent. Ce lien avec les cadres originaux de FrameNet est crucial pour le projet à l'étude de base de données multilingue FrameNet (cf. section 9.3 sur les perspectives de travail).

Les notions nécessaires à la création du French FrameNet ayant été redéfinies, les trois chapitres qui suivent sont consacrés à la présentation de la création de la ressource. Ce travail s'est effectué en trois étapes : la sélection et adaptation de cadres BFN pour nos domaines, le travail sur le lexique de déclencheurs de ces cadres, et enfin, l'annotation en corpus d'occurrences de ces déclencheurs. La première étape consiste à délimiter nos domaines en termes de cadres, en examinant les cadres existants et en les modifiant ou en en créant de nouveaux le cas échéant. La deuxième est la création d'un lexique pour les cadres préalablement sélectionnés. Il s'agit d'associer à chaque cadre un ensemble de lemmes susceptibles de le déclencher. Ces deux étapes sont décrites chapitre 5, et appliquées chapitre 6, tandis que le chapitre 7 traite de la troisième étape : l'annotation en corpus de déclencheurs de cadres et de leurs remplisseurs de rôles.

CHAPITRE 5

CADRES ET LEXIQUE

Sommaire

5.1 La polysémie **86**
5.1.1 Les tests 87
5.1.1.1 Test logique 87
5.1.1.2 Critère d'interprétation multiple 88
5.1.1.3 Test de co-prédication (= critère de cooccurrence compatible) 89
5.1.1.4 Test d'identité 90
5.1.1.5 Indépendance des relations lexicales 90
5.1.1.6 Critère de dérivation différentielle 90
5.1.1.7 Critère de cooccurrence différentielle 91
5.1.1.8 Le test définitionnel 92
5.1.1.9 Interaction entre les tests 92
5.1.2 Application à des cas d'école 93
5.1.2.1 *Convaincre* et *dissuader* (SUASION vs. EVENTIVE_COGNIZER_AFFECTING) 93
5.1.2.2 *Convaincre* et *persuader* (FR_COGNIZER_AFFECTING.VERACITY vs. FR_COGNIZER_AFFECTING.TO_DO) 94
5.1.2.3 Constater et remarquer (BECOMING_AWARE vs. STATEMENT) . 96
5.2 Travail conjoint cadres et lexique **97**
5.2.1 Modélisation d'un domaine 98
5.2.1.1 Définition et tests d'appartenance 98
5.2.1.2 Travail sur les cadres 100
5.2.2 Lexique des cadres 102
5.2.2.1 Lexiques automatiques 102
5.2.2.2 Travail sur ces lexiques 103

Nous appellerons *lexique d'un cadre* l'ensemble des mots pouvant évoquer ledit cadre. Employé seul, le mot lexique (comme dans « le lexique du projet ») désigne l'ensemble des couples déclencheur-cadre d'une ressource de type FrameNet.

Comme nous l'avons vu au chapitre 4, la première partie du projet ASFALDA consiste en la spécification d'un sous-ensemble de cadres pour lesquels le corpus serait plus tard annoté. Le plan initial du projet prévoyait grossièrement d'effectuer le travail de préparation de l'annotation en deux parties, la première concernant la modélisation en cadres des domaines choisis, et la deuxième la création du lexique de ces cadres. Nous allons voir que nous avons fini par effectuer la modélisation des domaines en même temps que le lexique, puisque chacune de ces étapes a influé sur l'autre

et que nous avons effectué de nombreux allers-retours entre les deux. Notons aussi que le travail d'annotation, qui sera décrit au chapitre suivant, a aussi amené des modifications, mais cela a été plus marginal. Ce sont ces étapes concomitantes de modélisation des domaines et de lexique des cadres qui sont décrites dans ce chapitre, après une première partie destinée à expliciter la notion de polysémie, à laquelle nous avons souvent fait appel.

5.1 La polysémie

Nous présentons en 5.2 le travail de délimitation des domaines en cadres et celui de création du lexique de ces cadres. Nous y verrons qu'il nous est arrivé de modifier des cadres BFN, et que le travail sur le lexique peut parfois amener à fusionner des cadres ou à en diviser. Il nous intéresse ici de voir qu'au cours du travail sur le lexique, nous avons souvent rencontré trois types de situations problématiques.

Granularité trop fine de deux cadres (potentielle polysémie artificielle) Deux cadres BFN nous semblent pouvoir être évoqués par le même ensemble de déclencheurs, ou tout du moins présentent un bon nombre de déclencheurs communs. C'est le cas que nous étudions en 5.1.2.1.

Couverture du cadre insuffisante Au sein d'un cadre, plusieurs ou tous les déclencheurs ne peuvent être annotés convenablement à l'aide des rôles du cadre. Les déclencheurs en question peuvent généralement couvrir un sens proche de celui du cadre, et ce sens ne doit pas être couvert par un autre cadre (on aurait alors affaire au problème de granularité trop fine). Nous étudions un tel cas en 5.1.2.2.

Granularité trop grossière d'un cadre À l'examen d'un déclencheur potentiel en contexte, il est difficile d'évaluer si différentes occurrences partagent le même sens, et donc évoquent ou non le même cadre. C'est le problème qui apparaît en 5.1.2.3.

Ces trois cas de figure posent tous le problème de l'identification du statut mono- ou polysémique d'un mot. Si pour un (ou un ensemble de) déclencheur(s) on a bien affaire à plusieurs sens, on va pour le problème de couverture du cadre insuffisante créer un cadre adapté au deuxième sens, tandis qu'on se contentera de clarifier le guide d'annotation pour les deux autres problèmes. En revanche, si on n'a affaire qu'à un seul sens, on fusionnera les cadres impliqués dans le problème de granularité trop fine de deux cadres, élargira le rôle trop étroit ou créera un rôle supplémentaire dans le problème de couverture du cadre insuffisante, et gardera le déclencheur uniquement dans le lexique d'un des cadres pour le problème de granularité trop grossière. On a donc besoin de tests de polysémie pour pouvoir apporter une réponse cohérente à ces problèmes qui touchent à la fois à l'appariement déclencheurs-cadre, à la modification et création de cadres, et à la désambiguisation en contexte lors de l'annotation.

Comme nous l'avons mentionné section 3.3.2.2, Fillmore (2007) aborde rapidement le problème de la potentielle polysémie des déclencheurs : pour lui, l'entreprise FrameNet a propension à multiplier les sens plutôt qu'à les amalgamer. Il mentionne que BFN cherche à maximiser la régularité des points suivants pour les déclencheurs d'un même cadre :

— relations à autres lexèmes (antonymie, synonymie…)
— relation déclencheurs/patrons de valence
— réalisations syntaxiques particulières
— possibilités d'inférence différentes
— ambiguités non associées au type de complément
— relations morphologiques à des mots de la même famille

C'est donc ces points qui sont examinés pour décider si tous les emplois d'un lexème doivent ou non être rangés dans le même cadre (si ce n'est pas le cas, on a affaire à plusieurs sens). Plusieurs exemples de cas de délimitation de plusieurs sens distincts pour un même lexème sont présentés,

mais nous n'avons trouvé nulle part dans la littérature FrameNet de procédure détaillée pour déterminer le statut (polysémique ou non) d'un lexème.

5.1.1 Les tests

NB Pour décider des potentielles modifications de périmètre des cadres, des fusions et des divisions, nous présentons dans cette section et celle qui suit des méthodes et exemples d'étude fine de polysémie. Il est néanmoins important de noter qu'avant de recourir à ces tests nous avons d'abord tâché d'employer des ressources externes. Ainsi, le travail lexicographique déjà effectué par les créateurs et créatrices du Réseau Lexical Français (RLF) de Lux-Pogodalla et Polguère (2011) aurait pu nous éviter d'avoir à effectuer nous même la délimitation des sens. Nous avons eu accès à une visualisation de cette ressource avant sa sortie effective, mais avons buté sur le fait que le statut du travail sur les lexies n'était pas accessible. Il nous a donc été impossible de déterminer si une absence de distinction de sens signifiait absence de polysémie ou bien travail en cours. C'est pourquoi nous avons choisi de nous pencher nous même sur la notion de polysémie.

La notion de polysémie, qui a généré une très importante littérature, est en elle-même une notion controversée. Nous n'examinerons pas ici de manière poussée la littérature concernant les problèmes de polysémie : nous avons seulement besoin d'outils pour créer des frontières de cadre, sans nous positionner sur la question de s'il y a ou non réellement polysémie. Cette section est donc consacrée à l'examen des tests de polysémie inventoriés dans la littérature et à leur applicabilité au problème qui nous occupe. Geeraerts (1993) comme Lewandowska-Tomaszczyk (2007) distinguent trois types de tests : le test logique, les tests linguistiques et le test définitionnel. Par opposition, Cruse (2000) inventorie les tests en fonction de la propriété qu'ils testent selon lui : le test linguistique du zeugme teste l'antagonisme de deux lectures, tandis que les autres testent la discrétion des lectures. À la suite de Goossens (2011), nous lions aussi les tests aux « critères de délimitation des lexies » de Mel'čuk *et al.* (1995), qui se présentent encore différemment mais correspondent grosso modo aux mêmes concepts et peuvent apporter un complément d'information.

Nous détaillons ici les différents tests avant de préciser comment nous les utiliserons en interaction les uns avec les autres, et notamment en cas de conflits.

5.1.1.1 Test logique

Si une assertion employant un mot peut être vraie et fausse du même référent, alors le mot est polysémique. Si *X et non X* est vraie, alors *X* possède au moins deux sens distincts. L'exemple (52) est acceptable si on interprète un des *bar* comme le lieu et l'autre comme le meuble.

(52) C'est un bar mais ce n'est pas un bar. [1]

Cette propriété correspond à ce que Cruse (2000) appelle *conditions de vérité indépendantes*; il teste cette indépendance en créant une situation hypothétique pour laquelle la réponse à une question fermée utilisant le mot à tester pourrait à la fois être oui et non. *Est-ce que c'est un bar ?*

Geeraerts (1993) attribue la paternité du test à Quine (1960), et évoque les cas dans lesquels le test n'est pas si simple à appliquer : ceux dans lesquels l'assertion *X* utilisée par le test est vraie pour une lecture, ininterprétable pour l'autre (plutôt que vraie pour l'une et fausse pour l'autre). Pour reprendre l'exemple de Geeraerts, dans *this book is sad*, la lecture de *sad* comme *qui ressent de la tristesse* est impossible car *book* viole les restrictions de sélection de ce sens de *sad* (on attend un être doué de conscience). L'assertion peut donc être vraie ou fausse pour le sens *qui cause de la tristesse* de l'adjectif *sad*, mais sera ininterprétable pour l'autre sens. L'application du test logique est, pour Geeraerts, impossible dans ce type de cas. Nous allons voir que ce problème correspond à

1. Repris de l'exemple (13) de Goossens (2011, p. 12).

des situations dans lesquelles nous souhaiterions être en mesure d'appliquer le test logique mais ne le pourrons pas.

Dans la littérature, on trouve la plupart du temps les tests appliqués à des noms. Pour les besoins de notre travail, nous cherchons à ce qu'ils le soient aussi à des verbes, puisque la plupart de nos déclencheurs sont des verbes. Pour construire le test logique avec un verbe, il faut coordonner deux occurrences de ce verbe (dont l'une niée) avec les mêmes arguments. Si l'un des arguments n'est pas partagé, la phrase pourra être acceptable quel que soit le statut mono- ou polysémique du verbe : *Alex mange une pomme et ne mange pas une poire* est satisfaisable, pourtant, le verbe manger n'est pas polysémique. Cela paraît évident mais pose problème dans les cas où les deux sens potentiels ne peuvent pas apparaître avec un même type d'argument. C'est le cas par exemple du verbe *amuser* dont on pourrait postuler un sens où *un humain ou un autre être capable de volition effectue des actions dans le but de faire rire quelqu'un* et un sens dans lequel *une situation fait rire quelqu'un*. Coordonner les deux sens du verbe nécessite l'emploi d'arguments différents, puisque chacun des sens a des restrictions de sélection différentes.

Construire une version de la phrase où les arguments sont omis ne fonctionne pas non plus. En reprenant l'exemple des deux sens postulés du verbe *amuser*, on utilise le passif pour pouvoir omettre le sujet dans la forme « question » du test. *Les enfants ont-ils été amusés ?* On peut effectivement répondre à cette question par oui et par non dans une même situation, mais pour ce faire, on reconstruit nécessairement l'argument manquant : *oui, les enfants ont été amusés par une situation* mais *non, les enfants n'ont pas été amusés intentionnellement par une personne*. Cela revient donc à appliquer le tests avec des arguments différents. Le test logique ne permet donc pas du tout de tester de candidats polysèmes dont les restrictions de sélection diffèrent, ce qui correspond aussi aux conclusions de Geeraerts (1993) que nous avons évoquées plus haut.

Cette non applicabilité est encore plus évidente pour les verbes dont les deux lectures sont associées à une sous-catégorisation différente. C'est le cas par exemple du verbe *jouer*, pour lequel nous considérons ici les sens *pratiquer (par exemple un instrument)* et *interpréter (par exemple un rôle ou une pièce)*. Le premier prend pour argument un syntagme prépositionnel en *de* (*jouer du piano*) tandis que le deuxième prend pour objet un syntagme nominal (*jouer Macbeth*). Le type syntaxique d'argument sous-catégorisé par chacun des sens étant différent, il n'est pas possible de coordonner deux occurrences de *jouer* avec les mêmes arguments et d'avoir accès à une lecture différente pour chacune. Comme plus haut avec les types sémantiques, la question *Est-ce que la troupe a joué ?* oblige à reconstruire un énoncé avec des arguments différents *oui, elle a joué une pièce mais non, elle n'a pas joué d'un instrument*. Les candidats polysèmes à la sous-catégorisation différente excluent donc eux aussi le test logique, mais la différence de sous-catégorisation sera capturée comme signe de polysémie par le critère de cooccurrence différentielle.

5.1.1.2 Critère d'interprétation multiple

Geeraerts (1993) utilise un exemple de (Cruse, 1982, p. 70) pour présenter une variante du test logique : on peut lire (53) et comprendre que Charles s'est déplacé dans l'espace, ou alors qu'il a changé d'avis, mais les deux lectures sont impossibles simultanément.

(53) Charles has changed his position.
« Charles a changé de position. »

Pour Cruse (2000), il s'agit là d'un effet de l'antagonisme des deux lectures, qui sont donc en compétition et ne peuvent être traitées qu'une seule à la fois. Cela correspond au critère 1 de Mel'čuk *et al.* (1995), qui présentent cinq critères de délimitations des lexies. Ce critère, dit d'*interprétation multiple*, spécifie que si l'on peut, à partir d'un mot, créer un énoncé ambigu qui peut donner lieu à deux interprétations distinctes mais ne pouvant apparaître en même temps, c'est un indice pour

considérer qu'on a affaire à deux sens distincts [2].

Néanmoins, tel que formulé, ce test permet d'établir l'existence de sens multiples pour des mots clairement monosémiques. Ainsi, mettons que l'on se demande si *manger* peut avoir deux sens, « manger rapidement/dévorer » et « manger lentement/savourer ». À la lecture de (54a), on peut comprendre que (54b) est vraie ou que (54c) est vraie mais pas les deux en même temps. Il nous semble donc que pour être utilisable, ce test doit être bordé : il ne peut s'appliquer qu'aux lectures pour lesquelles on ne trouve pas de définition subsumante (cf. le test définitionnel que nous présenterons un peu plus bas).

(54) a. Alex a mangé une pomme.
b. Alex a dévoré une pomme.
c. Alex a savouré une pomme.

5.1.1.3 Test de co-prédication (= critère de cooccurrence compatible)

La notion d'antagonisme des sens avec laquelle Cruse (2000) explique le test précédent donne lieu au test de co-prédication. Selon ce test, forcer plusieurs sens d'un mot sur une même occurrence de ce mot résulte en un zeugme ou effet comique : Cruse en tire la conséquence que le zeugme peut donc être symptome de polysémie. Ainsi (55) et (56) produisent-elles un effet de zeugme alors que ce n'est pas le cas de (57).

(55) ?Arthur and his driving licence expired last Thursday. *(Cruse, 1982)*
« ?Arthur et son permis de conduire ont expiré jeudi dernier. »

(56) ?Alex loue un appartement et tous les dieux.

(57) Sam mange une pomme et une poire.

Le test de co-prédication correspond au critère de cooccurrence compatible de Mel'čuk *et al.* (critère 3), mais ces derniers sont plus restrictifs : le test fonctionne pour eux uniquement dans le sens *l'existence d'un exemple non zeugmatique pointe vers un seul sens*, et pas *l'existence d'un exemple zeugmatique pointe vers plusieurs sens*. Ainsi, pour eux, s'il est possible de créer une phrase à cooccurrence compatible des deux lectures qui n'est pas zeugmatique, alors on a affaire à un seul sens dont la définition consiste en une disjonction des deux lectures, mais si l'on ne peut pas créer de phrase non zeugmatique, le critère reste sans effet.

Cruse (2000) souligne que ce test est difficile à appliquer, notamment parce que les jugements sur la qualité zeugmatique ou non d'un énoncé ne sont pas toujours clairs. De notre côté, le jugement de normalité d'un énoncé étant parfois sujet à discussion, nous n'utiliserons ce test que dans les cas comme (58) où l'énoncé est carrément inacceptable ou ininterprétable.

(58) #Alex loue un appartement pour se faire de l'argent et pour avoir un endroit où vivre.

Nous ajouterons néanmoins la contrainte que le contexte utilisé pour forcer les deux lectures sur le mot ne présente pas d'arguments incompatibles, afin que des exemples comme (59) ne fassent pas deux sens de *manger*.

(59) #Alex mange une pomme à toute vitesse et très lentement

2. Ce critère doit pour Mel'čuk *et al.* être appliqué en conjonction avec les autres critères, en particulier le critère de cooccurrence compatible présenté supra, pour prendre une décision sur la délimitation d'une ou plusieurs lexies.

5.1.1.4 Test d'identité

Lewandowska-Tomaszczyk (2007), comme Mel'čuk *et al.* (1995), présente le test d'identité qui va suivre comme une variante du test de co-prédication supra. En revanche, pour Cruse (2000), le test de co-prédication témoigne de l'antagonisme des lectures tandis que le test d'identité témoigne de leur discrétion.

Le test peut se définir ainsi : dans une phrase employant le mot dont on soupçonne qu'il a deux sens, puis une expression anaphorique (ou une ellipse), si le mot et l'expression qui le reprend renvoient forcément au même sens, c'est un indice pour considérer que le mot a bien deux sens distincts.

(60) Alex loue un appartement à Paris et Sam (le fait) à Lyon.

(61) Ce groupe comprend les mathématiques et il comprend aussi la logique.

(62) Alex mange une Golden et une Williams.

Pour interpréter (60), on est obligé de considérer qu'Alex et Sam sont soit propriétaires, soit locataires d'appartements. La phrase devient inacceptable si Alex est propriétaire à Paris et Sam locataire à Lyon (ou l'inverse). De même pour (61), le groupe doit soit à la fois maîtriser les mathématiques et la logique, soit à la fois contenir les deux disciplines. En revanche en (62), nos sens potentiels « manger (une pomme) » et « manger (une poire) » apparaissent chacun pour une occurrence de *manger* sans aucun problème, ce qui pointe bien vers un seul sens pour ce mot.

5.1.1.5 Indépendance des relations lexicales

Parmi les tests de discrétion de deux lectures, Cruse (2000) présente l'indépendance des relations lexicales : si deux sens potentiels d'un mot ont des ensembles de relations lexicales indépendants, cela pointe vers une polysémie. Ainsi le verbe *comprendre* a-t-il dans un de ses sens pour synonymes *appréhender* et *connaître* et pour antonymes *ignorer* et *méconnaître*, tandis que l'autre de ses sens a pour synonymes *inclure* et *comporter* et pour antonymes *exclure* et *omettre*.

Ce test nous paraît problématique car il est possible de le falsifier avec notre exemple des sens « manger vite » et « manger lentement » pour le verbe *manger* : manger (vite) a pour synonyme *dévorer*, tandis que manger (lentement) a pour synonyme *savourer*. Il nous faut donc soit fournir une restriction supplémentaire, comme *les synonymes trouvés ne doivent pas être co-hyponymes d'un mot qui serait aussi synonyme du mot étudié dans son sens le plus large*, ou *le test ne s'applique pas aux sens potentiels pouvant être couverts par une définition englobante* (cf. test définitionnel un peu plus bas).

On pourrait imaginer d'utiliser le test dans l'autre sens : si deux sens potentiels ne présentent pas d'ensembles indépendants de relations lexicales, alors il ne s'agit pas de sens distincts. Cela est néanmoins problématique, puisqu'on peut imaginer des cas de « trous » de lexique dans lesquels un mot n'aurait simplement pas de synonyme ni d'antonyme.

En ce qui nous concerne, nous utiliserons donc ce test dans le même sens que Cruse, mais uniquement dans les cas où les sens potentiels ne peuvent pas être couverts par une définition englobante. Ce test ne pourra donc en aucun cas pointer vers une autre direction que celle dans laquelle pointe le test définitionnel, et ne servira donc qu'à en renforcer les résultats.

5.1.1.6 Critère de dérivation différentielle

Mel'čuk *et al.* proposent aussi un critère de dérivation différentielle (critère 5) : si deux sens potentiels d'un même mot présentent des ensembles disjoints de dérivés morphologiques ou lexicaux, alors on a affaire à deux sens distincts.

C'est le cas par exemple pour *compter*, dont le sens « calculer » peut être nominalisé en *compte* alors que ce n'est pas le cas du sens « faire confiance ». De même pour *comprendre*, dont seul le sens cognitif donne compréhension.

5.1.1.7 Critère de cooccurrence différentielle

Le critère de cooccurrence différentielle de Mel'čuk *et al.* (critère 4) énonce que si deux sens potentiels d'un mot présentent deux ensembles disjoints de cooccurrents (morphologiques, syntaxiques ou lexicaux), les sens potentiels sont effectivement distincts. Kleiber (2008) illustre ce test avec le verbe *regarder*, dont le sens perceptif (63) peut être mis au passé, ainsi qu'au passif mais pas le sens locatif (64).

(63) a. Alex regarde la montagne.
b. Alex a regardé la montagne.
c. La montagne a été regardée par Alex.

(64) a. L'église regarde la montagne.
b. #L'église a regardé la montagne.
c. #La montagne a été regardée par l'église.

Comme le fait remarquer Goossens (2011), ce critère permet de prendre en compte « la combinatoire au niveau syntagmatique, avec un grand nombre de précisions importantes : type de propositions, nombre d'actants, etc. ». Cela inclut en particulier les cas où les deux sens potentiels présentent une différence de sous-catégorisation. Ainsi nous avons vu plus haut l'exemple du verbe *jouer* : dans sa lecture qui peut être paraphrasée par *pratiquer*, il prend pour argument un syntagme prépositionnel en *de* (*jouer du violoncelle*), alors que dans sa lecture *interpréter*, il prend pour objet un syntagme nominal (*jouer "Un tramway nommé Désir"*). Cette possibilité du test de cooccurrence différentielle fait qu'il nous fournira de très utiles indices sur le statut polysémique ou non des verbes.

Ce test exige néanmoins d'être manié avec précaution. Par exemple dans (65), aussi tiré de Kleiber (2008), *intentionnellement* est de toutes façons exclu par les restrictions de sélection de *l'église* : l'agrammaticalité de (65b) ne devrait donc pas être utilisée pour déterminer deux sens distincts au verbe *regarder*. Une version alternative où un nom propre est utilisé au lieu de l'église est aussi difficile à interpréter : le fait que (65c) ne puisse être lue que comme un cas de *regarder* locatif pourrait simplement être dû au fait que la notion d'intentionnalité fait partie de ce sens, ce qui force l'interprétation en sa faveur. Il ne faudrait donc probablement pas faire porter ce test sur la présence de cooccurrents faisant partie intégrante de la définition d'un des sens postulés.

(65) a. Alex regarde intentionnellement la montagne.
b. #L'église regarde intentionnellement la montagne.
c. #Alex regarde-locatif intentionnellement la montagne.

Quoi qu'il en soit, pris à la lettre, ce critère nous paraît à risque de créer beaucoup de polysémie, dès qu'un mot présente par exemple une variation de sens liée à son association à des arguments de classes sémantiques différentes, potentiellement exprimés par des arguments syntaxiques différents. Nous avons par ailleurs constaté une pratique FrameNet de parfois définir au sein d'un même cadre des rôles de types sémantiques différents en relation d'exclusion mutuelle. Ainsi nous avons vu en 3.2.1.3 que pour beaucoup de cadres décrivant un évènement pouvant être déclenché par un agent (être sensible et intentionnel) ou une cause (situation), les deux type d'initiateur de l'évènement correspondaient à deux rôles en relation d'exclusion mutuelle. Ces deux raisons nous ont amenés au choix de n'utiliser le critère de cooccurrence différentielle que comme indice de la possible présence de plusieurs sens, en aucun cas suffisant pour justifier la présence d'un déclencheur dans deux cadres différents.

5.1.1.8 Le test définitionnel

Tous les auteurs examinés s'accordent sur l'existence du test définitionnel : un mot est polysémique si plus d'une seule définition est nécessaire pour rendre compte de son sens. Plus précisément pour Cruse (2000), l'absence de définition unifiée des deux lectures *sans disjonction* pointe vers deux sens discrets. Ravin et Leacock (2000a) formulent aussi ce test comme l'impossibilité de définir un seul ensemble de conditions nécessaires et suffisantes pour couvrir tous les concepts exprimés par le mot.

Cruse (2000) considère que le test définitionnel est un critère utile si on s'en sert avec les autres tests, mais que manié seul, il peut poser des problèmes, comme le fait que certains mots ont des disjonctions dans leur définition sans selon lui être polysémiques : une princesse est la fille d'un-e monarque ou l'épouse de l'enfant d'un-e monarque. Lewandowska-Tomaszczyk (2007) pointe quant à elle une critique de Geeraerts (1993) qui indique que dans les cas d'autohyponymie, où un des sens est inclus dans l'autre, le test définitionnel ne peut fonctionner puisque la référence plus restreinte est forcément incluse dans la définition plus large. C'est le cas par exemple pour le mot anglais *egg*, qui peut à la fois prendre le sens général de « gamète de nombreux animaux » (*ovule* en français) et le sens restreint de « gamète d'un oiseau, avec un jaune et une coquille » (*œuf* en français).

5.1.1.9 Interaction entre les tests

Goossens indique à propos de l'ensemble des tests que « la plupart des auteurs utilisent ces tests de manière cumulative ». Ces tests se recoupent de toutes façons au moins en partie : il paraît logique de supposer que deux sens aux ensembles disjoints de relations lexicales appellent deux définitions différentes, et seront donc associés à des cooccurrents, au moins lexicaux, propres à chacun des sens.

Les différents travaux examinés n'accordent pas toujours le même statut à tous les tests, mais tous s'accordent à dire qu'ils donnent souvent des résultats conflictuels. Tous les auteurs n'en tirent pas les mêmes conclusions, mais l'idée de l'existence d'un continuum allant du vague à la polysémie semble faire consensus. Cruse (2000), par exemple, définit des indices de distinction de sens parmi ceux vus supra : l'antagonisme est pointé par le critère d'interprétation multiple et le test de co-prédication, et le caractère discret par le test d'identité, le test logique, l'existence de relations lexicales indépendantes, et la distinction définitionnelle. À l'inverse, il propose des indices d'unité de sens pour deux sens potentiels : leur non-antagonisme, c'est-à-dire quand deux lectures opèrent simultanément, mais aussi le fait de pouvoir coordonner des prédicats sélectionnant les deux lectures sans effet de zeugme, comme c'est le cas en (66).

(66) This book is riveting —I wish the print wasn't so small.
« Ce livre est passionnant — je préfèrerais que sa police ne soit pas si petite. »

Cruse utilise cette idée d'indices pour considérer un spectre de distinctivité avec deux extrêmes : les deux lectures possèdent tous les indices de discrétion et d'indépendance et aucun indice d'unité *vs.* les deux lectures ne possèdent aucun indice de discrétion et d'indépendance et tous les indices d'unité.

Les critères de Mel'čuk *et al.* spécifient aussi une direction (distinction ou unité) dans laquelle utiliser les tests, mais sauf spécification explicite, chacun d'eux est supposé être suffisant, et les cas où les critères divergent ne sont pas mentionnés. La table 5.1 présente l'effet (unité ou distinction) de chaque résultat des tests pour Mel'čuk *et al.* comme pour Cruse.

Bien que les tests soient utilisés de manière à peu près similaire, on observe des différences dans plusieurs cas. Il nous faut néanmoins choisir un mode de fonctionnement pour décider de l'appartenance d'un lexème à un ou plusieurs cadres : nous utiliserons donc les tests comme des indices de polysémie plutôt que des conditions toutes nécessaires ou chacune suffisante. Les tests n'étant pas tous applicables dans les mêmes conditions, nous utiliserons pour chaque cas à traiter

Test	Résultat	Mel'čuk *et al.*	Cruse
Test logique	Une assertion employant le mot peut être vraie et fausse du même référent.	-	distinction
	Une assertion employant le mot ne peut pas être vraie et fausse du même référent.	-	-
Critère d'interprétation multiple	On peut créer un énoncé ambigu qui donne lieu a deux interprétations distinctes mais ne pouvant apparaître en même temps.	distinction	distinction
	Les deux lectures opèrent simultanément.	-	unité
Cooccurrence compatible	On peut coordonner des prédicats sélectionnant les deux lectures sans effet de zeugme.	unité	unité
	Coordonner des prédicats sélectionnant les deux lectures produit un effet de zeugme.	-	distinction
Test d'identité	Dans une phrase *X fait P avec L et X' fait P avec L aussi*, les deux P renvoient forcément au même sens.	-	distinction
	On peut créer une phrase *X fait P avec L et X' fait P avec L aussi* où le premier P réfère à un des sens et le deuxième à l'autre.	unité	-
Indép. des relations lexicales	Les deux sens potentiels ont deux ensembles disjoints de relations lexicales.	-	distinction
	Pas d'indépendance des relations lexicales.	-	-
Dérivation différentielle	Les deux sens potentiels ont deux ensembles disjoints de dérivés (morphologiques ou lexicaux).	distinction	-
	Pas de dérivation différentielle.	-	-
Cooccurrence différentielle	Les deux sens potentiels ont deux ensembles disjoints de cooccurrents (morphologiques, syntaxiques ou lexicaux).	distinction	-
	Pas de cooccurrence différentielle.	-	-
Test définitionnel	Une seule définition ne suffit pas à rendre compte du sens du mot.	distinction	distinction
	Une seule définition permet de rendre compte du sens du mot.	unité	-

TABLE 5.1: Application des tests de polysémie selon Mel'čuk *et al.* (1995) et Cruse (2000)

ceux qui nous paraîtront effectivement applicables. La table 5.1 nous permettra de voir en chacun des indices d'unité ou de distinction. Nous prendrons ensuite les décisions en fonction du côté vers lequel penche la balance. Nous nous laissons par ailleurs la liberté de prendre en compte des critères autres tels que la cohérence avec les données BFN.

5.1.2 Application à des cas d'école

Les tests de polysémie présentés supra nous ont guidés pour l'adaptation au français des cadres BFN. Dans cette section, nous présentons à titre d'exemple d'abord le cas des cadres EVENTIVE_COGNIZER_AFFECTING et SUASION, que nous avons été amenés à redécouper sur la base des tests de polysémie de leurs déclencheurs français, puis le cas de certains déclencheurs des cadres BECOMING_AWARE et STATEMENT dont les tests nous ont permis de mettre au jour qu'ils avaient bien deux sens distincts.

5.1.2.1 *Convaincre* et *dissuader* (SUASION vs. EVENTIVE_COGNIZER_AFFECTING)

BFN propose deux cadres, EVENTIVE_COGNIZER_AFFECTING et SUASION, qui décrivent respectivement les situations où un évènement convainc un être sensible de quelque chose ou de faire quelque chose et celles où un être sensible convainc délibérément un autre être sensible de quelque chose ou de faire quelque chose. Ils sont tous deux unis par une relation d'héritage où EVENTIVE_COGNIZER_AFFECTING est le parent de SUASION. Ces deux cadres peuvent en français être évoqués au moyen du même ensemble de déclencheurs : *apprendre, convaincant, convaincre,*

décider, *dissuader*, *dissuasion*, *persuader* et *persuasion*. Cette situation est pour nous un indice de polysémie artificielle, nous procédons donc à une étude fine de polysémie pour établir s'il convient pour le French FrameNet de fusionner ces deux cadres ou s'il y a effectivement lieu d'utiliser deux cadres distincts pour cet ensemble de mots. Nous appliquons donc ici les tests décrits supra sur deux mots de cet ensemble : *convaincre* et *dissuader*.

Tout d'abord, il est possible d'établir une définition couvrant les deux cas : *un élément provoque l'émergence d'une position cognitive chez un être sensible.*

Ensuite, comme nous l'avons vu supra, le **test logique** n'est pas applicable aux prédicats dont les deux sens supposés ne présentent pas les mêmes restrictions de sélection, or *convaincre* et *dissuader* demandent dans l'usage SUASION un sujet de type être sensible, et pour le sens EVENTIVE_COGNIZER_AFFECTING un sujet de type évènementiel.

Nous avons aussi dit dans notre description des tests que le critère d'interprétation multiple ne nous semblait pas applicable aux lectures pour lesquelles une définition subsumante existait. Comme le test définitionnel vient de nous le montrer, c'est le cas ici, ainsi nous n'appliquerons pas le critère d'interprétation multiple à ce cas.

Coordonner deux contextes forçant respectivement les interprétations SUASION et EVENTIVE_COGNIZER_AFFECTING comme nous l'avons fait en (67) ne nous semble pas provoquer de zeugme, et nous ne sommes pas parvenus à construire de phrase forçant la cooccurrence des interprétations et inacceptable pour cette raison. Il y a donc **cooccurrence compatible**.

(67) a. Sam et l'aspect du ciel ont convaincu Alex qu'il allait pleuvoir.
b. Sam et l'aspect du ciel ont dissuadé Alex de sortir.

Les exemples (68) nous permettent d'appliquer le **test d'identité** : les arguments *Sam* et le beau temps forcent dans ces deux phrases l'interprétation du verbe comme un de ses sens potentiels, puis le deuxième. L'acceptabilité de ces phrases semble indiquer, du moins selon Mel'čuk *et al.*, un signe d'unité.

(68) a. Le beau temps a convaincu Alex, et Sam l'a fait aussi.
b. Sam a dissuadé Alex, et la pluie l'a fait aussi.

Nous n'observons pas pour nos verbes de **relations lexicales** valables pour l'un des sens et pas l'autre.

Le nom *dissuasion* ne semble à première vue associé qu'au sens SUASION de *dissuader*, mais l'existence du terme *dissuasion nucléaire* montre que ce n'est pas vraiment le cas. Nous n'observons donc pas d'**ensembles disjoints de dérivés**.

Enfin, nous n'observons pas non plus pour nos verbes d'**ensembles disjoints de cooccurents**, à part des mots comme *intentionnellement* qui découle de la définition du sens SUASION et ne peut donc pas être utilisé pour le test.

Parmi les éléments mis au jour par ces tests, aucun ne pointe vers l'existence de deux sens distincts (un évènement convainc un être sensible de quelque chose vs. un être sensible convainc délibérément un autre être sensible de quelque chose) pour *convaincre* comme pour *dissuader*, mais plusieurs pointent vers l'unité de ces sens. Comme il en va de même pour les autres déclencheurs potentiels des cadres BFN EVENTIVE_COGNIZER_AFFECTING et SUASION, nous avons décidé de fusionner ces derniers en un même cadre : FR_COGNIZER_AFFECTING.

5.1.2.2 *Convaincre* et *persuader* (FR_COGNIZER_AFFECTING.VERACITY vs. FR_COGNIZER_AFFECTING.TO_DO)

En observant les emplois en corpus des déclencheurs de notre nouveau cadre, on s'aperçoit qu'ils peuvent être divisés en deux groupes : ceux pour lesquels l'être sensible forme une position

sur la véridicité d'un contenu mental, et ceux pour lesquels cette position porte sur la détermination de l'être sensible à effectuer une action. On ne retrouve pas les mêmes constructions dans les deux groupes : les emplois portant sur l'opportunité d'effectuer une action sont complétés par des propositions infinitives et complétives au subjonctif tandis que ceux portant sur la véridicité d'un contenu propositionnel sont associés aux complétives à l'indicatif et aux syntagmes prépositionnels. De plus, certains déclencheurs du cadre ne permettent qu'un seul de ces sens potentiels : on *dissuade* forcément quelqu'un de faire quelque chose (influence sur la détermination à effectuer une action), tandis qu'on peut seulement apprendre à quelqu'un que quelque chose est vrai ou faux (influence sur la position quant à la véridicité d'un contenu).

Pour déterminer si deux cadres distincts devraient rendre compte de cette différence, nous avons donc appliqué nos tests à deux déclencheurs du cadre permettant d'exprimer l'un ou l'autre de ces sens : *convaincre* et *persuader*. Nous appellerons les sens suspectés TO-DO et VERACITY.

Tout d'abord, nous avons dit qu'un des éléments qui nous avait mis la puce à l'oreille sur la possibilité de VERACITY et TO-DO comme sens distincts était les restrictions de sélection et les contraintes syntaxiques que chacun d'eux impose. Le sens VERACITY n'accepte que les compléments de type propositionnel, contenu mental, tandis que TO-DO ne peut être complété que par un argument de type action. De plus, comme illustré en (69), les deux sens potentiels sont chacun associé à des constructions particulières. La lecture TO-DO n'est par exemple disponible qu'avec un complément de type infinitive, ou complétive au subjonctif (comme en (69c) qui n'est d'ailleurs acceptable que si Alex a un pouvoir ou une autorité sur la venue de Marie, et se comprend *Sam a convaincu Alex de faire que Marie vienne.*), et réciproquement, un complément de ce type rend une lecture VERACITY impossible. De même, un complément de type syntagme prépositionnel force une lecture VERACITY. Il y a donc clairement **cooccurrence différentielle**.

(69) a. Sam a convaincu.VERACITY Alex que Marie venait/viendrait.
b. #Sam a convaincu.TO-DO Alex que Marie venait/viendrait.
c. Sam a convaincu.TO-DO Alex que Marie vienne.
d. #Sam a convaincu.VERACITY Alex que Marie vienne.
e. Sam a convaincu.VERACITY Alex de la venue de Marie.
f. #Sam a convaincu.TO-DO Alex de la venue de Marie.

Pour ce qui est du test définitionnel, il est possible d'établir une **définition** couvrant les deux cas : *un élément provoque l'émergence d'une position cognitive chez un être sensible.* Dans le cas VERACITY, cette position cognitive porte sur la véridicité d'un contenu propositionnel, tandis que dans le cas TO-DO, elle porte sur la véridicité du fait qu'il est opportun d'effectuer une action ou en tout cas de faire en sorte qu'une éventualité future ait lieu (par la suite, nous parlerons de type sémantique « action » pour l'éventualité à faire arriver aussi bien que pour l'action à effectuer).

Comme c'était déjà le cas plus haut, le **test logique** n'est pas applicable : *convaincre* et *persuader* demandent dans l'usage TO-DO un argument de type action, et pour le sens VERACITY un argument de type propositionnel.

Il en va de même pour le **critère d'interprétation multiple** puisque comme nous venons de le voir, il existe une définition couvrant les deux sens.

En (70), nous avons coordonné deux contextes forçant respectivement les interprétations VERACITY et TO-DO. Ces phrases ne semblent pas très heureuses (ni vraiment zeugmatiques), mais cela pourrait simplement être dû au fait que nous avons dû pour les construire coordonner des éléments de catégories syntaxiques différentes. Il est donc difficile de dire si le **test de cooccurrence compatible** pointe vers l'unité ou vers l'existence de deux sens.

(70) a. ?Sam a convaincu Alex qu'il allait pleuvoir et de rentrer.
b. ?La vue des gros nuages a persuadé Alex qu'il allait pleuvoir et de rentrer.

Le **test d'identité** n'est applicable qu'avec un contexte qui permet l'emploi des deux sens suspectés. Changer ce contexte revient à tester la cooccurrence compatible plutôt que l'identité. Pour les mêmes raisons que nous n'avons pu appliquer le test logique, nous considérons ici le test d'identité inapplicable.

Comme nous l'avons évoqué plus haut, *persuader* et *convaincre* quelqu'un sont synonymes du verbe *apprendre* uniquement au sens VERACITY, et antonymes de *dissuader* uniquement au sens TO-DO. Il y a donc **indépendance des relations lexicales**.

De plus, comme on peut le voir en (71), les résultatifs *convaincu*, *persuadé* et la nominalisation *conviction* sont valables uniquement pour le sens VERACITY. Il y a donc **dérivation différentielle**.

(71) a. Alex est convaincue/persuadée qu'il va faire beau.
b. *Alex est convaincu/persuadé d'aller à la plage. [3]
c. Sa conviction qu'il va faire beau est forte.
d. *Sa conviction d'/pour aller à la plage est forte.

Les seuls tests présentés qui nous ont paru applicables pointent vers une distinction entre les sens « convaincre quelqu'un de la véracité d'une proposition » et « convaincre quelqu'un de la nécessité d'effectuer une action ». De plus, être capable d'effectuer cette distinction nous paraît pertinent dans le cadre de notre projet, puisqu'un seul de ces sens appartient réellement au domaine des positions cognitives [4]. Nous prenons donc la décision de séparer FR_COGNIZER_AFFECTING en deux sous-cadres : FR_COGNIZER_AFFECTING.VERACITY et FR_COGNIZER_AFFECTING.TO_DO.

5.1.2.3 Constater et remarquer (BECOMING_AWARE vs. STATEMENT)

Un certain nombre de déclencheurs se trouvent à la fois dans le lexique proposé à la validation des cadres BECOMING_AWARE, qui décrit la situation où un élément (Phenomenon) entre dans le modèle du monde d'un Cognizer, et STATEMENT, dans lequel un Speaker utilise la langue pour exprimer un Message. Ces deux cadres ont entre autres en commun les déclencheurs verbaux *constater*, *noter*, et *remarquer*. Dans de nombreux cas, en contexte, il est difficile ou même impossible de déterminer si l'on a affaire à l'un ou l'autre des cadres. Il est donc légitime de se demander si les deux sens sont réellement distincts pour ces déclencheurs. Pour tenter de répondre à cette question, nous appliquons ici encore les tests décrits supra, cette fois-ci sur les verbes *constater* et *remarquer*.

Le **test logique** est cette fois-ci applicable : dans une situation où Alex aurait vu de ses yeux qu'il faisait beau dehors mais ne l'aurait annoncé à personne, la réponse à (72) serait à la fois *oui* et *non*, ce qui pointe vers une polysémie pour les deux verbes.

(72) a. Est-ce qu'Alex a constaté qu'il faisait beau ?
b. Est-ce qu'Alex a remarqué qu'il faisait beau ?

Nous passons donc au **critère d'interprétation multiple**. A l'examen de (73), il nous semble que les deux phrases peuvent donner lieu à deux interprétations où seul l'un des sens suspectés est présent, mais une lecture ou les deux le sont paraît aussi possible. Nous utilisons donc une particularité des cadres portant les sens auxquels nous nous intéressons. BECOMING_AWARE comporte un rôle Evidence, qui correspond à l'indice ayant permis au Cognizer d'intégrer le Phenomenon dans son modèle du monde. De son côté, STATEMENT intègre un rôle Medium : l'objet physique ou le canal utilisé pour exprimer le Message. En (74), on est obligé d'interpréter *le rapport* soit comme Medium, soit comme Evidence : les deux lectures sont possibles mais en aucun cas simultanément.

3. Cet exemple n'est pas inacceptable en soi mais force l'interprétation passif processif où Alex est en train d'être convaincue/persuadée. La lecture qui nous intéresse, pour laquelle *convaincue/persuadée* est résultatif, est donc inaccessible.

4. Nous verrons néanmoins en 6.4 que nous avons par extension intégré au domaine des positions cognitives les cadres correspondant à la version .TO-DO des cadres du domaine.

(73) a. Alex a constaté qu'il pleuvait.
b. Alex a remarqué qu'il pleuvait.

(74) a. Alex a constaté dans le rapport que le chômage était en hausse.
b. Alex a remarqué dans le rapport que le chômage était en hausse.

De même, on peut voir en (75) qu'une phrase avec nos verbes où ces deux rôles sont exprimés n'est pas interprétable. Il n'y a donc pas de **cooccurrence compatible** des deux sens.

(75) a. #Alex a constaté à haute voix et dans les statistiques que le chômage était en hausse.
b. #Alex a remarqué publiquement et dans les statistiques que le chômage était en hausse.

Le **test d'identité** pointe aussi vers deux sens : (76) ne peut être comprise que si les deux occurrences du verbe renvoient au même sens.

(76) a. Alex a constaté qu'il faisait beau et Sam a constaté qu'il faisait beau aussi.
b. Sam a remarqué qu'il faisait beau et Alex a remarqué qu'il faisait beau aussi.

Les sens potentiels présentent deux **ensembles disjoints de relations lexicales** : pris dans le sens BECOMING_AWARE, les deux verbes sont synonymes de *découvrir* et *réaliser*, tandis qu'au sens STATEMENT, ils sont proches de *dire* et *murmurer*.

Les deux sens présentent des **ensembles disjoints de cooccurrents** : des expressions telles que *à haute voix* et *publiquement* ne sont pas compatibles avec le sens BECOMING_AWARE.

Les deux sens présumés du verbe *constater* ne présentent pas d'**ensembles disjoints de dérivés**. En revanche, la nominalisation *remarque* du verbe *remarquer* ne s'applique qu'à son sens STATEMENT.

Enfin, pour ce qui est du **test définitionnel** : si on considère que le sens STATEMENT implique forcément aussi le sens BECOMING_AWARE, on peut dire que les deux sont unis par la définition *prendre conscience d'un phénomène*. Mais les tests de cooccurrence compatible ont fait ressortir qu'il était possible de *constater* au sens STATEMENT sans *constater* au sens BECOMING_AWARE (ce qui est d'ailleurs confirmé par le fait que (77) est acceptable). Il nous semble donc impossible d'unir les deux sens en une seule définition sans disjonction.

(77) Le diplomate a constaté publiquement qu'une décision avait été prise mais il n'en pense pas un mot.

L'ensemble des tests pointe vers l'existence de deux sens distincts, les cadres associés à ces sens resteront donc tels quels et seront bien déclenchés par les mots examinés. Le problème de difficile désambiguisation en contexte qui nous avait fait nous pencher sur cette étude de polysémie devra donc être pallié par une grande précision de la documentation sur la marche à suivre pour l'annotation de ces déclencheurs (cf. quelques exemples annexe A). Nous verrons en 5.2.2.2 que pour les cas de désambiguïsation difficile, nous avons choisi de préciser un cadre par défaut. Dans les cas comme celui qui nous occupe, le cadre cognitif a toujours été priorisé par rapport au cadre langagier : nous avons considéré qu'un cas de divergence entre message et position cognitive comme dans l'exemple (77) serait nécessairement explicite.

5.2 Travail conjoint cadres et lexique

Comme mentionné dans le chapitre précédent, sept domaines notionnels ont été sélectionnés pour leur intérêt théorique ou applicatif. Les domaines, brièvement décrits en 4.1.2.1, sont la cause, la communication langagière, les positions cognitives, les transactions commerciales, le jugement/évaluation, les relations spatiales, et le temps. Les étapes décrites dans le présent chapitre

ont été effectuées pour tous ces domaines, mais seuls les quatre premiers ont été annotés dans le corpus. Les chiffres présentés dans ce chapitre concerneront donc tous les domaines, tandis que ceux du chapitre 7 ne porteront que sur les domaines de la cause, de la communication langagière, des positions cognitives et des transactions commerciales. Deux ou trois responsables ont été désignés pour chaque domaine, et c'est ensemble qu'ils ont effectué les différentes tâches décrites dans la suite de ce chapitre. Nous avons, en ce qui nous concerne, particulièrement travaillé sur les domaines des transactions commerciales et des positions cognitives.

5.2.1 Modélisation d'un domaine

Le travail de modélisation d'un domaine implique d'abord de fournir une définition précise du périmètre sémantique dudit domaine. ll faut ensuite sélectionner les cadres BFN appartenant au domaine, sachant qu'un même cadre peut appartenir à plusieurs domaines. Pour ce faire, nous avons utilisé deux heuristiques principales. Pour chaque domaine, nous avons cherché des cadres existants évidents pour le domaine (que nous avons repérés car ils sont évoqués par des mots anglais typiques du domaine), puis suivi les relations qui les liaient à d'autres cadres pour ajouter ces derniers aux cadres à considérer. L'autre heuristique a consisté à ajouter aux cadres à considérer les cadres BFN comportant un nom de rôle estimé typique du domaine (comme le rôle Speaker pour la communication langagière), ainsi que les cadres liés à eux par une relation. Cette technique génère plus de bruit que la première mais permet de repérer des cadres que la première heuristique n'aurait pas nécessairement mis en évidence. Le périmètre défini à la première étape permet ensuite de déterminer parmi tous les cadres repérés lesquels appartiennent réellement au domaine. Enfin, un certain travail sur les cadres restants est nécessaire afin d'éviter de possibles redondances. L'ensemble de ce travail est décrit plus avant et exemplifié dans les deux sous-sections qui suivent.

5.2.1.1 Définition et tests d'appartenance

Les domaines étaient initialement conçus autour d'un concept, qu'il nous a fallu au début du projet définir plus précisément afin de pouvoir circonscrire au mieux le domaine. Les responsables de chaque domaine ont donc ensemble décidé d'une définition sémantique pour ce domaine, qui permettrait potentiellement par la suite de trouver des tests formels à même d'en exclure les cadres qui n'y appartenaient pas.

Comme nous l'avons vu en 4.1.2.1, le domaine des transactions commerciales a par exemple été défini comme :

> *l'ensemble des cadres décrivant une situation au cours de laquelle a lieu un échange d'argent contre : biens ou services (propriété), usage de biens (location), ou travail (paiement de salaire).*

Différents cadres ont été examinés à l'aune de cette définition. Nous sommes partis des cadres évoqués par des déclencheurs nous semblant prototypiques de notre définition, et avons élargi notre recherche à tous les cadres en relation avec ces cadres, puis à ceux en relation avec ceux de ces cadres que nous avions conservés, et ainsi de suite. L'ensemble des cadres examinés pour le domaine des transactions commerciales est représenté figure 5.1.

Ont ainsi été intégrés au domaine des cadres prototypiques de la transaction commerciale, comme COMMERCE_SELL (*sell.v*, *auction.v*, ...) qui décrit une situation dans laquelle un protagoniste cède des biens à un autre protagoniste en échange d'argent, mais aussi des cadres intégrant une autre dimension que la transaction commerciale. C'est le cas pour IMPORTING (*import.v*, *importation.v*, ...), qui réfère à la situation dans laquelle un protagoniste reçoit des biens venus d'une zone d'exportation, séparée par une limite ou frontière de la zone d'importation : ce cadre inclut nécessairement un acte d'achat de la part de l'importateur, mais cet achat n'est pas tout ce qui constitue la sémantique du cadre. À l'inverse, certains cadres examinés ont été exclus du domaine

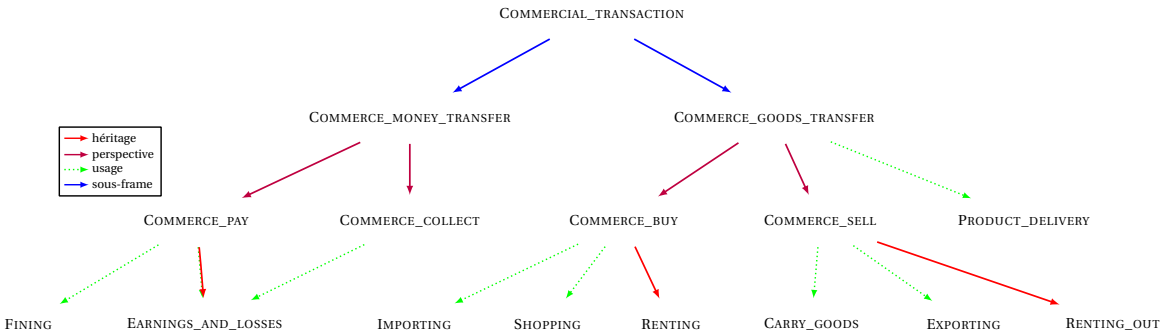


FIGURE 5.1: Ensemble des cadres examinés pour le domaine des transactions commerciales

car bien qu'y faisant référence, ils n'incluaient pas de transaction commerciale. C'est le cas du cadre SHOPPING, qui décrit la situation dans laquelle un protagoniste cherche des biens à acheter. Ce cadre présuppose une intention d'effectuer une transaction commerciale, mais ladite transaction n'y intervient pas nécessairement.

Dans certains cas, une deuxième heuristique a dû être développée pour permettre de rassembler tous les cadres BFN candidats à l'appartenance au domaine. C'est le cas du domaine des positions cognitives qui, nous l'avons vu brièvement en 4.1.2.1, est défini comme :

> *l'ensemble des cadres pour lesquels le cœur de la sémantique du prédicat doit porter sur la position cognitive d'un être doué de conscience par rapport à un contenu propositionnel. Par* position cognitive*, on entend position (gradable) sur la valeur de vérité d'un contenu, sa réalité dans le modèle du monde d'un être doué de conscience.*

Cette définition requiert que tout cadre appartenant au domaine des positions cognitives réponde aux trois critères suivants :

— avoir un rôle correspondant à l'être doué de conscience de la définition : il s'agit donc d'un rôle de type sémantique *sentient*.
— avoir un rôle correspondant au contenu propositionnel (c'est-à-dire auquel on peut assigner une valeur de vérité) sur lequel le premier participant a une position cognitive.
— la position cognitive du protagoniste sur le contenu doit être au cœur de la sémantique du cadre.

La méthode décrite et exemplifiée supra de partir de lexèmes anglais nous paraissant typiques du domaine (pour les positions cognitives : *think, believe, consider*...), repérer à quels cadres ils sont associés, et suivre les relations cadre-à-cadre nous a ramené quelques candidats. Nous avons néanmoins rapidement observé que certains de ces cadres n'étaient liés entre eux par aucune courte chaîne de relations. C'est le cas par exemple des cadres AWARENESS (*think, believe, suppose*...), qui réfère à la position cognitive d'un être doué de conscience sur un contenu mental, et CATEGORIZATION (*consider, deem, perceive*...), qui décrit une situation où un être doué de conscience considère qu'un élément appartient à une certaine catégorie. CATEGORIZATION réfère donc à la position cognitive de son protagoniste sur la prédication de la catégorie sur l'élément, et appartient à ce titre au domaine des positions cognitives, mais nous n'aurions pu l'identifier en partant simplement du cadre AWARENESS.

Nous avons alors conçu une procédure alternative. Ayant remarqué que le protagoniste de la plupart des cadres jusqu'alors repérés pour le domaine était appelé *Cognizer*, nous avons capturé tous les cadres ayant un rôle du nom de Cognizer. Nous avons ensuite déterminé lesquels d'entre eux disposaient d'un rôle de type contenu propositionnel. Les types sémantiques des différents rôles des cadres n'étant pas toujours spécifiés (c'est le cas pour beaucoup de rôles de type contenu), nous avons considéré tous les rôles dont le remplisseur pouvait être propositionnel (on peut lui assigner une valeur de vérité), comme dans le cas d'une complétive/infinitive ou d'une question,

ouverte ou fermée. Il ne nous restait ensuite plus qu'à éliminer du domaine les cadres pour lesquels la position cognitive du protagoniste sur le contenu ne nous paraissait pas centrale. Nous avons alors pu examiner les cadres en relation avec les cadres conservés à l'issue de cette procédure, et ainsi de suite.

Cette heuristique consistant à utiliser certaines régularités dans le nommage des rôles pour repérer des cadres appartenant potentiellement à nos domaines a été appliquée seulement pour les domaines paraissant peu cohésifs en matière de relations entre cadres.

5.2.1.2 Travail sur les cadres

NB Dorénavant, pour les cadres du domaine des positions cognitives, nous utiliserons le terme *cogniteur* pour l'être doué de conscience qui a ou acquiert une certaine position mentale sur un contenu propositionnel. C'est le mot que nous avons choisi pour traduire *Cognizer*, le nom du rôle référant à ce participant dans la plupart des cadres.

Nous avons vu au chapitre précédent, section 4.2.4.3, que pour éviter une annotation redondante, nous ne souhaitions globalement annoter que les rôles noyaux ou presque (les remplisseurs des rôles non-noyaux ne sont annotés que lorsqu'ils sont sous-catégorisés par le déclencheur). Cette orientation rend encore plus importante à nos yeux la distinction des rôles selon leur statut de nucléarité. La phase d'annotation du lexique des cadres décrite en 5.2.2.2, combinée aux critères de nucléarité définis section 4.2.3.1, a pu parfois mener à un remaniement des statuts nucléaires ou non de rôles desdits cadres.

De même, la phase d'annotation du lexique des cadres nous a parfois permis de remarquer que les déclencheurs proposés pour un même cadre présentaient des divergences quant à la nucléarité de certains de leurs rôles. L'existence de divergences sur ce point — ou tout autre point présenté en 4.2.4.2 comme critère de délimitation d'un cadre — nous a menés à parfois diviser un cadre en plusieurs. Ainsi, pour le cadre COMMERCE_PAY, des déclencheurs comme *payer* se mélangeaient à d'autres comme *verser*. *Payer* implique nécessairement un échange de l'argent déboursé contre un bien ou un service, qui correspond au rôle Goods. En revanche, avec *verser*, la présence de ce Goods n'est pas obligatoire. Nous avons donc créé un cadre supplémentaire : FR_GIVING_MONEY, qui sera évoqué par les déclencheurs comme *verser* pour lesquels le rôle Goods n'est pas noyau.

Par ailleurs, comme nous l'avons exploré en 4.1.2, la méthode domaine-par-domaine permet de repérer les cadres similaires les uns aux autres, et ainsi de se pencher de près sur ce qui les distingue deux à deux afin d'éviter de possibles redondances et donc une polysémie « artificielle », c'est-à-dire qui ne serait induite que par le prisme de la création de cadres, ce qui n'a pas lieu d'être. Cette manière de procéder nous a donc parfois amenés à fusionner deux cadres.

Lorsque nous avons rencontré des cadres BFN qui nous paraissaient très similaires, nous avons tenté de trouver ce qui pouvait les différencier, comparant à la fois leurs définitions, leurs ensembles de rôles noyaux et les relations entre ces rôles, les déclencheurs pouvant les évoquer, et enfin, les annotations de ces déclencheurs. Procéder ainsi nous a parfois permis de discerner une différence entre les deux cadres et de la préciser dans leurs définitions afin de faciliter la tâche des annotatrices, mais il nous est aussi arrivé de tomber sur des cadres entièrement redondants. Ainsi, les cadres SUBJECTIVE_INFLUENCE et INFLUENCE_OF_EVENT_ON_COGNIZER, à la croisée des domaines Positions cognitives et Cause. SUBJECTIVE_INFLUENCE (*drive.v*, *influence.v*, *inspire.v*, ...) réfère aux situations où un agent, une situation ou une entité a une influence sur un cogniteur, qui peut se manifester par ce que le cogniteur s'engage dans une action, modifie son comportement ou crée un produit. De son côté, INFLUENCE_OF_EVENT_ON_COGNIZER (*guide.v*, *influence.v*, ...) décrit les cas où une situation ou un protagoniste créent des conditions qui poussent un cogniteur à exécuter une action, ou l'influencent dans une activité dans laquelle il est déjà engagé. Les définitions de ces deux cadres sont très semblables, et leurs rôles compatibles, bien qu'INFLUENCE_OF_EVENT_ON_COGNIZER semble plus restreint. De même, nous n'avons pas pu discerner de différence entre leurs ensembles

de déclencheurs qui permettaient tous l'expression des rôles des deux cadres, ni leurs annotations (peu nombreuses pour INFLUENCE_OF_EVENT_ON_COGNIZER). Aucune relation n'étant de plus spécifiée pour ces deux cadres, nous avons considéré qu'il s'agissait simplement d'une incohérence, que notre méthodologie en domaines nous avait permis de repérer. Nous avons donc déterminé de fusionner les deux cadres en un seul, en amalgamant aussi ses rôles équivalents (Agent avec Entity et Protagonist, Situation avec Situation, et Behavior avec Activity).

D'autres modifications de cadres ont parfois aussi eu pour raison des différences de lexique entre le français et l'anglais. Nous avons d'ailleurs vu en 5.1, et nous redirons en 5.2.2.2 que l'étude du statut mono- ou polysémique de déclencheurs potentiels en français nous avait parfois menés à fusionner ou diviser des cadres BFN. La table 5.2 indique le nombre de cadres du French FrameNet (version 1.2) ayant subi des modifications de chacun des types décrits plus haut par rapport aux cadres du BFN. Nous n'y avons inclus que les cadres des quatre domaines sur lesquels a été menée la phase d'annotation.

L'ensemble du travail sur les cadres décrit ici nous a permis de creuser de manière approfondie ce qui différenciait chacun de nos cadres des autres cadres conservés dans le domaine. Nous avons pris le parti de profiter de ce travail pour préciser ces différences dans la partie « définition » de nos cadres. Pour chaque cadre sont listés ses cadres les plus proches, avec un lien vers leur visualisation et une description des caractéristiques qui les distinguent du cadre courant. Un exemple est présenté figure 5.2.

Nos distinctions cadre à cadre étaient au départ inscrites dans des guides spécifiques à chaque domaine, mais la méthode décrite ci-dessus s'est avérée plus pratique, entre autres parce que certains cadres chevauchent plusieurs domaines. Ce changement de support en cours de route fait que la description des caractéristiques distinctives de chaque cadre proche ne figure pas encore dans tous nos cadres.

Domaine notionnel	Nb. de cadres	Nb de nouveaux cadres	Nb. total de cadres modifiés [5]	Fusions	Divisions	Modifs. de rôles
Cause	11	2	5	2	0	3
Communication langagière	52	8	23	4	7	12
Positions cognitives	46	6	27	6	11	13
Transactions commerciales	19	5	3	0	0	3
Total	109 [6]	18	45	12	11	25

TABLE 5.2: Nombre de cadres (lexicalisés) dans la structure du French FrameNet (v. 1.2), répartis par domaine notionnel, avec comptes de nouveaux cadres et de cadres modifiés par rapport au FrameNet de Berkeley

5. Lorsque plusieurs cadres French FrameNet proviennent d'une fusion de cadres BFN suivie d'une division de ce cadre sur un autre axe (comme dans l'exemple présenté en 5.1.2.2), ces cadres sont comptés à la fois dans la colonne des fusions et dans celle des divisions de cadres. La colonne *Nb. total de cadres modifiés* ne revient donc pas toujours à la somme des trois colonnes suivantes. Notons que ce cas ne touche que trois cadres French FrameNet, tous dans le domaine des positions cognitives.

6. Comme certains cadres appartiennent à plusieurs domaines, les totaux sont inférieurs à la somme des nombres pour chaque domaine.



Coming_to_believe

Domains: Cognitive_positions,Verbal_communication

Definition:

A person (the Cognizer) comes to believe something (the Content), sometimes after a process of reasoning. This change in belief is usually initiated by a person or piece of Evidence. Occasionally words in this domain are accompanied by phrases expressing Topic, i.e. that which the mental Content is about.

Based on the most recent census I have CONCLUDED that most Americans sleep too much.

ASFALDA French FrameNet comment: The reasoning from some Evidence or through some Means is the key aspect of that frame, to distinguish it from FR_Becoming_aware. Contrary to the above definition, for the LUs of that frame, the change in belief cannot be initiated by another person (which is the case for FR_Cognizer_affecting.veracity).

Distinctions with other frames:

≠ FR_Becoming_aware: Evidence is only a core role for Coming_to_believe: FR_Becoming_aware's LUs do not imply that the Cognizer's stance on the Content was based on anything, whereas Coming_to_believe's do.

≠ FR_Cognizer_affecting.veracity: Coming_to_believe's Cognizer's stance cannot be initiated by a human volitional source, whereas FR_Cognizer_affecting.veracity's can.

≠ FR_Awareness-Certainty-Opinion: Coming_to_believe is inchoative of FR_Awareness-Certainty-Opinion, which is a stative frame.

FIGURE 5.2: Capture d'écran de la définition du cadre COMING_TO_BELIEVE et de ses caractéristiques distinctives

5.2.2 Lexique des cadres

Une fois les domaines bien définis et délimités, et leurs cadres sélectionnés, il nous a fallu doter ces cadres de lexiques de déclencheurs qui les évoqueraient. Nous sommes pour cela partis de lexiques obtenus automatiquement, que nous avons manuellement validés et enrichis. C'est ce qui est décrit dans les sections suivantes.

5.2.2.1 Lexiques automatiques

Nous avons utilisé deux lexiques automatiques issus respectivement des travaux de Padó (2007) et de Mouton *et al.* (2010).

Mouton *et al.* ont obtenu un lexique par traduction de l'anglais vers le français des déclencheurs des cadres BFN, à l'aide de dictionnaires bilingues. Plusieurs heuristiques de filtrage et d'enrichissement on été appliquées, et le lexique français obtenu comporte 15 132 déclencheurs répartis entre 727 cadres.

Padó a, comme on l'a vu en 3.4.5, projeté des annotations sur des corpus parallèles, créant ainsi des ensembles d'annotations FrameNet de l'allemand et du français. Jouant sur différents paramètres, il a pu extraire des annotations françaises obtenues différents lexiques FrameNet français. Nous avons utilisé celui présentant la meilleure précision, qui compte 6 770 déclencheurs pouvant évoquer 482 cadres. Le lexique obtenu étant un lexique de formes fléchies non catégorisées, nous avons effectué une étape de catégorisation en nous aidant du Lefff (Sagot, 2010), un lexique morphologique et syntaxique à large couverture du français. Lorsque le candidat déclencheur correspondait à plusieurs entrées du Lefff, une seule a été conservée, suivant un ordre de priorité sur les catégories que nous avons établi (verbe > nom > adjectif > tout le reste). Certains mots de la version du lexique utilisée ne correspondaient à aucune entrée du Lefff, nous leur avons donc assigné la catégorie *unknown*. Cette manière de procéder a fait que certains mots ont disparu du lexique (les formes conjuguées de verbes qui y étaient déjà présents).

Nous avons utilisé l'union du résultat de ce travail et du lexique de Mouton *et al.* comme point de départ du reste du travail de création du lexique de la ressource.

5.2.2.2 Travail sur ces lexiques

Le travail sur le lexique de chaque cadre a été effectué séparément par chacun des deux responsables de chaque domaine notionnel, à l'aide de l'outil *Asfalinks* (présenté dans (Candito *et al.*, 2014a)), développé par Gaël de Chalendar dans le cadre du projet ASFALDA.

Ce travail comprend deux aspects : le filtrage et l'enrichissement du lexique automatique décrit en 5.2.2.1. Afin d'éliminer le bruit, chacun des responsables d'un domaine a dû, en s'aidant de la définition du cadre concerné, retirer du lexique les candidats déclencheurs n'évoquant en fait pas ce cadre. D'autre part, afin de combler le silence, chacun a cherché à ajouter des déclencheurs manquants, à la fois par introspection et à l'aide d'une procédure systématique. Nous avons considéré des synonymes des déclencheurs conservés, souvent repérés à l'aide du Dictionnaire Électronique des Synonymes du CRISCO (Morel et François, 2015). Nous avons aussi considéré les verbes correspondant aux déclencheurs nominaux, et certains déverbaux des déclencheurs verbaux (les nominalisations pouvant, dans un de leur sens, référer à l'éventualité décrite par le verbe, mais pas les noms d'agent par exemple).

Pour certains domaines, les responsables ont pu s'aider de lexiques pré-existants. Pour le domaine Cause, les responsables ont utilisé le LEXCONN de Roze *et al.* (2012), un lexique de connecteurs de discours du français. Les responsables du domaine Jugement/évaluation ont utilisé un lexique subjectif du français développé dans le cadre du projet CASOAR de mesure de l'impact des structures de discours sur l'analyse d'opinion (Benamara *et al.*, 2011; Chardon *et al.*, 2013). Le lexique des cadres du domaine du temps a quant à lui bénéficié du travail sur l'annotation automatique TimeML du français de Parent *et al.* (2008) et Bittar *et al.* (2011).

Les deux responsables de chaque domaine ont ensuite travaillé ensemble à l'adjudication des déclencheurs de chacun des cadres de leur domaine. Cette tâche consistait pour eux à valider les déclencheurs pour lesquels ils avaient effectué le même choix et s'entendre sur les déclencheurs potentiels pour lesquels existaient des différences entre leurs deux versions (soit parce qu'ils n'étaient pas d'accord sur le statut à un déclencheur proposé par le lexique automatique, soit parce que tous deux n'avaient pas ajouté les mêmes déclencheurs lors de la phase d'enrichissement du lexique).

Asfalinks, l'outil d'annotation, permet à la fois de valider ou refuser les déclencheurs proposés pour chaque cadre, d'ajouter d'autres déclencheurs, et d'effectuer l'adjudication décrite ci-dessus. Une capture d'écran de son utilisation pour l'adjudication du lexique du cadre CONTAINING est présentée figure 5.3. On peut y voir que le lexème *abriter* a été validé par l'ensemble des responsables du domaine comme évoquant bien le cadre, tandis que *cacher* a été refusé par tous. L'icône près du lexème *comporter* indique que les responsables ont émis un avis divergent, et le conflit doit être résolu par l'adjudication.

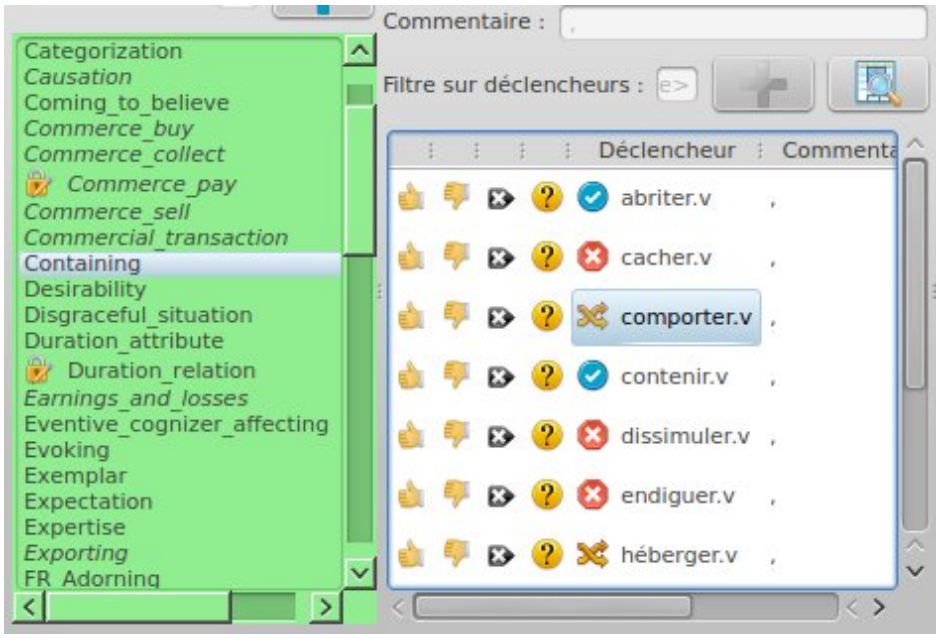


FIGURE 5.3: Capture d'écran de l'adjudication du lexique de CONTAINING à l'aide de l'outil Asfalinks

Le travail de délimitation des cadres et d'annotation de leur lexique ayant été effectué séparément par les responsables de chaque domaine, nous avons souhaité vérifier la cohérence de l'ensemble du lexique. Pour ce faire, nous avons examiné sa polysémie interne, c'est-à-dire tous les cas de déclencheurs associés à plus d'un cadre de nos domaines. Ce travail a consisté à expliciter dans quel cas un lexème ambigu évoquait quel cadre ; pour certains déclencheurs s'avérant difficiles à désambiguïser, les responsables des domaines ont écrit des guides informels de désambiguïsation à destination des annotatrices. Ces documents, dits *guides de polysémie*, intègrent généralement des définitions, des tests de désambiguïsation et parfois des règles par défaut. Quelques exemples de guides de polysémie sont présentés annexe A. Ce travail a été mené à l'aide de deux méthodes que nous avons utilisées conjointement. Pour vérifier l'existence de différents sens pour un lexème, nous avons consulté des dictionnaires, comme le TLFi, mais aussi d'autres ressources : Les verbes français (LVF) de Dubois et Dubois-Charlier (1997), les tables du lexique-grammaire de Gross (1975), ainsi qu'une version préliminaire du Réseau Lexical du Français (RLF) de Lux-Pogodalla et Polguère (2011). Parallèlement à cela, nous avons aussi observé les occurrences en corpus des lexèmes ambigus à l'aide d'un concordancier KWIC développé dans le cadre du projet. Lorsqu'un ensemble de déclencheurs apparaissait dans le lexique des deux mêmes cadres, une étude fine de polysémie comme celle menée en 5.1.2.1 nous a permis de repérer des cas de polysémie artificielle, où une granularité trop fine de nos cadres faisait que la différence sémantique qui nous paraissait claire entre lesdits cadres ne se retrouvait pas dans leurs déclencheurs. En plus de fournir des guides de polysémie à même de faciliter et rendre plus rapide l'annotation subséquente, l'examen de la polysémie nous a donc dans ces cas-là fait prendre conscience d'incohérences dans notre structure de cadres, mais aussi parfois de redondances inter-domaines. Nous avons alors pu clarifier les distinctions de cadres et modifier le lexique en conséquence.

Notons dès à présent que malgré nos précautions, l'annotation en corpus dont traite le chapitre suivant nous a souvent menés à revoir les distinctions que nous avions établies, à en spécifier de nouvelles, et parfois à modifier à nouveau le lexique de nos cadres et leur périmètre. Cela s'est en particulier produit dans les cas dont un exemple est fourni en 5.1.2.2, dans lesquels certaines occurrences de certains déclencheurs d'un cadre ne peuvent être annotées convenablement à l'aide des rôles dudit cadre. Les déclencheurs en question peuvent généralement couvrir un sens proche de celui du cadre, et ce sens n'est couvert par aucun autre cadre, ce qui fait que le problème n'est pas

nécessairement repéré dans la phase de vérification de cohérence décrite au paragraphe précédent. Une étude de polysémie permet dans ce type de cas de déterminer si un nouveau cadre doit être créé.

Un autre type de cas où les modifications ont été effectuées au cours de l'annotation est le cas que nous avons mentionné en 5.1 sous le nom de *problème de granularité trop grossière d'un cadre*. Dans ces cas, à l'examen d'un déclencheur potentiel en contexte, il est difficile d'évaluer si différentes occurrences partagent le même sens, et donc évoquent ou non le même cadre. Un exemple de tel cas est présenté en 5.1.2.3 avec l'étude de la polysémie potentielle des verbes *constater* et *remarquer*. Dans les guides de polysémie créés puis enrichis au cours de l'annotation, nous avons choisi de définir des cadres par défaut pour les cas de désambiguïsation difficile (dans les cas de compétition entre un cadre langagier et un cadre cognitif comme en 5.1.2.3, nous avons généralement choisi de privilégier le cadre cognitif). Cela correspond à un parti pris de désambiguïser même les cas limites plutôt que d'opter pour une annotation ambiguë où les deux cadres seraient annotés sur les occurrences limites.

Nous pensions originellement fournir une mesure de l'accord inter-annotateurs pour le travail de validation du lexique. Cependant, comme nous l'avons déjà mentionné, la constitution du lexique a été fortement impactée par la vérification de la nucléarité des rôles de chaque déclencheur potentiel, ainsi que le travail sur la polysémie potentielle des déclencheurs. Le lexique, de même que le périmètre des cadres, a évolué tout au long du travail, et ce jusqu'à l'annotation. En effet, l'observation systématique des déclencheurs en contexte nous a fait découvrir sans cesse de nouveaux sens potentiels qui nous ont amenés à effectuer des tests de polysémie et parfois à de nouveau changer les frontières de cadres. La mesure de l'accord inter-annotateurs n'aurait donc pas été indicative de grand chose, c'est pourquoi nous avons renoncé à en fournir une.

CHAPITRE 6

LE DOMAINE DES POSITIONS COGNITIVES

Sommaire

6.1 Lien avec la littérature **108**
6.1.1 Modes de connaissance 110
6.1.2 Factivité 112
6.2 Modélisation du domaine en cadres : vue d'ensemble **113**
6.3 Positions cognitives : le cœur du domaine **113**
6.3.1 Position cognitive sur un contenu 115
6.3.1.1 Fusion d'AWARENESS, CERTAINTY et OPINION 115
6.3.1.2 CATEGORIZATION et RELIGIOUS_BELIEF : redéfinitions de frontières de cadres 116
6.3.1.3 Deux cadres à la marge du groupe 116
6.3.2 Inchoatif de position cognitive 117
6.3.3 Position cognitive sur la probabilité d'un évènement futur 118
6.3.4 Réactivation mentale du passé 118
6.3.4.1 Fusion de MEMORY, REMEMBERING_EXPERIENCE et REMEMBERING_INFORMATION 119
6.3.4.2 EVOKING : un cadre causatif de réactivation mentale 120
6.3.5 Influence 120
6.3.5.1 Fusion de SUASION et EVENTIVE_COGNIZER_AFFECTING . . . 120
6.3.5.2 Division de REASONING 121
6.3.6 Accord 121
6.4 Positions cognitives portant sur des actions **122**
6.4.1 Influence portant sur une action à effectuer 123
6.4.1.1 Fusion de SUBJECTIVE_INFLUENCE et INFLUENCE_OF_EVENT_ON_COGNIZER 124
6.4.2 Accord portant sur une action à effectuer 124
6.4.3 Intégration du but au domaine 125
6.5 Positions cognitives de légitimité **126**
6.5.1 Cadres statifs de position sur la légitimité 126
6.5.2 Cadres causatifs de position sur la légitimité 127
6.5.3 Cadres portant sur la légitimité d'éventualités passées 128
6.6 Par extension : cadres de jugement **129**
6.7 Relations entre les cadres du domaine des positions cognitives **130**

Nous allons ici nous pencher sur le cas de la délimitation et création du lexique du domaine notionnel des positions cognitives. Nous avons vu en 4.1.2.1 que ce domaine se définissait comme l'ensemble des cadres pour lesquels le cœur de la sémantique du prédicat doit porter sur la position cognitive d'un être doué de conscience par rapport à un contenu propositionnel ; et que par *position cognitive*, on entendait position (gradable) sur la valeur de vérité d'un contenu, sa réalité dans le modèle du monde d'un être doué de conscience. Il s'agit là d'une définition initiale, valable pour les cadres au cœur du domaine des positions cognitives. Cependant, notons dès à présent que de proche en proche, nous avons été amenés à étendre le domaine pour inclure trois autres types de cas :

— les positions d'un être doué de conscience sur l'opportunité d'effectuer une action. Les cadres correspondant à ce cas sont traités en 6.4 ;
— les positions d'un être doué de conscience sur la légitimité ou le bien-fondé d'une éventualité (évènement ou état) ou d'un contenu. Les cadres correspondant à ce cas sont traités en 6.5 ;
— certains jugements de valeur portés par un être doué de conscience sur une éventualité ou un contenu. Les cadres correspondant à cette description appartiendraient plutôt au domaine Jugement/évaluation, qui n'a pas été annoté, mais nous verrons section 6.6 qu'ils sont néanmoins proches de certains cadres plus typiques de notre domaine.

Nous utiliserons généralement le mot *cogniteur* pour parler d'un être doué de conscience qui est le protagoniste des cadres du domaine, y compris ceux qui y ont été ajoutés par extension. Un cogniteur peut donc :

— arborer une certaine position quant à la valeur de vérité à assigner à une certaine proposition, que nous appellerons *contenu* ;
— arbore une certaine position quant à l'opportunité d'effectuer une action ;
— arborer une certaine position quant à la légitimité d'une éventualité ou d'un contenu ;
— porter un jugement de valeur sur une entité, une éventualité ou un contenu.

Le présent chapitre s'articule ainsi : dans un premier temps, nous évoquerons rapidement la littérature ayant trait aux prédicats inclus dans le domaine des positions cognitives. Puis, nous nous intéresserons aux problématiques pratiques de la description et du découpage de ce domaine en cadres, en décrivant les différents groupes de cadres que nous y avons inclus, et aux différences entre ces cadres.

6.1 Lien avec la littérature

Vu les critères de définition du statut des rôles établis au chapitre 4, la définition du domaine des positions cognitives que nous avons présentée ci-dessus implique des contraintes syntaxiques. Pour être des rôles noyaux, le cogniteur et le contenu doivent être sous-catégorisés par les déclencheurs des cadres du domaine, ou les gouverner dans le cas d'un déclencheur-modifieur. Nous tâchons dans la présente section de rapprocher la définition obtenue de plusieurs classes de prédicats décrites dans la littérature.

La notion qui nous semble la plus proche de nos prédicats de positions cognitives est ce que Saurí (2008) appelle *source introducing predicates* (prédicats introduisant une source), ou SIP, une sous-classe de ce qu'elle nomme *event selecting predicates* (prédicats sélectionnant une éventualité) ou ESP. Les ESP sont définis comme les prédicats qui sélectionnent un argument dénotant une éventualité (habituellement évènement ou état mais Saurí y inclut des contenus plus abstraits : proposition, fait, possibilité, issue, etc.). Parmi ces ESP, les SIP sont les prédicats qui permettent d'introduire une source par rapport à laquelle la factualité de l'éventualité sera évaluée — Saurí s'intéresse à la factualité des éventualités en discours, ou plus précisément au degré de certitude sur un contenu attribuable aux différentes sources mentionnées dans un texte. Les SIP ont donc

pour arguments une source et une éventualité, qu'on peut mettre en relation avec le cogniteur et le contenu de nos prédicats de positions cognitives. Nous ne considérons comme appartenant au domaine des positions cognitives qu'un sous-ensemble des SIP de Saurí :

— les prédicats de connaissance, qui incluent l'expression de l'état d'avoir des connaissances (comme *savoir*, *connaître*, *se rappeler*…), d'acquérir des connaissance (*découvrir*, *apprendre*, *réaliser*…), de perdre des connaissances (*oublier*), ou d'admettre des connaissances (*admettre*, *accepter*, *reconnaître*…) ;

— les prédicats de croyance et d'opinion, comme *penser*, *trouver*, *considérer*, *supposer* ou *prévoir* ;

— les prédicats de doute, comme *douter* ou *se demander* ;

— les prédicats exprimant une preuve, comme *prouver* ou *montrer* ;

— les prédicats exprimant une inférence, comme *inférer*, *déduire* ou *conclure*.

Sont exclus du domaine :

— les prédicats de perception comme *voir* et *entendre* : leur sens littéral n'a pas été retenu pour le domaine, mais nous avons considéré certains prédicats de perception comme des prédicats de la liste ci-dessus : *voir* peut par exemple fonctionner comme un prédicat de connaissance dans des phrases comme dans *Sam voit qu'il y a un problème* ;

— les prédicats de communication : pour le French FrameNet, ils appartiennent au domaine de la communication langagière. Notons au passage que l'aspect qui fait que Saurí intègre ces prédicats dans ces SIP, à savoir l'engagement du locuteur envers la factualité du contenu, n'est pas considéré dans notre traitement des prédicats de communication, qui reste très littéral ;

— les prédicats exprimant une réaction psychologique à une éventualité, comme *regretter* ou *aimer*.

Notre domaine des positions cognitives inclut aussi les prédicats causatifs de position cognitive, comme *convaincre* et *persuader*, qui n'apparaissent pas dans les classes de Saurí.

Nos prédicats de positions cognitives se recoupent aussi avec ce que l'on trouve souvent étudié sous le nom de *prédicats d'attitude propositionnelle*. Saurí (2008) les définit comme établissant « une relation entre un agent cognitif et un contenu propositionnel envers lequel l'agent cognitif possède une attitude — par exemple de croyance, connaissance, désir, regret, intention ou espoir »[traduction]. Vu la définition qui borne notre domaine des positions cognitives, nous nous intéressons uniquement à un sous-ensemble de ces prédicats : ceux qui ont trait à la croyance et à la connaissance. Cela dit, nous verrons plus bas que nous avons fini par étendre le domaine au delà de son périmètre initial, y intégrant en particulier des prédicats exprimant l'intention.

Sauri mentionne un certain nombre d'études des prédicats d'attitude propositionnelle dans plusieurs domaines et courants de recherche, de la philosophie à l'interface syntaxe-sémantique en passant par le cognitivisme. Elle indique aussi que le travail sur les prédicats d'attitude propositionnelle néglige souvent de s'intéresser au rôle que l'on a appelé le cogniteur : les prédicats sont plus souvents regroupés en fonction de ce qu'ils impliquent de l'attitude du locuteur les employant qu'en fonction de ce qu'ils impliquent de l'attitude de ce cogniteur. C'est en particulier le cas des nombreux travaux sur les modalités, c'est-à-dire l'attitude prise par les auteurs à l'égard du contenu représentatif de leurs énoncés, et sur l'évidentialité, c'est-à-dire le marquage de la source et de la fiabilité du contenu des assertions par les locuteurs ou les auteurs desdites assertions. Ces travaux n'ont pas trait au type de prédicats qui nous intéresse, ou ne les séparent pas de ceux ne permettant pas l'expression locale d'un cogniteur : les verbes sous-catégorisant un cogniteur ne sont pas distingués des autres, et sont souvent présentés uniquement à la première personne. Pourtant, on y retrouve certaines distinctions qui se recoupent avec le découpage en cadre du domaine des positions cognitives dont nous parlerons dans les sections suivantes. C'est pourquoi nous présentons ici certaines de ces distinctions qui nous semblent pertinentes pour le domaine des positions cognitives.

Les travaux que nous citons ci-après discernent plusieurs types de relations entre un cogniteur et une proposition (que ce cogniteur soit un participant sous-catégorisé par un élément de la phrase ou bien son auteur ou locuteur). Biber et Finegan (1989) considèrent deux types d'attitude pour un cogniteur : l'affect, qu'il définit comme l'expression d'attitudes personnelles comme les émotions, les sentiments, ou les humeurs (par exemple *je suis content que* ou *heureusement*) et l'évidentialité, définie comme l'expression d'attitudes envers la connaissance d'un contenu —sa fiabilité ou sa source par exemple (*impossible*, *certainement*, *je conclus que*...). Field (1997) parle pour ces mêmes notions de position affective et position épistémique. De même, Berman (2005) distingue attitude affective et attitude épistémique, et les distingue toutes deux de l'*attitude déontique*. Elle décrit cette dernière comme un point de vue évaluatif, de jugement, du locuteur envers un état de fait (exprimé par exemple par *devoir* au sens d'obligation), tandis que l'*attitude affective* exprime les émotions du locuteur quant à un état de fait. Enfin, l'*attitude épistémique* porte sur la relation entre le locuteur et le degré de probabilité d'un état de fait, ou les indices qui le poussent à croire en la réalité de cet état de fait. Ce sont ces définitions que nous gardons à l'esprit pour la suite de cette section. Nos prédicats de positions cognitives ont trait à l'attitude *épistémique* du cogniteur sur l'état de fait décrit par l'argument du prédicat que nous appelons *contenu* ou *proposition*. Nous examinerons donc uniquement les distinctions qui apparaissent dans la littérature ayant trait à l'attitude épistémique. Nous tenterons ensuite d'établir un rapport entre les sous-classes repérées et les cadres appartenant à notre domaine des positions cognitives.

6.1.1 Modes de connaissance

Chafe (1986), qui s'intéresse à l'évidentialité dans la conversation et les travaux académiques en anglais, présente plusieurs *modes de connaissance* d'un contenu asserté par un locuteur, liés à l'origine de l'information assertée. Dans les paragraphes qui suivent, nous décrivons très brièvement chacun de ces modes et mentionnons aussi leurs équivalents chez d'autres auteurs. Les exemples qui illustrent chaque mode sont repris de Chafe (1986).

Le premier mode de connaissance décrit par Chafe est la *croyance*. La source de l'information est dans ce mode insignifiante, effacée. Les exemples donnés pour l'anglais sont les expressions comme *I believe* ou *I think*.

Pour l'*induction* (ou inférence), le locuteur est parvenu au contenu qu'il exprime à partir d'indices. Chafe indique que l'induction est souvent signalée sans que la nature des indices ne soit précisée, c'est-à-dire avec l'adjectif *obvious* en (78).

(78) Well it was just obvious I couldn't work.
« Bon c'était simplement évident que je ne pouvais pas travailler. »

Chafe présente aussi un sous-type de l'induction : *les indices sensoriels*. Avec ce mode de connaissance, le locuteur indique l'indice à partir duquel il est arrivé au contenu de son énoncé. C'est le cas avec les verbes de perception dans les phrases comme (79) mais aussi avec les expressions comme *looks like* en (80).

(79) I feel something crawling up my leg.
« Je sens quelque chose me grimper sur la jambe. »

(80) She looks like she's asleep.
« On dirait qu'elle dort. »

L'ouï-dire regroupe les cas où le contenu est parvenu au locuteur par voie langagière. Ce mode de connaissance est souvent marqué par des verbes de communication, comme en (81), mais peut aussi être signifié par des marqueurs habituellement utilisés pour l'induction, comme en (82).

(81) They were using more verbs than English speaking kids have been said to learn.
« Ils utilisaient plus de verbes que ce qu'on dit que les enfants anglicistes apprennent. »

(82) Apparently at the end of a paper she gave, she pulled out a giant hairbrush.
« Apparemment, à la fin d'un examen qu'elle a fait passer, elle a sorti une brosse à cheveux géante. »

Enfin, la déduction : dans ce mode de connaissance, le locuteur marque que le contenu énoncé dérive de l'application d'un modèle de ses connaissances à une situation. Le modal *should*, illustré en (83), est un marqueur typique de déduction.

(83) He or she should take longer to respond following exposure to inconsistent information than when exposed to no information at all.
« Il ou elle devrait mettre plus longtemps à réagir après avoir été exposé-e à des informations inconsistantes qu'après n'avoir été exposé-e à aucune information. »

Les types de position épistémique décrits par Mushin (2001) reprennent et complètent les modes de connaissance de Chafe. Les positions par expérience personnelle, par inférence, et par report correspondent respectivement aux indices sensoriels, à l'induction et à l'ouï-dire de Chafe. Mushin y ajoute une *position factuelle*, pour laquelle sont absentes à la fois la source et le statut (degré de fiabilité) de l'information. Les phrases de « vérité générale », qu'on reconnaît à ce qu'il n'est pas possible d'y spécifier de source pour le contenu, correspondent à ce type de position. L'exemple (84), traduit de Mushin, illustre cette modalité. Enfin, la *position par imagination* couvre les cas où le locuteur marque l'inscription du contenu qu'il énonce dans un monde fictif de sa propre création. L'entrée dans cette position peut par exemple se faire au moyen d'expressions conventionnelles comme *il était une fois*.

(84) a. Deux et deux font quatre.
b. ??Deux et deux font quatre, d'après mon professeur de mathématiques.
c. ??J'ai l'impression que deux et deux font quatre.

Palmer (1986) distingue quant à lui trois types de modalité épistémique qui semblent se recouper avec les modes de connaissance de Chafe : le type *dubitatif* permet d'exprimer l'incertitude, le type *déductif* indique l'inférence à partir d'indices observables et le type *assomptif* exprime l'inférence à partir de ce qui est déjà connu. L'auteur indique qu'en anglais, ces trois types peuvent être marqués par les modaux *may*, *must* et *will* respectivement, et illustre cela des exemples en (85).

(85) a. John may be in his office.
« John est peut-être dans son bureau. »
b. John must be in his office.
« John doit être dans son bureau. »
c. John'll be in his office.
« John sera dans son bureau. »

Bien que pensés pour décrire la position du locuteur sur le contenu qu'il énonce, les types d'évidentialité que nous venons de présenter peuvent aussi être appliqués au cogniteur de nos prédicats de positions cognitives. Ainsi un type de source pour la position peut être inclus dans le prédicat. Il peut s'agir d'induction, comme c'est le cas avec *découvrir*, de déduction comme avec *déduire* ou encore d'ouï-dire, comme avec *apprendre une information de quelqu'un*. Cette source peut aussi être absente, effacée, comme avec *savoir* et *penser*. De même, différents prédicats permettent d'exprimer différents degrés de confiance du cogniteur en la véridicité du contenu : *supposer* indique un degré bien moins fort de croyance qu'*être sûr* ou *savoir*.

Une fois les prédicats permettant la réalisation locale d'un cogniteur isolés, leur découpage en cadres BFN concorde au moins grossièrement avec les concepts derrière les types d'évidentialité inventoriés. BECOMING_AWARE et COMING_TO_BELIEVE regroupent par exemple tous deux les prédicats évoquant les modes de connaissance inductif et déductif tandis que les cadres comme AWARENESS et OPINION regroupent ceux qui se rapprochent du mode de connaissance de la croyance.

Nous verrons en 6.4 et 6.5 que notre domaine des positions cognitives a été complété, par extension, des cadres portant sur l'opportunité d'effectuer des actions, et d'autres décrivant des jugements sur la légitimité d'états de faits. Il est intéressant de noter que ces cadres correspondent respectivement aux attitudes déontique et affective que nous avons évoquées supra.

6.1.2 Factivité

Au sein des prédicats d'attitude propositionnelle, une distinction souvent citée est la factivité, c'est-à-dire la position de l'auteur quant à la véridicité de la proposition argument du prédicat. Nous décrivons ici brièvement cette distinction, mais mentionnons dès à présent que pour le French FrameNet, elle n'interviendra pas dans la délimitation des cadres : en effet, comme nous l'avons annoncé en 4.2.4.2, le partage des présuppositions n'a pas été retenu comme critère d'appartenance à un même cadre dans notre ressource.

Les verbes factifs, classe définie par Kiparsky et Kiparsky (1970), sont les verbes dont le complément est présupposé vrai, c'est-à-dire que l'auteur exprime en les utilisant sa conviction de la véridicité du complément, et ce indépendamment de toute négation ou interrogation dans laquelle ce complément s'insèrerait. Ainsi (86d) est vrai pour l'auteur de (86a), mais aussi de (86b) et (86c).

(86) a. Alex sait que la terre est ronde.
b. Alex ne sait pas que la terre est ronde.
c. Est-ce qu'Alex sait que la terre est ronde ?
d. La terre est ronde.

Une partie des verbes de notre domaine des positions cognitives ayant ce type de comportement sont ceux que Karttunen (1971) appelle *semi-factifs*. Ces derniers perdent leur factivité dans certaines constructions, dont les constructions hypothétiques à la première personne. Ainsi (87a) présuppose (87c), contrairement à (87b), alors que (88a) et (88b) présupposent toutes deux (87c).

(87) a. Si je regrette un jour qu'Alex ne soit pas venue, mon chagrin ne durera qu'un temps.
b. Si j'apprends un jour qu'Alex n'est pas venue, mon chagrin ne durera qu'un temps.
c. Alex n'est pas venue.

(88) a. Si Sam regrette un jour qu'Alex ne soit pas venue, son chagrin ne durera qu'un temps.
b. Si Sam apprend un jour qu'Alex n'est pas venue, son chagrin ne durera qu'un temps.

Mørdrup pointe que les semi-factifs français sont tous sémantiquement proches les uns des autres, soit synonymes d'*apprendre quelque chose*, soit de *faire apprendre quelque chose*. Il s'agit donc de verbes inchoatifs ou causatifs de positions cognitives.

On pourrait imaginer que la factivité joue un rôle dans les appartenances de cadre. Cela devrait théoriquement être le cas dans le FrameNet de Berkeley, puisque les déclencheurs d'un même cadre sont censés partager les mêmes présuppositions. En pratique, cela ne semble jamais avoir été appliqué pour la factivité. Le cadre AWARENESS mélange des déclencheurs factifs et non factifs, et bien qu'une volonté de changer cette situation soit inscrite dans la définition du cadre, elle ne

semble à l'heure de l'écriture de cette thèse pas avoir été actée. Le cadre BECOMING_AWARE mélange quant à lui déclencheurs factifs (comme *realize*) et non factifs (comme *note*) et n'indique rien de particulier dans sa définition. Pour le French FrameNet, la question n'a pas à se poser puisque pour rester le plus compatible possible avec les données BFN, nous avons acté en 4.2.4.2 que le partage des présuppositions n'était pas une obligation pour tous les déclencheurs d'un même cadre French FrameNet.

6.2 Modélisation du domaine en cadres : vue d'ensemble

Dans les sections qui suivent, nous décrivons succinctement les cadres obtenus dans le domaine, proposant le cas échéant une justification des modifications opérées par rapport au BFN version 1.5. Nous l'avons mentionné en introduction : bien que le cœur du domaine des positions cognitives — et notre point de départ pour sa délimitation en cadres — soit la position d'un cogniteur sur la véridicité d'une proposition, de proche en proche, nous avons par la suite étendu le domaine à d'autres types de positions :

— les positions d'un être doué de conscience sur l'opportunité d'effectuer une action (cf. 6.4) ;
— les positions d'un être doué de conscience sur la légitimité ou le bien-fondé d'une éventualité (évènement ou état) ou d'un contenu (cf. section 6.5) ;
— certains jugements de valeur portés par un être doué de conscience sur une éventualité ou un contenu (cf. section 6.6).

Nous verrons dans les sections pointées ci-dessus les raisons qui nous ont poussés à intégrer ces cas au domaine, mais notons dès à présent qu'il semble exister un assez grand parallélisme entre les positions sur la véracité d'un contenu, sur sa légitimité, et sur l'opportunité d'effectuer une action. La table 6.1, qui rassemble tous les cadres de notre domaine classés selon leurs caractéristiques distinctives, met en évidence ce parallélisme. Elle permet aussi une visualisation globale du domaine, et pourra aider à la lecture de la suite du chapitre. Rappelons également que les cadres obtenus sont visualisables en ligne [1] et contiennent l'explicitation des caractéristiques distinctives des cadres proches entre eux.

6.3 Positions cognitives : le cœur du domaine

Nous avons vu en 5.2.1.1 que la définition du domaine requerrait que tout cadre appartenant au domaine des positions cognitives, ou du moins au cœur de ce domaine, réponde aux trois critères suivants :

— avoir un rôle correspondant au cogniteur (rôle de type sémantique *sentient*)
— avoir un rôle correspondant au contenu propositionnel (c'est-à-dire auquel on peut assigner une valeur de vérité) sur lequel le cogniteur a une position cognitive.
— la position cognitive du cogniteur sur le contenu doit être au cœur de la sémantique du cadre.

À partir de ces critères, nous avons obtenu un certain nombre de cadres, répartissables en plusieurs groupes :

— les positions cognitives statives sur un contenu ;
— les cadres rassemblant les prédicats inchoatifs de position cognitive ;
— les positions cognitives portant sur la probabilité d'un évènement futur ;
— les « réactivations mentales » du passé ;
— les cadres rassemblant les prédicats causatifs de position cognitive ;
— l'accord de plusieurs cogniteurs sur une position cognitive.

1. http://asfalda.linguist.univ-paris-diderot.fr/frameIndex.xml

TYPE	GROUPE	PARTICULARITÉ	CADRES
VÉRACITÉ	statif		FR_AWARENESS-CERTAINTY-OPINION, CATEGORIZATION, RELIGIOUS_BELIEF, FR_ATTRIBUTED_INFORMATION, FR_QUESTIONING_ONESELF, FR_ATTRIBUTING_CAUSE
	inchoatif		FR_BECOMING_AWARE, COMING_TO_BELIEVE
	causatif		FR_COGNIZER_AFFECTING.VERACITY, FR_PROVING, (FR_ATTEMPT_SUASION.VERACITY)
	accord entre deux cogniteurs	statif	BE_IN_AGREEMENT_ON_ASSESSMENT
		inchoatif	FR_MAKE_AGREEMENT_ON_ASSESSMENT
	portant sur le futur		FR_EXPECTATION, PREDICTING, FR_RELIANCE_ON_EXPECTATION
	portant sur le passé	statif	FR_MEMORY-REMEMBERING_EXPERIENCE-REMEMBERING_INFORMATION
		causatif	EVOKING
ACTION	statif		FR_PURPOSE, WILLINGNESS, FR_AGREE_OR_REFUSE_TO_ACT, RESPOND_TO_PROPOSAL, FR_AGREE_OR_REFUSE_TO_ACT_IN_FAVOR_OF, FR_MEANS_FOR_PURPOSE
	inchoatif		DECIDING
	terminatif		RENUNCIATION
	causatif		FR_COGNIZER_AFFECTING.TO_DO, FR_INFLUENCE_OF_EVENT_ON_COGNIZER-SUBJECTIVE_INFLUENCE, (FR_ATTEMPT_SUASION.TO_DO), (FR_REQUEST)
	accord entre deux cogniteurs	statif	BE_IN_AGREEMENT_ON_ACTION, FR_HAVING_COMMERCIAL_AGREEMENT
		inchoatif	MAKE_AGREEMENT_ON_ACTION
	portant sur le passé		REMEMBERING_TO_DO
LÉGITIMITÉ	statif		FR_TAKING_SIDES, FR_BEING_IN_FAVOR_OF, FR_EXPRESSING_SIDE
	causatif		FR_COGNIZER_AFFECTING.LEGITIMACY, (FR_ATTEMPT_SUASION.LEGITIMACY)
	portant sur le passé		FR_JUSTIFYING
JUGEMENT	mental		JUDGMENT
	possiblement exprimé à l'oral	polarité non-incluse	REGARD
		polarité incluse	FR_JUDGMENT_COMMUNICATION, FR_JUDGMENT_COMMUNICATION_NO_REASON

TABLE 6.1: Vue d'ensemble des cadres lexicaux du domaine des Positions cognitives

Notons que nous ne mentionnerons pas ici les cadres que nous avons examinés puis éliminés parce qu'ils ne remplissaient pas un de nos critères.

6.3.1 Position cognitive sur un contenu

Un premier groupe rassemble les cadres que nous avons identifiés comme typiques du domaine des positions cognitives. Ils évoquent tous un cogniteur ayant de manière stative, présumée invariante au moment de l'emploi du prédicat, une position quant à la valeur de vérité d'un contenu propositionnel.

6.3.1.1 Fusion d'AWARENESS, CERTAINTY et OPINION

Nous avons rapidement identifié trois cadres traitant de ce cas : AWARENESS, OPINION et CERTAINTY. D'emblée la distinction entre ces cadres a paru difficile à caractériser, (d'ailleurs, la définition d'AWARENESS dans les données 1.5 annonce un remodelage nécessaire, qui ne semble pas avoir été effectué dans les données 1.7 récemment sorties). Comme nous l'expliquons dans les paragraphes suivants, nous avons choisi de fusionner ces cadres, faute de pouvoir caractériser précisément leurs différences.

Les cadres BFN AWARENESS, OPINION et CERTAINTY regroupent en anglais des déclencheurs très similaires en sens, et plusieurs verbes se retrouvent dans le lexique de plusieurs de ces cadres. On peut citer par exemple *know*, qui fait à la fois partie du lexique d'AWARENESS et de CERTAINTY et *think*, qui fait à la fois partie du lexique d'OPINION et d'AWARENESS. Le verbe *believe*, quant à lui, est présent dans le lexique de chacun des trois cadres. Nous avons examiné ce qui pouvait distinguer ces cadres deux à deux :

CERTAINTY et OPINION La différence entre CERTAINTY [2] et OPINION [3] n'est pointée dans aucune de leurs définitions, et dans les données BFN 1.5, il n'existe pas entre eux de relation cadre à cadre, même lointaine. Nous avons observé leurs rôles et lexique respectifs, ainsi que leurs annotations pour tenter d'établir entre eux une différence, mais cela n'a pas été concluant.

AWARENESS et OPINION Si on se fie aux définitions, la différence entre les cadres AWARENESS [4] et OPINION tient au statut présupposé (et donc à la factualité) du rôle Contenu (présupposé pour AWARENESS mais pas pour OPINION). Cette distinction n'est cependant qu'annoncée comme un développement futur dans les définitions et n'a pas été actée dans la version 1.5 des données BFN, ni d'ailleurs dans les versions 1.6 et 1.7. Les critères de délimitation que nous avons retenus en 4.2.4 ne comportent pas le partage de présuppositions par tous les déclencheurs d'un même cadre. AWARENESS et OPINION doivent donc selon nos critères être fusionnés.

AWARENESS et CERTAINTY Dans le BFN 1.5, CERTAINTY et AWARENESS étaient en relation *using*, et rien dans leurs définitions ni dans leurs rôles noyaux n'indiquait ce qui pouvait les différencier. Nous avons donc décidé de fusionner AWARENESS et CERTAINTY. AWARENESS et OPINION

2. Définition de CERTAINTY dans BFN 1.5 : *This frame concerns a Cognizer's certainty about the correctness of beliefs or expectations. It only includes uses where a Cognizer is expressed.*

3. Définition de OPINION dans BFN 1.5 : *A Cognizer holds a particular Opinion, which may be portrayed as being about a particular Topic.*

4. Définition de AWARENESS dans BFN 1.5 : *A Cognizer has a piece of Content in their model of the world. The Content is not necessarily present due to immediate perception, but usually, rather, due to deduction from perceivables. In some cases, the deduction of the Content is implicitly based on confidence in sources of information (believe), in some cases based on logic (think), and in other cases the source of the deduction is deprofiled (know). Note that this frame is undergoing some degree of reconsideration. Many of the targets will be moved to the Opinion frame. That frame indicates that the Cognizer considers something as true, but the Opinion (compare to Content) is not presupposed to be true; rather it is something that is considered a potential point of difference. In the uses that will remain in the Awareness frame, however, the Content is presupposed.*

ayant déjà été fusionnés, nous avons obtenu un seul cadre au lieu de trois à l'origine : FR_AWARENESS-CERTAINTY-OPINION.

Notons que dans la version 1.7 du BFN sortie fin janvier 2017, la définition d'AWARENESS inclut une indication de ce qui distingue les deux cadres :

> *This frame is also distinct from the Certainty frame, in that it does not profile the relationship of the Cognizer to the Content, but rather presupposes it. In Certainty, the Degree of confidence or certainty is expressible as a separate frame element, as in the following : She absolutely knew that he would be there.*

Cette information ne nous était pas connue à l'époque de la création de FR_AWARENESS-CERTAINTY-OPINION, mais il nous semble *a posteriori* que notre décision aurait été la même : le profilage d'un rôle fait aussi partie des critères de délimitation de cadres BFN que nous n'avons pas retenus (cf. 4.2.4).

6.3.1.2 CATEGORIZATION et RELIGIOUS_BELIEF : redéfinitions de frontières de cadres

Le cadre CATEGORIZATION décrit le classement mental par un cogniteur (Cognizer) d'un élément (Item) dans une catégorie (Category). Ce classement peut aussi être physique (classer des livres par genre) tant qu'il implique un classement mental préalable. Nous considérons ce cadre comme relevant des positions cognitives car il s'agit pour le cogniteur d'une position cognitive sur la prédication de la catégorie sur l'élément catégorisé. Nous n'avons inclus dans le lexique de ce cadre que les déclencheurs sous-catégorisant l'élément et la catégorie (sous forme d'objet et de complément attributif de cet objet), sans alternance possible avec une complétive. Comme on peut le voir en (90), ce test nous fait associer le verbe *cataloguer* au cadre CATEGORIZATION. En revanche, *considérer*, y compris dans ses emplois avec attribut, est classé sous FR_AWARENESS-CERTAINTY-OPINION (cf. . (89)). Les dépendants de *considérer* qui auraient été annotés comme Item et Category avec CATEGORIZATION sont annotés comme remplissant le rôle Content de FR_AWARENESS-CERTAINTY-OPINION, possiblement de manière discontinue. Nous considérons que la syntaxe permettra de rendre compte de la prédication de la Category sur l'Item.

(89) a. Sam **considère** Alex (comme) sympathique.
b. Sam **considère** qu'Alex est sympathique.

FR_AWARENESS-CERTAINTY-OPINION: Cognizer, Content, Evidence, Topic

(90) a. Sam **catalogue** Alex comme sympathique.
b. *Sam catalogue qu'Alex est sympathique.

CATEGORIZATION: Category, Cognizer, Criteria, Item

Enfin, le cadre RELIGIOUS_BELIEF rassemble les croyances de type religieux. Sa définition mentionne que le contenu ou l'élément en lequel croit le cogniteur doit s'insérer dans un système constitué de plusieurs articles de foi. Nous avons considéré que pour qu'un déclencheur évoque ce cadre, son Contenu *X* devrait être une entité ou un concept faisant partie d'un système de croyance, et pouvoir être interprété comme *l'existence/le caractère sacré de X*. Cette idée d'existence ou de caractère sacré doit être implicite pour que RELIGIOUS_BELIEF soit évoqué. Ainsi, *Sam croit en dieu* évoquera RELIGIOUS_BELIEF tandis que *Sam croit en l'existence de dieu* évoquera FR_AWARENESS-CERTAINTY-OPINION.

6.3.1.3 Deux cadres à la marge du groupe

Nous avons ici examiné un groupe de cadres traitant de la position cognitive stative d'un cogniteur sur un contenu, qui comporte dans le French FrameNet trois cadres : FR_AWARENESS-CERTAINTY-OPINION, CATEGORIZATION et RELIGIOUS_BELIEF. De manière plus indirecte, nous sommes

également arrivés à inclure deux autres cadres de position cognitive stative : FR_ATTRIBUTED_INFORMATION et FR_QUESTIONING_ONESELF.

— Le cadre ATTRIBUTED_INFORMATION permet d'assigner une proposition à un locuteur qui l'aurait formulée. En français, il devrait regrouper des déclencheurs comme *d'après* et *selon*. Selon nous, ces prépositions peuvent être utilisées pour dénoter l'attribution d'une proposition sans que cette dernière ait été communiquée, entre autres parce qu'on peut clairement les utiliser pour exprimer un avis (*selon/d'après moi*). Nous avons donc modifié le cadre pour en faire FR_ATTRIBUTED_INFORMATION, qui inclut aussi les cas où une proposition est attribuée à une entité sans qu'il soit nécessairement clair que cette proposition a été communiquée.

— Un cadre FR_QUESTIONING_ONESELF a par ailleurs été créé pour les déclencheurs comme *se demander* et *s'interroger*, rencontrés lors de l'annotation des déclencheurs du cadre QUESTIONING, dans lequel un locuteur interroge un interlocuteur. FR_QUESTIONING_ONESELF décrit les situations où un cogniteur se pose à lui même une question : il s'agit là d'un processus purement cognitif. On peut par ailleurs pour certains déclencheurs avoir une nuance de doute. Ce cadre se rapproche de FR_AWARENESS-CERTAINTY-OPINION mais la position cognitive du cogniteur sur le contenu n'est pas fixée. Nous l'ajoutons néanmoins à ce groupe de cadres de position cognitive stative car c'est celui dont FR_QUESTIONING_ONESELF est le plus proche.

6.3.2 Inchoatif de position cognitive

Deux cadres BFN sont inchoatifs de position cognitive, c'est-à-dire qu'ils décrivent la situation où un Cogniteur commence à croire un Contenu, intègre ce Contenu à son modèle du monde. Il s'agit des cadres BECOMING_AWARE et COMING_TO_BELIEVE. La définition BFN 1.5 de BECOMING_AWARE signale que les deux cadres se ressemblent, et indique que les déclencheurs de COMING_TO_BELIEVE impliquent généralement un raisonnement à partir d'indices. Ce *généralement* n'indiquant pas de frontière précise entre les deux cadres, nous avons observé leurs déclencheurs et leurs rôles afin de tenter d'identifier cette frontière. Les déclencheurs de BECOMING_AWARE semblent en effet ne pas inclure de raisonnement mais certains de ceux de COMING_TO_BELIEVE, comme *learn* et *realize* peuvent ne pas inclure cette idée. Les déclencheurs de COMING_TO_BELIEVE ne semblent pas permettre de prendre pour Contenu une entité physique, alors que ceux de BECOMING_AWARE le peuvent généralement, mais de même, il y a des exceptions, comme la présence de *learn* et *find out*, qui n'autorisent pas ce type de complément, dans le lexique de BECOMING_AWARE. Quant aux rôles, ils sont comparables en types et en statuts, hormis le rôle Evidence : l'indice sur lequel se base un possible raisonnement est noyau dans COMING_TO_BELIEVE et pas dans BECOMING_AWARE.

Nous avons décidé de conserver ces deux cadres, et de préciser la différence entre eux. La croyance nouvellement acquise avec COMING_TO_BELIEVE est un accomplissement, elle a nécessairement été obtenue par un moyen (Means) et à partir d'indices (Evidence). Les rôles Means et Evidence sont tous les deux noyaux pour ce cadre, et tous ses déclencheurs (par exemple *conclure*, *déduire*, *inférer*...) impliquent l'idée de raisonnement à partir d'un indice. À l'inverse, les déclencheurs de BECOMING_AWARE (comme *découvrir*, *observer* et *réaliser*) n'impliquent aucun raisonnement ni origine pour la croyance. Dans ce cadre, le rôle Evidence n'est pas noyau. En contexte, l'idée de raisonnement à partir d'indices peut néanmoins être présente, et comme le rôle Evidence est sous-catégorisé par plusieurs déclencheurs (*s'apercevoir de l'existence du problème aux nombreuses lettres reçues*), il doit quand il est réalisé être annoté. Notons que BECOMING_AWARE a dû être modifié en FR_BECOMING_AWARE : comme aucun des déclencheurs du lexique de ce cadre ne sous-catégorise le rôle Means (moyen d'acquisition de la position cognitive par le cogniteur), ce rôle a dû être supprimé du cadre.

Le French FrameNet comporte donc deux cadres ayant trait aux positions cognitives à l'aspect inchoatif : COMING_TO_BELIEVE et FR_BECOMING_AWARE.

6.3.3 Position cognitive sur la probabilité d'un évènement futur

Un groupe de cadres du domaine partage la spécificité d'avoir un Contenu qui correspond nécessairement à une future éventualité. EXPECTATION diffère de FR_AWARENESS-CERTAINTY-OPINION en ce que les déclencheurs d'EXPECTATION incluent nécessairement l'idée que le Contenu est une éventualité future, tandis que même si en contexte, le Contenu de FR_AWARENESS-CERTAINTY-OPINION peut correspondre à une éventualité future, cela n'est pas inclus dans le sens des déclencheurs hors contexte. Ainsi l'exemple en (91a) sera annoté avec le cadre EXPECTATION tandis que (91b) le sera avec FR_AWARENESS-CERTAINTY-OPINION.

(91) a. Sam prévoit que les robots finiront par gouverner le monde.
b. Sam pense que les robots finiront par gouverner le monde.

Nous avons modifié le cadre EXPECTATION pour y ajouter le rôle noyau Time_of_eventuality, qui permet d'annoter les expressions temporelles portant sur le moment auquel le Cogniteur s'attend à voir arriver l'éventualité décrite par le Contenu (en gras dans l'exemple (92)). En effet, il s'agit d'après nos tests d'un rôle sous-catégorisé et sémantiquement obligatoire. Le cadre doté de son nouveau rôle se nomme donc FR_EXPECTATION.

(92) [...] l'équilibre d'exploitation pourrait être ENVISAGÉ **fin 1992-début 1993** [...]

Le Contenu du cadre PREDICTING doit lui aussi correspondre par défaut à une éventualité future, mais ce cadre diffère d'EXPECTATION en ce que ses déclencheurs doivent de plus pouvoir impliquer un acte de communication. À la différence de *prédire*, *anticiper* ne peut inclure d'acte de communication. C'est pourquoi (93a) est annoté comme PREDICTING alors que (93b) déclenche le cadre EXPECTATION. Le cadre possédant déjà un rôle Time_of_eventuality, il n'a pas été nécessaire de le modifier pour le français. Ce cadre appartient aussi au domaine des communications langagières.

(93) a. [...] personne n'avait **PRÉDIT** la violence de la crise monétaire de septembre [...]
PREDICTING: Eventuality, Speaker, Time_of_eventuality
b. Les torréfacteurs **ANTICIPENT** une poursuite du redressement des cours [...]
FR_EXPECTATION: Cognizer, Expectation, Time_of_eventuality, Topic

Enfin, le cadre RELIANCE_ON_EXPECTATION a lui aussi un Contenu correspondant à une éventualité dont un Cogniteur considère qu'elle va se réaliser dans le futur, mais contrairement à ce qu'on a pour EXPECTATION, ses déclencheurs impliquent que le Cogniteur considère ce Contenu assez probable pour pouvoir baser de potentiels projets dessus. Il s'agit de déclencheurs comme *escompter*, *compter sur*, ou encore *tabler*. Ce cadre a aussi été modifié par nos soins pour lui ajouter un rôle Time_of_eventuality pour les phrases comme (94). Le cadre a donc été renommé FR_RELIANCE_ON_EXPECTATION.

(94) [...] elle **TABLE** pour 1992 sur un bénéfice de 2,9 milliards de francs [...]
RELIANCE_ON_EXPECTATION: Cognizer, Expectation, Time_of_eventuality

Le French FrameNet comporte donc trois cadres ayant trait aux positions cognitives pour lesquels le Contenu est interprété comme nécessairement futur : FR_EXPECTATION, PREDICTING et FR_RELIANCE_ON_EXPECTATION.

6.3.4 Réactivation mentale du passé

Le French FrameNet intègre deux cadres ayant trait à la réactivation mentale d'un contenu déjà présent dans le modèle du monde du Cogniteur : FR_MEMORY-REMEMBERING_EXPERIENCE-RE-

MEMBERING_INFORMATION et EVOKING. Le premier de ces cadres est issu de la fusion de trois cadres BFN 1.5 : MEMORY, REMEMBERING_INFORMATION et REMEMBERING_EXPERIENCE.

6.3.4.1 Fusion de MEMORY, REMEMBERING_EXPERIENCE et REMEMBERING_INFORMATION

Les définitions des cadres MEMORY et REMEMBERING_INFORMATION (dont aucune ne mentionne l'autre cadre) n'indiquent rien qui les distingue. Le rôle Topic n'est noyau que pour MEMORY, bien qu'il existe aussi dans REMEMBERING_INFORMATION, ce qui paraît étrange puisque les déclencheurs annotés sont les mêmes. La seule différence au niveau des annotations est que MEMORY permet les exemples du type *remember Nentité* alors que ce n'est pas le cas pour REMEMBERING_INFORMATION. Ne trouvant pas de distinction correspondant aux critères que nous avons établis en 4.2.4, nous avons décidé de fusionner ces deux cadres.

REMEMBERING_INFORMATION partage plusieurs déclencheurs avec le cadre REMEMBERING_EXPERIENCE, et les définitions précisent que REMEMBERING_INFORMATION concerne l'axe *se souvenir d'un fait ou d'une information*, tandis que REMEMBERING_EXPERIENCE a trait au fait de *se souvenir d'une expérience ancrée dans le temps*. REMEMBERING_EXPERIENCE prévoit un rôle pour cette expérience ancrée dans le temps (Experience), mais aussi d'autres rôles noyaux, en exclusion mutuelle avec Experience. Ainsi, le rôle Salient_entity réfère à une entité saillante que se rappelle le cogniteur. Ce rôle peut être complété par un rôle State, évoquant l'état dans lequel était l'entité saillante que le cogniteur se rappelle, ou un rôle Impression, qui dénote une propriété que le cogniteur attribue à l'entité saillante dans son souvenir. State et Impression s'excluent mutuellement.

Dans les faits, la distinction entre les deux cadres n'est pas si claire, et on trouve dans chacun des cadres quelques annotations qui nous semble-t-il auraient aussi bien pu être dans l'autre. C'est le cas par exemple des phrases en (95), annotées avec REMEMBERING_INFORMATION, et de celles en (96), qui évoquent REMEMBERING_EXPERIENCE. Les annotations de MEMORY, elles, mélangent clairement les deux types de souvenirs, comme on peut le voir en (97).

(95) a. Is it possible to FORGET a memory?
« Est-il possible d'OUBLIER un souvenir? »

b. So then I would FORGET about it.
« Et ensuite je l'OUBLIERAIS. »

(96) a. I FORGOT about that.
« J'ai OUBLIÉ cela ».

b. She had FORGOTTEN just how strong he was.
« Elle avait OUBLIÉ à quel point il était fort. »

(97) a. I'm a Time Lord, REMEMBER?
« Je suis un Seigneur du Temps, tu T'en SOUVIENS? »

b. She REMEMBERED her experience in the supermarket.
« Elle SE REMÉMORA son expérience au supermarché. »

c. I just REMEMBER feeling at home.
« Je ME RAPPELLE seulement m'être sentie comme chez moi. »

d. I can REMEMBER the first time I got drunk on Scotch.
« Je ME RAPPELLE la première fois que je me suis saoulée au Scotch whisky. »

e. People should REMEMBER alcohol is dangerous whether you drink it in the pub or at the Christmas carol concert.
« Les gens devraient SE RAPPELER que l'alcool est aussi dangereux que l'on boive dans un bar ou au concert de chants de Noël. »

De plus, en français, nous n'avons pu trouver aucun déclencheur potentiel qui permettrait d'évoquer l'un des deux cadres seulement. Certaines constructions semblent globalement favoriser

l'une ou l'autre des interprétations : les infinitives au passé semblent évoquer le souvenir d'une expérience (*je me rappelle être montée sur la table et avoir fait un discours*), tandis que les complétives sont interprétées comme un souvenir factuel (*je me rappelle que je suis montée sur la table et que j'ai fait un discours*). En pratique, la distinction n'est pas toujours aussi claire. Est-il évident que *se rappeler être allé chercher du pain* évoque le souvenir physique de l'expérience d'aller chercher du pain et pas simplement l'information que oui, le pain a bien été acheté ? Lorsque le complément est une entité, dans certains cas, il semble clair que *se rappeler le costume sombre de Jean-Pierre Foucault* doit être interprété comme un souvenir visuel et pas comme *le fait que JP Foucault portait un costume sombre*, mais dans d'autres la distinction est plus difficile. Comment savoir si *se rappeler Sam* signifie *se rappeler qui est Sam* comme le propose la définition de REMEMBERING_INFORMATION ou bien *se rappeler le visage de Sam* ? Nous reconnaissons l'existence de la nuance entre se souvenir d'une information et se souvenir d'une expérience, mais il nous semble que cette distinction n'est pas toujours faisable. Nous avons donc décidé de fusionner le cadre REMEMBERING_EXPERIENCE avec les cadres REMEMBERING_INFORMATION et MEMORY, créant ainsi le cadre FR_MEMORY-REMEMBERING_EXPERIENCE-REMEMBERING_INFORMATION. Ce dernier a pour rôles, en plus du Cogniteur, un sujet pour le souvenir (Topic), un contenu mental (Mental_content) et une entité saillante (Salient_entity), tous trois en ensemble nucléaire.

6.3.4.2 EVOKING : un cadre causatif de réactivation mentale

De son côté, le cadre EVOKING décrit une situation dans laquelle un stimulus entraine le cogniteur à penser à un élément (situation, entité...) du passé. Ce cadre est en français évoqué par des déclencheurs comme *évoquer*, *rappeler* ou *remémorer*. Dans ce cadre causatif de réactivation mentale, le stimulus ne peut être humain et volitionnel. Ainsi, (98a) est annoté avec EVOKING, tandis que (98b) l'est avec le cadre FR_TELLING et (98c) le cadre FR_STATEMENT-MANNER-NOISE.

(98) a. [...] Ma sorcière bien aimée (12h30) **RAPPELLE** quelques souvenirs à leurs parents [...]

EVOKING: Cognizer, Phenomenon, Stimulus

b. cet échange de vues inédit a permis à M. Balladur de **REMÉMORER** aux employeurs qu'eux aussi sont comptables de l'évolution du nombre de chômeurs.

FR_TELLING: Addressee, Medium, Message, Speaker, Topic

c. M. Mitterrand **RAPPELLE** [...] les bienfaits d'une CEE consolidée par les accords de Maastricht [...]

FR_STATEMENT-MANNER-NOISE: Addressee, Medium, Message, Speaker, Topic

6.3.5 Influence

On trouve aussi dans BFN un groupe de cadres à cheval entre le domaine de la cause et celui des positions cognitives. Il s'agit de cadres causatifs de changement de position cognitive.

6.3.5.1 Fusion de SUASION et EVENTIVE_COGNIZER_AFFECTING

EVENTIVE_COGNIZER_AFFECTING et SUASION traitent tous deux d'un changement de position cognitive, position qui peut avoir trait à l'opportunité selon le cogniteur d'effectuer une action. La définition de SUASION [5] est centrée sur l'idée d'une action que le locuteur tente de convaincre

5. « Définition de SUASION dans BFN 1.5 : A Speaker expresses through language his wish to get the Addressee to act and as a result the Addressee forms an intention do so. [...] In the case of convince and persuade, the act that the Speaker wishes the Addressee to perform can be the adoption of a belief, which may be expressed as a that-clause or as an of-PP

le cogniteur d'effectuer, mais précise que cette action peut être l'adoption d'une croyance. Une partie au moins des annotations des déclencheurs de ces cadres (les cas dans lesquels l'influence porte sur un contenu à accepter comme vrai) aurait donc sa place dans le domaine des positions cognitives. Les deux cadres paraissent semblables en tout point hormis le fait que ce qui provoque le changement de position cognitive est une situation ou un évènement pour EVENTIVE_COGNIZER_AFFECTING et un être doué de conscience qui agit délibérément (généralement au moyen du langage) pour SUASION. En français, nous n'avons recensé aucun lemme permettant l'une de ces interprétations mais pas l'autre. Nous avons, comme détaillé en 5.1.2.1, appliqué des tests de polysémie aux mots qui nous semblaient pouvoir évoquer ces cadres, et en avons déduit que nous n'avions pas affaire à des sens assez distincts pour conserver une séparation en cadres. Nous avons donc fusionné EVENTIVE_COGNIZER_AFFECTING et SUASION en un cadre FR_COGNIZER_AFFECTING, qui rassemble en exclusion mutuelle un rôle *Evidence* et un rôle *Persuader*, correspondant respectivement aux situations et aux agents causatifs d'un changement de position cognitive chez un cogniteur.

6.3.5.2 Division de REASONING

Dans le cadre BFN REASONING, un être doué de conscience présente un contenu et un support à un interlocuteur afin de tenter de convaincre cet interlocuteur de croire à la réalité du contenu, ou d'effectuer une action. Pour nous, les déclencheurs de ce cadre pouvaient être séparés en deux groupes : ceux qui impliquent que l'interlocuteur est nécessairement convaincu, comme *prove* ou *demonstrate*, et les autres, comme les verbes *argue* et *reason*. Nous avons donc créé pour le French FrameNet une version restreinte de ce cadre, FR_PROVING, qui correspond à REASONING auquel on aurait ajouté la contrainte que l'interlocuteur devait à la fin de l'évènement être convaincu par le locuteur. Ce cadre diffère de FR_COGNIZER_AFFECTING en ce que tous les déclencheurs de FR_PROVING doivent impliquer que l'opération de causation de changement de position se fait de manière agentive, par le langage. Les versions françaises des déclencheurs originels de REASONING pour lesquels l'interlocuteur n'est pas nécessairement convaincu évoqueraient pour leur part le cadre ATTEMPT_SUASION, qui se trouve donc à la marge du domaine puisque rien n'oblige à ce qu'un changement de position cognitive ait lieu.

Nous avons donc dans le French FrameNet plusieurs cadres causatifs de changement de position cognitive, ou proches de ces cadres : FR_PROVING, FR_COGNIZER_AFFECTING et ATTEMPT_SUASION. Nous verrons dans les sections suivantes que nous avons en fait divisé ces deux derniers en plusieurs cadres.

6.3.6 Accord

Enfin, le BFN comporte un cadre BE_IN_AGREEMENT_ON_ASSESSMENT, qui décrit l'accord (ou désaccord) de deux cogniteurs sur un contenu, c'est-à-dire la similarité de leurs positions cognitives quant à ce contenu. Le lexique de ce cadre est constitué de déclencheurs censés permettre une réalisation collective, comme en (99), ou séparée des cogniteurs, comme en (100). Comme on peut le voir dans ces exemples, des rôles différents sont utilisés pour annoter ces cogniteurs.

(99) a. Sam et Alex sont D'ACCORD sur l'intérêt de cette idée.
b. Sam et Alex S'OPPOSENT sur la question de l'avenir de l'hôpital.

(100) a. Sam est D'ACCORD avec Alex sur l'intérêt de cette idée.
b. Sam S'OPPOSE à Alex sur la question de l'avenir de l'hôpital.

BE_IN_AGREEMENT_ON_ASSESSMENT: Cognizer1, Cognizer2, Cognizers, Opinion, Topic

complement.[...]

Les déclencheurs de ce cadre peuvent permettre de réaliser localement le contenu exact sur lequel sont d'accord les cogniteurs (*Sam et Alex sont d'accord qu'il va pleuvoir*), la question implicite à laquelle l'opinion des cogniteurs répond, ou le sujet sur lequel porte l'opinion partagée. Ces trois rôles, respectivement Opinion, Question et Topic, sont en exclusion mutuelle. La différence entre les rôles Question et Topic ne nous paraissant pas très claire ni dans la définition, ni dans les annotations BFN, nous avons spécifié pour le French FrameNet que nous n'utiliserions pas le rôle Question, qui a donc été retiré du cadre.

Nous verrons en 6.4 que nous avons aussi inclus dans le domaine un cadre proche de BE_IN_AGREEMENT_ON_ASSESSMENT mais dans lequel entre les cogniteurs porte sur une action à effectuer, ainsi que des versions causatives de ces deux cadres.

6.4 Positions cognitives portant sur des actions

En observant les cadres que nous venons de présenter, nous avons rapidement remarqué une certaine régularité : plusieurs de ces cadres semblaient proches d'un ou plusieurs autres cadres, avec pour principale différence d'avoir trait à une certaine résolution du cogniteur d'effectuer une action donnée, plutôt qu'à une position cognitive sur la véracité d'un contenu comme c'est le cas pour les cadres vus supra. On peut néanmoins voir cette idée de résolution à effectuer une action comme une position cognitive portant sur l'opportunité d'effectuer cette action. Dans certains cas, comme par exemple REMEMBERING_TO_DO et BE_IN_AGREEMENT_ON_ACTION, ces cadres portant sur une action existaient déjà, pouvant parfois même être évoqués par des déclencheurs présents aussi dans le lexique de leur version Position cognitive. Une analyse détaillée des tests de polysémie sur les verbes *convaincre* et *persuader* sur cet axe action vs. véracité d'un contenu est proposée en 5.1.2.2, et ses résultats s'appliquent globalement à tous les prédicats pouvant déclencher deux interprétations suivant ces axes. On observe dans ces cas-là généralement des cooccurrences différentielles au niveau des constructions syntaxiques : une infinitive donne nécessairement lieu à une interprétation *action*, comme en (101), alors qu'une complétive à l'indicatif évoque la version *véracité d'une position cognitive*, comme en (102). On retrouve aussi généralement une indépendance des relations lexicales (les ensembles de déclencheurs des deux cadres ne se recoupent pas entièrement). En revanche, on peut formuler une définition commune aux deux cadres, puisque comme nous l'avons dit, on peut considérer que les cadres d'action traitent d'une position cognitive sur l'opportunité d'effectuer une action. Quant aux autres tests de polysémie, ils ne sont pas toujours faciles à appliquer ou à interpréter étant donnés que les compléments favorisant l'une ou l'autre des interprétations sont syntaxiquement différents et souvent non coordonnables ou incompatibles.

(101) a. Alex a convaincu Sam d'aller chercher du pain.
b. Sam s'est rappelé d'aller chercher du pain.
c. Sam est d'accord avec Alex pour aller chercher du pain.

(102) a. Alex a convaincu Sam que la terre était ronde.
b. Sam s'est rappelé que la terre était ronde.
c. Sam est d'accord avec Alex que la terre est ronde.

Un élément est venu faciliter notre décision : nous considérons utile de pouvoir distinguer les « véritables » positions cognitives, portant sur des contenus propositionnels jugés vrais ou faux, des actions qu'un cogniteur jugerait ou non nécessaire d'accomplir. Ces dernières seraient en effet marginales au domaine des positions cognitives, auquel elles n'appartiendraient que par extension. Nous avons donc décidé de conserver la distinction entre les cadres portant sur une action et ceux portant sur un contenu propositionnel lorsqu'elle existait, et de diviser les cadres existants sur cet axe lorsque cela s'avèrerait nécessaire, c'est-à-dire lorsque les tests de polysémie pointeraient dans cette direction. Dans ce dernier cas, les nouveaux cadres émergeant de cette division ont été suffixés

en *.veracity* pour la version traitant de la position cognitive d'un cogniteur sur un contenu, et en *.to_do* pour la version ayant trait à une position sur une action. Nous récapitulons ici les cadres *.to_do* venus compléter le domaine des positions cognitives.

Nous avons plus haut mentionné le cadre REMEMBERING_TO_DO, version *.to_do* de FR_MEMORY-REMEMBERING_EXPERIENCE-REMEMBERING_INFORMATION. Il rassemble les prédicats, comme *songer* et *se rappeler*, qui peuvent décrire un cogniteur se remémorant une action à effectuer. Une spécificité de ce cadre est qu'il implique aussi que le cogniteur effectue ladite action, comme on peut le voir avec *Alex a songé à aller chercher le pain.*

6.4.1 Influence portant sur une action à effectuer

Nous avons détaillé en 5.1.2.2 l'application des tests de polysémie aux verbes *convaincre* et *persuader*, dont nous avons vu qu'ils donnaient lieu à la division de FR_COGNIZER_AFFECTING en deux cadres : FR_COGNIZER_AFFECTING.VERACITY, qui couvre les cas où quelqu'un ou quelque chose cause chez un cogniteur l'adoption d'une position cognitive sur un contenu propositionnel, et FR_COGNIZER_AFFECTING.TO_DO, pour les cas qui ont trait à l'influence de quelqu'un ou quelque chose sur le jugement porté par un cogniteur sur l'opportunité d'effectuer une action. Nous avons vu que ces deux cadres partageaient des déclencheurs, comme *convaincre* et *persuader*, mais ils présentent aussi chacun des déclencheurs qui lui sont propres. *Apprendre*, tel qu'utilisé en (103a) ne peut avoir trait qu'à la véracité d'un contenu, tandis que *dissuader*, exemplifié en (103b), ne peut porter que sur une action qu'un cogniteur a été découragé d'effectuer. Notons ici que FR_COGNIZER_AFFECTING.VERACITY inclut un déclencheur qui donne lieu à des annotations qui rappellent fortement BECOMING_AWARE. En effet, *apprendre* permet toutes les constructions visibles en (104). Malgré des exemples comme (104c) où la cause de la nouvelle position cognitive passe complètement à l'arrière-plan, la position cognitive dénotée par *apprendre* a forcément une origine, puisque (104d) n'est pas acceptable. Cette origine peut être une source ayant communiqué le contenu au cogniteur par le biais du langage, tandis que cela n'est pas possible pour les déclencheurs de BECOMING_AWARE. Ainsi nous considérons qu'*apprendre* a sa place en FR_COGNIZER_AFFECTING.VERACITY, que nous utilisons aussi bien pour annoter (104c) que (104a) ou (104b).

(103) a. Alex a appris à Sam que la terre était ronde.
b. Alex a dissuadé Sam de venir.

(104) a. Alex a appris à Sam que la terre était ronde.
b. Sam a appris d'Alex que la terre était ronde.
c. Sam a appris que la terre était ronde.
d. #Un beau jour, sans rien faire, lire ou entendre de spécial, Sam a appris que la terre était ronde.

Sur le même modèle, ATTEMPT_SUASION se décline en deux cadres : FR_ATTEMPT_SUASION.VERACITY et FR_ATTEMPT_SUASION.TO_DO, qui ont chacun un ensemble bien distinct de déclencheurs : *arguer*, *étayer*, *accréditer*... pour FR_ATTEMPT_SUASION.VERACITY et *conseiller*, *décourager*, *exhorter*... pour FR_ATTEMPT_SUASION.TO_DO. Ces cadres diffèrent des FR_COGNIZER_AFFECTING en ce que leurs déclencheurs n'impliquent pas de changement de position du cogniteur, seulement une tentative à cet effet de la part du persuadeur (par ailleurs nécessairement agentif). Mentionnons ici le cadre REQUEST, dans lequel un locuteur demande à un interlocuteur de faire en sorte qu'une éventualité advienne dans le futur. Ce cadre est très proche de FR_ATTEMPT_SUASION.TO_DO, mais le locuteur formule ici un ordre ou une supplication plutôt que de tenter de persuader l'interlocuteur, et il nous a semblé que la nuance était assez claire pour conserver les deux cadres distincts. Pour le French FrameNet, les rôles de REQUEST ont été remaniés pour se rapprocher de ceux de FR_ATTEMPT_SUASION.TO_DO. Le cadre French FrameNet se nomme FR_REQUEST, et peut être évoqué par des déclencheurs comme *demander*, *réclamer* et *implorer*.

6.4.1.1 Fusion de SUBJECTIVE_INFLUENCE et INFLUENCE_OF_EVENT_ON_COGNIZER

Enfin, puisque des cadres *.to_do* sont intégrés au domaine, il nous a paru cohérent d'examiner aussi le cadre SUBJECTIVE_INFLUENCE. Dans ce dernier, un agent, une situation ou une entité exerce une influence sur un cogniteur, l'encourageant ainsi potentiellement à effectuer une action. Pour French FrameNet, ce cadre a été restreint aux cas où cette influence n'est pas volitionnelle, entraînant ainsi la suppression du rôle Agent. C'est ce qui différencie ce cadre de FR_ATTEMPT_SUASION.TO_DO. Les déclencheurs de SUBJECTIVE_INFLUENCE pouvant aussi déclencher FR_ATTEMPT_SUASION.TO_DO présentent une particularité : employés avec un sujet agentif, ils n'impliquent pas que l'influence ait réellement eu effet, simplement que le sujet agentif a effectué des actions dans le but d'exercer cette influence. En revanche, avec un sujet évènementiel, ou de type entité non agentive, ils impliquent que le cogniteur a subi l'influence de ce sujet, influence portant sur leur activité ou les poussant vers une action. Rien n'oblige, en revanche, à ce que l'influence ait effectivement causé leur réalisation de l'action. On peut observer ce contraste en (105).

(105) a. Sam a incité/encouragé/poussé Alex à rentrer.
b. La tempête a incité/encouragé/poussé Alex à rentrer.

En examinant les données BFN, nous avons rencontré un cadre extrêmement semblable à SUBJECTIVE_INFLUENCE : INFLUENCE_OF_EVENT_ON_COGNIZER. Les deux cadres ne sont liés par aucune relation, et en observant les définitions, les seules différences semblent être que SUBJECTIVE_INFLUENCE permet une influence générale plutôt que visant une action ou activité, et que INFLUENCE_OF_EVENT_ON_COGNIZER implique que le cogniteur fasse effectivement quelque chose. Nous avons décidé de fusionner les deux cadres, créant ainsi FR_INFLUENCE_OF_EVENT_ON_COGNIZER-SUBJECTIVE_INFLUENCE, lequel ne comporte aucune de ces deux particularités. Pour nous, l'implication de la réalisation d'une action est ce qui différencie FR_INFLUENCE_OF_EVENT_ON_-COGNIZER-SUBJECTIVE_INFLUENCE de FR_COGNIZER_AFFECTING.TO_DO, comme on peut le voir en (106) ; et un déclencheur ne permettant pas l'expression d'une action ou activité n'évoque selon nous pas FR_INFLUENCE_OF_EVENT_ON_COGNIZER-SUBJECTIVE_INFLUENCE mais EXPERIENCER_OBJ qui n'est dans aucun de nos domaines, et n'est à ce titre pas annoté.

(106) a. # L'orage a DISSUADÉ Alex de sortir mais elle l'a fait quand même.
FR_COGNIZER_AFFECTING.TO_DO: Action, Cognizer, Medium, Persuader, State_of_affairs
b. L'orage a DÉCOURAGÉ Alex de sortir mais elle l'a fait quand même.
FR_INFLUENCE_OF_EVENT_ON_COGNIZER-SUBJECTIVE_INFLUENCE: Action, Activity, Cognizer, Entity, Situation

6.4.2 Accord portant sur une action à effectuer

Pour finir, BE_IN_AGREEMENT_ON_ACTION décrit une situation dans laquelle deux parties partagent un engagement envers une obligation. La définition de ce cadre précise explicitement que l'obligation contractée par les parties doit être symétrique ou réciproque, mais nous avons choisi d'utiliser une interprétation plus large : pour nous, l'obligation ne concerne pas toujours les deux parties comme c'est le cas en (107a). Il peut arriver que l'accord porte sur une action qu'une seule des parties effectuera : c'est le cas en (107b), où chaque partie doit être d'accord pour que n'importe laquelle d'entre elles rompe la confidentialité. Comme pour BE_IN_AGREEMENT_ON_ASSESSMENT, les parties doivent pouvoir être exprimées séparément aussi bien que collectivement pour qu'un déclencheur appartienne au lexique de ce cadre.

(107) a. [...] le Koweït et l'Espagne **D'ACCORD** pour trouver une solution.

b. [...] la version complète ne peut être divulguée, explique-t-on au siège de ZDS, chaque partie devant être **D'ACCORD** pour rompre la confidentialité [...]

BE_IN_AGREEMENT_ON_ACTION: Obligation, Parties, Party_1, Party_2, Topic

Mentionnons ici que BE_IN_AGREEMENT_ON_ACTION est parent par une relation d'héritage d'un cadre HAVING_COMMERCIAL_AGREEMENT, de même sens que BE_IN_AGREEMENT_ON_ACTION à l'exception près que l'accord doit être de nature commerciale — le cadre appartient donc à ce titre aussi au domaine des transactions commerciales. Ce cadre, non lexical dans le BFN, a été modifié pour le French FrameNet, devenant ainsi FR_HAVING_COMMERCIAL_AGREEMENT, cadre lexical auquel nous avons ajouté des rôles spécifiques pour vendeur et acheteur, ainsi que pour la somme d'argent négociée dans le contrat.

MAKE_AGREEMENT_ON_ACTION est un cadre BFN inchoatif de BE_IN_AGREEMENT_ON_ACTION, bien qu'ils soient tous deux signalés comme étant en relation d'usage dans BFN. Comme pour BE_IN_AGREEMENT_ON_ACTION, la définition de ce cadre prévoit une obligation symétrique ou réciproque, et comme pour BE_IN_AGREEMENT_ON_ACTION, nous avons choisi de considérer aussi les obligations asymétriques.

L'existence de ce cadre nous a amenés à réfléchir à la nécessité d'un cadre inchoatif de BE_IN_AGREEMENT_ON_ASSESSMENT. La plupart des déclencheurs de MAKE_AGREEMENT_ON_ACTION permettent aussi une version *.veracity*, et on a affaire au même type de polysémie que ce que nous avons vu supra : chacune des interprétations est fortement associée à une construction syntaxique particulière, et le lexique des deux cadres ne se recoupe pas exactement. Nous avons donc créé un cadre FR_MAKE_AGREEMENT_ON_ASSESSMENT, qui permettra d'annoter les phrases comme (108).

(108) a. [...] nous sommes **TOMBÉS D'ACCORD** sur des propositions de réformes économiques concrètes et utiles, fondées sur des mesures microéconomiques.

b. [...] nous avons **CONVENU** que le Zimbabwe serait le cinquième point [...]

FR_MAKE_AGREEMENT_ON_ASSESSMENT: Cognizer_1, Cognizer_2, Cognizers, Opinion, Topic

6.4.3 Intégration du but au domaine

Encore plus à la marge du domaine, nous avons décidé d'intégrer les cadres évoquant le but ou l'intention d'un cogniteur. En effet, bien qu'elle intègre aussi d'autres composantes, on peut arguer qu'une intention implique une position cognitive stative sur l'opportunité d'effectuer une action.

Dans le cadre PURPOSE, un agent souhaite ou projette d'accomplir une action, d'atteindre un but. Si on considère que cela revient à considérer opportun d'effectuer cette action, on a là une version *.to_do* de FR_AWARENESS-CERTAINTY-OPINION. Ce cadre comportait des rôles noyaux qui n'avaient pas lieu d'être pour les déclencheurs du français, et a donc été modifié pour le French FrameNet en FR_PURPOSE. Le cadre a à cette occasion été enrichi d'un rôle Medium, en ensemble nucléaire avec l'agent. Ce rôle permet d'annoter aussi les phrases comme (109) et plus simplement celles du type (110).

(109) Le Plan préconise un OBJECTIF de 37 heures hebdomadaires d'ici à 1997.

(110) [...] le gouvernement a annoncé son OBJECTIF de réduire le déficit de 30 000 milliards de lires [...]

FR_PURPOSE: Agent, Medium, Goal, Time_of_eventuality, Means

Le cadre DECIDING décrit un cogniteur prenant une décision à propos d'une entité ou d'une action à effectuer (ou mesure à prendre). Nous avons utilisé ce cadre pour créer FR_DECIDING, qui comporte en plus un rôle *Topic* pour le sujet sur lequel porte la décision. Nous avons de plus spécifié

que cette décision pouvait porter sur une action à effectuer par le cogniteur ou sur toute situation sur laquelle ledit cogniteur aurait autorité, mais pas une entité choisie parmi d'autres. Ce dernier cas évoquerait le cadre BFN CHOOSING, qui n'appartient pas à nos domaines. FR_DECIDING, qui peut être évoqué par des déclencheurs tels que *décider*, *se résoudre* ou *prendre le parti*, est inchoatif de FR_PURPOSE.

Le cadre RENUNCIATION, quant à lui, décrit le cas où un locuteur (Speaker) déclare ne plus vouloir avoir affaire à un contenu (Content). Ce cadre a été modifié pour mieux correspondre à la version française de son déclencheur typique : *renoncer*. FR_RENUNCIATION décrit les situations dans lesquelles un cogniteur ne considère plus une éventualité particulière (Eventuality) comme son objectif, ou accepte de ne plus profiter d'un privilège (Benefit). Ce cadre est donc terminatif de FR_PURPOSE.

Un cadre WILLINGNESS proche de FR_PURPOSE a aussi été intégré au domaine. Dans ce cadre, un cogniteur est disposé à effectuer une action. Il n'y est en revanche pas particulièrement déterminé, n'ayant pas pris de décision à cet égard, et c'est ce qui le distingue de FR_PURPOSE. Ce cadre est évoqué par des déclencheurs tels que *prêt*, *disposé* et *réticent*.

Dans le cadre FR_AGREE_OR_REFUSE_TO_ACT, un cogniteur accepte ou refuse d'effectuer une action qui lui a été proposée. Il s'agit de la version « communicative » (cf. section 4.2.5) de WILLINGNESS : *Alex a accepté d'aller chercher le pain* peut être paraphrasé par *Alex a communiqué qu'elle était disposée à aller chercher le pain*. Ce cadre, qui peut être évoqué par des déclencheurs comme *accepter*, appartient aussi au domaine de la communication langagière. Lorsque la proposition à laquelle répond un locuteur n'est pas explicitement une action, c'est le cadre RESPOND_TO_PROPOSAL et pas FR_AGREE_OR_REFUSE_TO_ACT qui est évoqué. Les déclencheurs de ce cadre sont par exemple *repousser* et *rejeter*.

FR_AGREE_OR_REFUSE_TO_ACT est en relation d'héritage avec FR_AGREE_OR_REFUSE_TO_ACT_IN_FAVOR_OF, dans lequel un cogniteur accepte ou refuse d'effectuer une action pour un bénéficiaire. Ce cadre peut être évoqué par des déclencheurs comme *consentir* et *refuser* dans leur construction sous-catégorisant un à-objet (*Le patron d'Alex lui a consenti/refusé une augmentation*).

Enfin, un cadre FR_MEANS_FOR_PURPOSE a été créé pour permettre d'annoter les cas où une action est menée par un agent déprofilé pour lui permettre d'accomplir un but. C'est le cas des phrases comme *l'enquête de police se poursuit AFIN DE savoir qui a fait quoi*. D'autres déclencheurs typiques de ce cadre sont *dans le but de*, *pour* et *en vue de*.

6.5 Positions cognitives portant sur la légitimité d'un contenu ou d'une entité

6.5.1 Cadres statifs de position sur la légitimité

L'examen de certains déclencheurs, en particulier *accepter*, qui évoque aussi FR_AWARENESS-CERTAINTY-OPINION et plusieurs cadres partageant des déclencheurs avec BE_IN_AGREEMENT_ON_ACTION nous ont amenés à examiner le cadre TAKING_SIDES. Dans ce dernier, un cogniteur a un point de vue statif, positif ou négatif, sur une question (Issue), un des participants du débat sur cette question (Side), ou une action effectuée par ce participant (Action). Nous avons décidé d'intégrer ce cadre au domaine des positions cognitives, à la fois parce qu'il traite d'une position cognitive quant à la légitimité d'un contenu ou d'une personne et parce que plusieurs déclencheurs pouvant l'évoquer évoquent déjà d'autres cadres de nos domaines et que nous aurions à les annoter de toutes façons. Nous avons à partir de ce cadre créé FR_TAKING_SIDES, qui en diffère par le fait que nous avons fusionné en un même rôle Content (contenu) les rôles Issue et Action, dont la différence ne nous a pas semblée claire au vu des annotations BFN 1.5. Nous avons de plus spécifié que tous les déclencheurs de FR_TAKING_SIDES devraient être capables d'inclure l'acte de communication de cette position par le cogniteur à un interlocuteur.

Pour les déclencheurs correspondant à la définition de FR_TAKING_SIDES mais ne pouvant pas impliquer d'acte de communication, nous avons créé le cadre FR_BEING_IN_FAVOR_OF. Ce dernier rassemble donc les déclencheurs qui réfèrent à une situation où un cogniteur a un point de vue positif ou négatif sur un contenu, c'est-à-dire une position cognitive sur la légitimité de ce contenu, mais ne peuvent signifier que cette position cognitive a été communiquée à un interlocuteur. Ainsi, (111a) est annoté en FR_TAKING_SIDES tandis que (111b) évoque FR_BEING_IN_FAVOR_OF.

(111) a. Les Cointreau **CONTESTENT** la création de Rémy et associés en 1986 par les Hériard Dubreuil [...]

FR_TAKING_SIDES: Cognizer, Content, Medium, Side

b. Les grandes compagnies américaines **VOYAIENT D'UN MAUVAIS ŒIL** l'incursion du britannique sur leur territoire [...]

FR_BEING_IN_FAVOR_OF: Cognizer, Content, Side

Il existe des prédicats communicationnels qui correspondent à la définition de FR_TAKING_SIDES mais dont le sens n'inclut pas la polarité de la position sur la légitimité du contenu. Ces prédicats sous-catégorisent un marqueur de polarité. C'est le cas par exemple de *prendre parti* : on prend parti *pour* ou *contre* quelqu'un ou quelque chose. Pour ces déclencheurs communicatifs de positions sur la légitimité d'un contenu, nous avons créé le cadre FR_EXPRESSING_SIDE, qui prévoit — en plus des rôles de FR_TAKING_SIDES — un rôle pour le marqueur de polarité susmentionné (Polarity). Ces marqueurs de polarité sont eux-même des déclencheurs de FR_BEING_IN_FAVOR_OF, comme on peut le voir dans les exemples construits présentés en (112). Néanmoins, nous verrons section 7.1.2.1 que lorsqu'une préposition qui pourrait déclencher un cadre apparaît en position régie par un prédicat, elle n'est pas annotée comme déclencheur. Ainsi, dans l'exemple (113a), *se prononcer* évoque FR_EXPRESSING_SIDE, mais *pour*, considéré comme régi, n'évoque pas FR_BEING_IN_FAVOR_OF comme il le ferait en (113b). Cela permet d'éviter une annotation redondante.

(112) a. Sam est **CONTRE** le projet de loi.

b. Alex est **EN FAVEUR** de l'avortement libre et gratuit.

FR_BEING_IN_FAVOR_OF: Cognizer, Content, Side

(113) a. [...] la commune **SE PRONONCE** pour une nouvelle convention de balayage des caniveaux et de nettoyage des avaloirs des eaux pluviales sur la voie publique.

FR_EXPRESSING_SIDE: Content, Polarity, Side, Speaker, Topic

b. La commune est **POUR** une nouvelle convention de balayage des caniveaux et de nettoyage des avaloirs des eaux pluviales sur la voie publique.

FR_BEING_IN_FAVOR_OF: Cognizer, Content, Side

6.5.2 Cadres causatifs de position sur la légitimité

Après avoir décidé d'intégrer à notre domaine les cadres vus à la section précédente (cf. 6.5.1), nous avons examiné des prédicats ressemblant à ceux d'autres cadres du « cœur de notre domaine » mais impliquant une position sur la légitimité d'un contenu plutôt que sur sa véridicité.

Ainsi, nous avons créé un pendant à FR_COGNIZER_AFFECTING.VERACITY et FR_COGNIZER_AFFECTING.TO_DO : FR_COGNIZER_AFFECTING.LEGITIMACY pour les cas où quelqu'un ou quelque chose cause chez un cogniteur une position cognitive sur la légitimité d'un contenu. Il regroupe les déclencheurs *convertir* et *rallier*, et permet d'annoter des exemples comme (114).

(114) Cette participation a RALLIÉ la CUT elle-même à l'idée de la privatisation.

FR_COGNIZER_AFFECTING.LEGITIMACY: Action, Cognizer, Medium, Persuader, State_of_affairs

Les ATTEMPT_SUASION ont eux aussi une version *legitimacy* : FR_ATTEMPT_SUASION.LEGITIMACY rassemble les déclencheurs décrivant une situation dans laquelle un être doué de parole (le Persuader) tente de convaincre un interlocuteur (Addressee) de la légitimité — ou non-légitimité — d'une situation (State_of_affairs). Ce cadre, qui rassemble les déclencheurs comme *défendre* et *prôner*, permet d'annoter les exemples comme (115). FR_ATTEMPT_SUASION.LEGITIMACY diffère de FR_TAKING_SIDES en ce que ses déclencheurs encodent intrinsèquement l'idée que le Persuader souhaite *convaincre* l'interlocuteur de la légitimité du contenu et pas simplement affirmer croire en cette légitimité.

(115) [...] René Capitant [...] PRÊCHAIT à ses « compagnons » la révolte contre la « nouvelle société » de Jacques Chaban-Delmas.
FR_ATTEMPT_SUASION.LEGITIMACY: Addressee, Medium, Persuader, Polarity, State_of_affairs

6.5.3 Cadres portant sur la légitimité d'éventualités passées

FR_TAKING_SIDES doit aussi être mis en contraste avec le cadre JUSTIFYING, dans lequel un agent fournit une raison visant à légitimer un acte qu'il a commis ou omis, ou une situation dont il est responsable. À partir de ce cadre, nous avons créé FR_JUSTIFYING, dans lequel l'acte ou la situation, parfois difficiles à distinguer, sont englobés par un même rôle Eventuality. L'agent a été renommé Speaker, et la raison fournie par ce dernier pour justifier l'Eventuality a été renommée Sufficient_reason. Enfin, le Speaker n'est pas nécessairement responsable de l'Eventuality : un nouveau rôle Responsible_entity permet de couvrir les cas dans lesquels un Speaker justifie les actions d'une autre personne, comme c'est le cas en (116). Ce cadre diffère de FR_TAKING_SIDES en ce qu'avec FR_JUSTIFYING, la situation qu'on cherche a légitimer doit s'être déjà produite, ou une décision en ce sens déjà avoir été prise, tandis que FR_TAKING_SIDES tend à légitimer des actions futures, des décisions non actées.

(116) Sam a DÉFENDU Alex en expliquant qu'elle n'avait pas fait exprès.
FR_JUSTIFYING: Addressee, Eventuality, Medium, Responsible_entity, Speaker, Sufficient_reason

Dans le cadre DESERVING l'existence d'une situation est une raison suffisante pour qu'une action soit considérée comme justifiée, légitime (*L'heure tardive justifie le départ d'Alex*). JUSTIFYING semble correspondre à l'attribution d'un évènement DESERVING à un locuteur : si *Alex justifie son départ par l'heure tardive*, alors *Alex considère que l'heure tardive justifie son départ*, ou du moins dit que c'est le cas. Les exemples supra montrent que les deux cadres peuvent être évoqués par les mêmes déclencheurs, il nous a donc paru informatif d'inclure par extension le cadre DESERVING au domaine, malgré l'absence de cogniteur. Nous l'avons néanmoins modifié pour permettre la justification d'éventualités aussi bien que d'actions : notre version du cadre se nomme FR_DESERVING.

Notons ici qu'un cadre FR_ATTRIBUTING_CAUSE assez proche de JUSTIFYING a été créé pour le French FrameNet et est considéré appartenir au domaine des positions cognitives aussi bien qu'à celui de la cause. Dans ce cadre, un cogniteur présente une éventualité comme ayant été causée par une cause ou par une entité. Il s'agit en fait de l'expression par un locuteur d'une position cognitive stative sur un lien de causalité. Ce cadre peut aussi être évoqué par *justifier*, mais il couvre les cas où la justification est présentée comme une cause et pas une légitimation de l'éventualité. L'exemple (117) présente des exemples d'utilisation du déclencheur *justifier* par les trois cadres présentés supra.

(117) a. Dans sa réponse, le ministère des postes et télécommunications **JUSTIFIE** la filialisation du transport de fonds par un souci de clarification.

FR_JUSTIFYING: Addressee, Eventuality, Medium, Responsible_entity, Speaker, Sufficient_reason

b. La fréquence des défauts, note-t-on au secrétariat d'état, n'est pas « dramatique » mais **JUSTIFIE** une vigilance accrue.

FR_DESERVING: Eventuality, Sufficient_reason

c. Certains ont **JUSTIFIÉ** les premiers succès de l' offensive allemande en 1914 par l'incompétence de ces officiers dont près de la moitié d'entre eux a été limogée par Joffre pour incompétence.

FR_ATTRIBUTING_CAUSE: Attributor, Cause, Known_state_of_affairs, Responsible_party

6.6 Par extension : cadres de jugement

De position sur la légitimité d'un contenu à jugement positif ou négatif d'un contenu, il n'y a qu'un pas. De fil en aiguille, nous avons intégré quatre cadres du domaine Jugement/évaluation (qui n'a pas été annoté) à notre domaine des positions cognitives.

JUDGMENT décrit le jugement de valeur porté par un cogniteur sur une entité (Evaluee). Une raison (Reason) au jugement est de plus sous-catégorisée par les déclencheurs de ce cadre. JUDGMENT diffère de FR_BEING_IN_FAVOR_OF et FR_TAKING_SIDES en ce qu'avec JUDGMENT, le jugement porte sur la valeur d'une entité et pas sa légitimité. JUDGMENT comprend des déclencheurs tels qu'*admirer*, *approuver*, *mépriser* et permet d'annoter des cas comme l'exemple (118).

(118) [...] le général a exprimé son « **ADMIRATION** » aux manifestants [...]
JUDGMENT: Cognizer, Evaluee, Reason, Topic

Le cadre REGARD est très proche de JUDGMENT. Il en diffère par le fait que ses déclencheurs, comme *juger* et *impression*, n'incluent pas la polarité de l'évaluation portée par le cogniteur. Un rôle *Judgment* permet de décrire la position de l'Evaluee sur l'échelle d'approbation du cogniteur.

(119) Les magistrats de la Cour des comptes [...] portent un **JUGEMENT** très sévère sur la création en 1987 par La Poste de la filiale spécialisée Sécuripost.
REGARD: Cognizer, Evaluee, Judgment, Reason

Dans JUDGMENT_COMMUNICATION, l'équivalent avec communication du cadre JUDGMENT, un locuteur (Communicator) communique à un interlocuteur (Addressee) un jugement sur une entité (Evaluee). Comme JUDGMENT, le cadre permet d'exprimer une raison pour le jugement. Nous avons à partir de JUDGMENT_COMMUNICATION créé un cadre FR_JUDGMENT_COMMUNICATION dans lequel l'entité jugée et le sujet sur lequel porte le jugement sont en ensemble nucléaire. Nous avons de plus ajouté au cadre un rôle Message, qui permet de décrire le jugement et s'avère utile pour les nominalisations qui réfèrent au message prononcé par le locuteur, comme on peut le voir en (120).

(120) La **CRITIQUE** de Sam selon laquelle Alex ne fait jamais la vaisselle ne paraît pas infondée.
FR_JUDGMENT_COMMUNICATION: Addressee, Communicator, Evaluee, Medium, Message, Reason, Topic

Pour finir, nous avons créé un cadre FR_JUDGMENT_COMMUNICATION_NO_REASON, qui regroupe les déclencheurs décrivant une situation similaire à celle de FR_JUDGMENT_COMMUNICATION mais ne permettant pas de réalisation locale de la raison pour laquelle le locuteur porte un tel jugement sur l'Evaluee. Ainsi, *médire* évoque FR_JUDGMENT_COMMUNICATION_NO_REASON tandis

que *critiquer* évoque FR_JUDGMENT_COMMUNICATION. En effet, on peut *critiquer Sam d'avoir oublié d'aller chercher le pain* mais pas **médire de Sam d'avoir oublié d'aller chercher le pain.*

L'ensemble des cadres que nous avons décrits dans ces sections est présenté plus haut, table 6.1, qui les trie en récapitulant leurs principales caractéristiques. Notons ici que nous avons aussi créé sept cadres non lexicaux permettant de faire le lien entre nos cadres de position cognitive stricto sensu, ceux portant sur l'opportunité d'effectuer une action, et ceux traitant de la légitimité d'une situation. Ces cadres (FR_ATTEMPT_SUASION, FR_BE_IN_AGREEMENT, FR_COGNIZER_AFFECTING, FR_EXPRESSING_STANCE, FR_MAKE_AGREEMENT, FR_PERSONAL_STANCE et FR_REMEMBERING) ne sont pas représentés dans la table susmentionnée, qui ne reprend que les cadres lexicaux, mais ils apparaissent (en grisé) dans les figures 6.1 et 6.2 qui exposent les relations qui lient les cadres du domaine entre eux.

6.7 Relations entre les cadres du domaine des positions cognitives

Nous avons observé toutes les relations déjà définies entre les cadres BFN que nous avons conservés, ainsi qu'entre les cadres BFN que nous avons modifiés pour en faire des cadres spécifiques au French FrameNet. À cette occasion, nous avons aussi quand cela était pertinent défini de nouvelles relations entre nos cadres.

Nous avons tenté d'appliquer systématiquement et formellement les critères FrameNet décrits en 3.2.1.4 pour déterminer une relation entre deux cadres. Comme nous l'avons précisé en 4.2.5, nous n'avons pas conservé les relations d'usage ni de *see also* pour le French FrameNet. Il est aussi arrivé plusieurs fois que nous supprimions une relation en place dans le BFN, par exemple parce que les contraintes sur les rôles n'étaient plus respectées avec nos versions des cadres. Prenons l'exemple des cadres BFN RESPOND_TO_PROPOSAL (un locuteur répond positivement ou négativement à la proposition d'un interlocuteur) et AGREE_OR_REFUSE_TO_ACT (un agent accepte ou non d'effectuer une action proposée par un interlocuteur), qui sont en relation d'héritage. Pour le French FrameNet, AGREE_OR_REFUSE_TO_ACT est devenu FR_AGREE_OR_REFUSE_TO_ACT car le rôle de l'interlocuteur n'est réalisable localement pour aucun déclencheur de ce cadre (*accepter, consentir, refuser...*). Un cadre doit hériter des rôles de son parent, et FR_AGREE_OR_REFUSE_TO_ACT n'a plus de rôle pour l'interlocuteur, noyau dans RESPOND_TO_PROPOSAL. Il ne peut donc plus hériter de ce dernier.

De même, il est arrivé que nous ajoutions entre des cadres BFN des relations qui n'avaient pas été spécifiées, ou changions le type d'une relation entre deux cadres. Cela s'est souvent produit avec des relations de type *usage*, dont Ovchinnikova *et al.* (2010) précisent qu'elle est souvent utilisée pour lier deux cadres dans des cas où une autre relation aurait pu convenir. Ainsi pour le BFN, le cadre MAKE_AGREEMENT_ON_ACTION utilise le cadre BE_IN_AGREEMENT_ON_ACTION, alors que nous avons pour le French FrameNet considéré que MAKE_AGREEMENT_ON_ACTION était causatif de BE_IN_AGREEMENT_ON_ACTION : *convenir* d'une action est causatif d'un *accord* quant à cette action.

Lorsque cela s'appliquait, nous avons aussi ajouté les relations que nous avons définies pour le French FrameNet aux cadres concernés. Ainsi, les cadres JUSTIFYING et DESERVING ne sont dans le BFN liés par aucune relation. Dans le premier de ces cadres, un agent fournit une raison visant à légitimer un acte qu'il a commis ou omis, ou une situation dont il est responsable (*Alex justifie son départ par l'heure tardive*), tandis que dans le deuxième, l'existence d'une situation est une raison suffisante pour qu'une action soit considérée comme justifiée (*L'heure tardive justifie le départ d'Alex*). Nous avons pour le French FrameNet défini une relation d'*attribution* entre nos versions des deux cadres : FR_JUSTIFYING, dont on voit un exemple en (121a), est attributif de FR_DESERVING, exemplifié en (121b). Dans FR_JUSTIFYING, la légitimation ou justification est attribuée à un participant supplémentaire, le Speaker.

(121) a. Dans sa réponse, le ministère des postes et télécommunications **JUSTIFIE** la filialisation du transport de fonds par un souci de clarification.

FR_JUSTIFYING: Addressee, Eventuality, Medium, Responsible_entity, Speaker, Sufficient_reason

b. La fréquence des défauts, note-t-on au secrétariat d'état, n'est pas « dramatique » mais **JUSTIFIE** une vigilance accrue.

FR_DESERVING: Eventuality, Sufficient_reason

Rappelons ici que nous avons choisi de ne pas conserver les relations entre les cadres du French FrameNet et les cadres BFN n'appartenant pas à nos domaines. Si ces cadres avaient été adaptés pour le français, ils auraient possiblement été modifiés. Sans connaissance de ces potentielles modifications, il nous est impossible de vérifier que les contraintes sur les relations entre cadres sont respectées.

Les figures 6.1 et 6.2 permettent de visualiser l'ensemble des relations liant entre eux des cadres du domaine des positions cognitives, ainsi que les cadres d'autres domaines auxquels nos cadres sont liés. Les cadres non lexicalisés y figurent en gris. Les cadres lexicalisés apparaissent en noir lorsqu'ils appartiennent au domaine des positions cognitives (ce qui ne les empêche pas d'appartenir aussi à d'autres domaines). Les cadres qui apparaissent en bleu sont ceux qui n'appartiennent pas au domaine des positions cognitives.

FIGURE 6.1: Représentation des relations au sein du domaine des positions cognitives

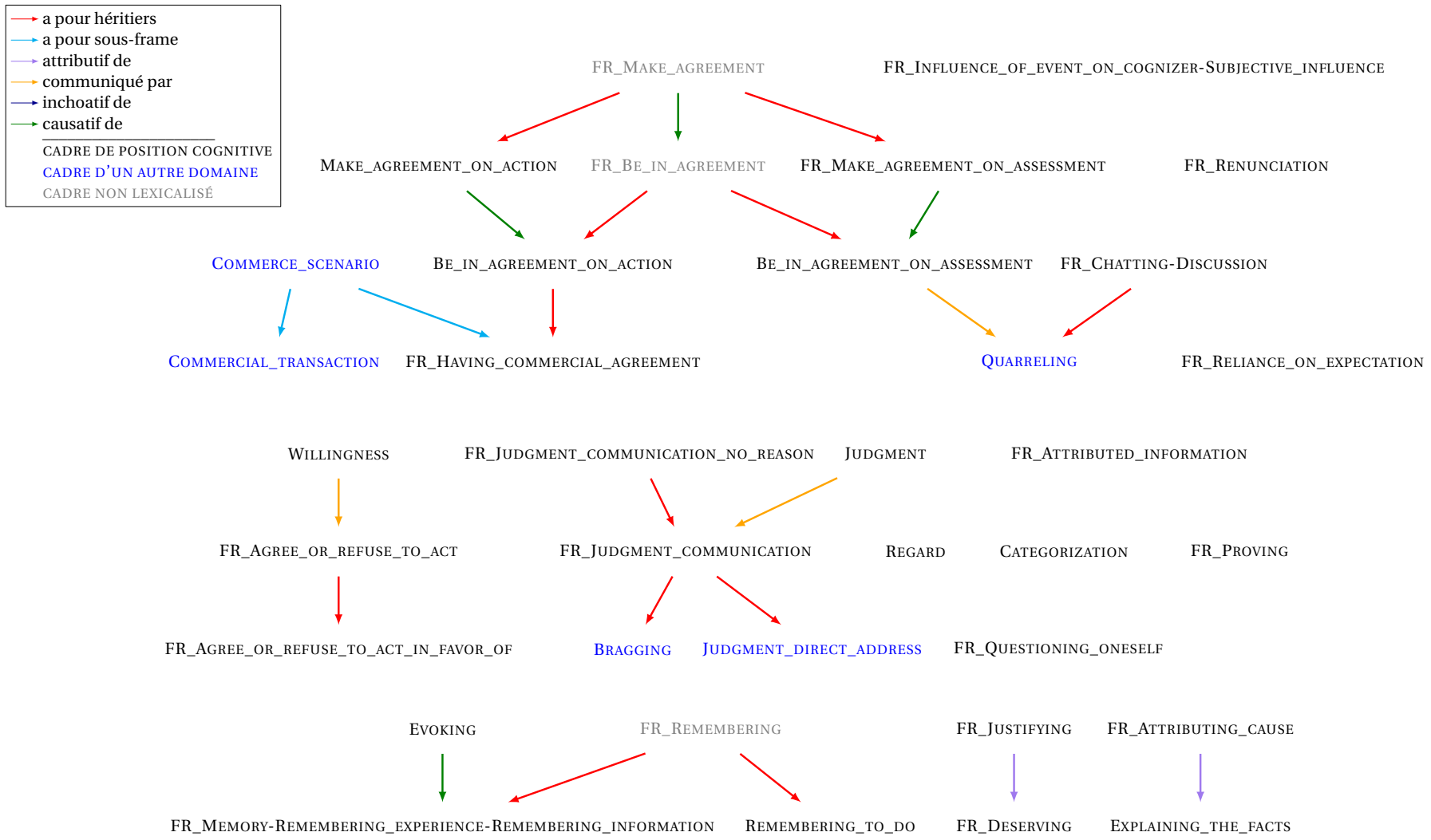


FIGURE 6.2: Suite de la représentation des relations au sein du domaine des positions cognitives

CHAPITRE 7

ANNOTATION

Sommaire

7.1 Principes généraux pour l'annotation **136**
7.1.1 Annotation des cadres 136
7.1.1.1 Annotation indépendante de la factualité et de la spécificité . 136
7.1.1.2 Emplois métaphoriques et figurés 136
7.1.1.3 Ellipses de déclencheur 137
7.1.2 Annotation des rôles et de leurs remplisseurs 139
7.1.2.1 Contenu des remplisseurs de rôles 139
7.1.2.2 Traitement des anaphores 140
7.1.2.3 Localité syntaxique 140
7.1.2.4 Saturation des remplisseurs prédicatifs 142
7.1.2.5 Instanciation nulle 147
7.1.2.6 Astuce dans le cas de déclencheurs qui dépendent syntaxiquement d'un de leurs arguments sémantiques 149
7.1.3 Constructions à verbes supports 151
7.2 Déroulement du travail d'annotation **153**
7.2.1 Corpus utilisés 153
7.2.2 Pré-annotation automatique 154
7.2.2.1 Pré-annotation des verbes pronominaux 155
7.2.2.2 Pré-annotation des déclencheurs composés 156
7.2.3 Annotation 157
7.2.4 Adjudication 160
7.2.5 Post-traitement 160
7.2.6 Documents et outils utilisés tout au long du cycle d'annotation 162
7.3 Points spécifiques **163**
7.3.1 Constructions à attribut de l'objet 163
7.3.1.1 Différentes catégories d'attributs 163
7.3.1.2 Complément vs. modifieur 164
7.3.1.3 Typologie sémantique 167
7.3.1.4 Traitement dans le FrameNet de Berkeley 169
7.3.1.5 Choix de traitement pour le French FrameNet 173
7.3.2 Déclencheurs nominaux 174
7.3.2.1 Emplois référentiels vs. emplois prédicatifs 175
7.3.2.2 Annotation des noms de participants 175
7.3.2.3 Liste non exhaustive des emplois prédicatifs d'un nom 176
7.3.2.4 Nominalisations ambiguës 177

Après avoir étudié les premières étapes de notre travail de création du French FrameNet, à savoir la délimitation d'un réseau de cadres et la constitution du lexique de ces cadres, nous nous penchons dans ce chapitre sur l'annotation en cadres et en rôles de corpus arborés, le French Treebank (Abeillé *et al.*, 2003) et le Sequoia Treebank (Candito et Seddah, 2012). Dans la première section, nous abordons les principes généraux que nous avons définis et suivis, avant de décrire le déroulement du travail d'annotation dans une deuxième section. Dans la troisième section, nous nous intéressons à plusieurs points spécifiques pour lesquels trouver un traitement harmonieux a nécessité une étude plus approfondie.

7.1 Principes généraux pour l'annotation

Nous décrivons ici les lignes directrices que nous avons posées et suivies quant à l'annotation, en nous limitant aux points qui nous paraissent les plus significatifs. Pour davantage d'informations, se référer au guide d'annotation fourni en annexe (annexe B).

7.1.1 Annotation des cadres

7.1.1.1 Annotation indépendante de la factualité et de la spécificité

Nous avons choisi d'annoter toutes les occurrences de cadres évoqués par nos déclencheurs, et ce sans tenir compte de la factualité de l'éventualité décrite. Cela signifie que l'on annote dès lors que l'idée du cadre est évoquée, même si elle ne correspond pas à une réalisation effective. Un déclencheur au futur ou au conditionnel, ou enchâssé dans une construction changeant sa polarité (une négation par exemple) sera annoté au même titre qu'une occurence du même déclencheur dans un contexte renforçant sa factualité, comme une phrase affirmative au présent ou au passé. En effet, nous avons considéré que l'information sur la factualité du cadre pourrait, au moins dans une certaine mesure, être calculée compositionnellement, et qu'il n'était donc pas de notre ressort de tenir compte de ces distinctions.

Annoter dès lors que l'idée d'un cadre est évoquée signifie aussi que nous considérons que les occurrences génériques de déclencheurs, pour lesquelles le déclencheur ne réfère à aucun élément spécifique du monde, évoquent tout de même le cadre concerné par cette idée. Nous annotons de même les déclencheurs au sein d'entités nommées dès lors que ces noms d'entités sont compositionnels et que les déclencheurs y apparaissent dans le sens voulu.

7.1.1.2 Emplois métaphoriques et figurés

Le traitement BFN de la métaphore s'articule autour de la différence entre métaphores productives et métaphores lexicalisées, où les métaphores productives sont celles pour lesquelles :

— tous les synonymes du déclencheur en question présentent la même alternation entre lecture littérale et lecture métaphorique ;
— chaque rôle du cadre de la lecture métaphorique correspond à un rôle du cadre de la lecture littérale ;
— et la phrase ne peut être comprise qu'en faisant appel au sens littéral.

Lorsque ces critères sont remplis, le déclencheur métaphorique est annoté avec le cadre de son sens littéral, et signalé grâce à un marquage spécial. Dans le cas contraire, on considère avoir affaire à une métaphore lexicalisée, c'est-à-dire un autre sens du déclencheur (ou d'une expression polylexicale contenant le déclencheur). C'est alors cet autre sens qui est pris en compte et annoté par un cadre si le sens est couvert par nos domaines, ou comme *Other_sense* dans le cas contraire [1].

1. Nous verrons en 7.2 que nous avons utilisé un pseudo-cadre appelé *Other_sense* pour marquer les cas où un

Par contraste, Burchardt *et al.* (2009b) indiquent que pour le projet SALSA de FrameNet de l'allemand, les expressions métaphoriques ont été annotées à l'aide de deux cadres : le cadre *source*, pour le sens littéral, et le cadre *cible* pour le sens figuré. La terminologie provient de la notion de transfert métaphorique de Lakoff et Johnson (1980), et l'intérêt de la double annotation est de permettre de modéliser ce transfert métaphorique en étudiant par exemple la correspondance des rôles du cadre source et ceux du cadre cible.

Pour le French FrameNet, nous avons choisi de procéder comme le BFN, et d'annoter uniquement le sens source des métaphores productives tout en les signalant d'un marquage spécifique appelé *METAPHOR*. Cette solution n'est pas incompatible avec celle appliquée pour SALSA. Le marquage METAPHOR permettra aisément de repérer les cas de métaphores et d'y ajouter les cadres cibles dans un futur travail.

Nous avons réutilisé les deux premiers critères de Ruppenhofer *et al.* (2006) pour distinguer polysémie et métaphore, à savoir la substituabilité avec des synonymes et la surjection des rôles du cadre littéral sur ceux de l'emploi métaphorique. À ces critères nous avons ajouté celui qu'à l'inverse, si le sens de l'emploi en question était distinct, il devait pouvoir s'utiliser dans d'autres contextes linguistiques.

Dans la phrase (122), ne pas annoter ARRIVING sur *gagnent* signifierait qu'il s'agit d'un autre sens, où *gagner* signifie en quelque sorte *contaminer*. Or, on peut remplacer gagner par ses synonymes au sein d'ARRIVING, comme *arriver*, *atteindre* ou *débarquer*. De plus, le cadre qui couvrirait *contaminer* aurait deux rôles : l'élément contaminant, et le corps contaminé. L'élément contaminant correspond au rôle Theme d'ARRIVING et le corps contaminé à son rôle Goal. Notre troisième critère ne s'applique pas puisqu'il fonctionne uniquement dans l'autre sens, mais le dernier critère BFN, selon lequel le locuteur a nécessairement à l'esprit le sens source pour interpréter la phrase confirme bien le diagnostic de métaphore productive.

(122) Les jours suivants, les manifestations **GAGNENT** d'autres villes de Transylvanie
[...]
ARRIVING : Goal, Theme

7.1.1.3 Ellipses de déclencheur

Certaines coordinations mettent en jeu une ellipse du prédicat, et la partie elliptique peut correspondre sémantiquement à une autre instance d'événement. On choisit dans ces cas d'annoter deux instances du même cadre, sur la même occurrence de déclencheur, et de signaler la deuxième instance de cadre d'un marquage spécial. Il existe plusieurs types d'ellipse de déclencheur.

« Head gapping »

Il s'agit des cas comme l'exemple construit (123) : des coordinations où le verbe n'est pas répété et on a deux sujets différents. On a alors à la fois deux instances de cadres et deux jeux de remplisseurs différents. On procède donc à l'annotation de chacune des deux instances sur le déclencheur, comme nous l'avons fait en (124a) et (124b), avec en plus un marquage nommé *HEADGAP* sur la deuxième instance de cadre.

déclencheur potentiel n'évoque aucun cadre de nos domaines. L'utilisation de ce cadre factice permet de faire la différence entre les occurrences de déclencheurs examinées au cours de la phase d'annotation dont on a jugé qu'elles n'évoquaient aucun cadre de nos domaines (annotées *Other_sense*) et les occurrences de déclencheurs qui n'ont jamais été examinées (pas d'annotation de cadre). Cette distinction est nécessaire si l'on souhaite que les données soient utilisables pour l'apprentissage automatique, puisque sans elle, la distribution des sens est faussée.

(123) Alex achète des baskets et Sam une casquette.

(124) a. Selon les statistiques de la direction de l'industrie touristique, le marché des cadeaux-souvenirs et "achats à rapporter" (montres, chaussures, vêtements...) a atteint l'an dernier dans l'Hexagone 10,4 milliards de francs, dont 6,4 milliards **ACHETÉS** par les Français en vacances et 4,2 milliards par les touristes étrangers.

b. Selon les statistiques de la direction de l'industrie touristique, le marché des cadeaux-souvenirs et "achats à rapporter" (montres, chaussures, vêtements...) a atteint l'an dernier dans l'Hexagone 10,4 milliards de francs, dont 6,4 milliards **ACHETÉS** par les Français en vacances et 4,2 milliards par les touristes étrangers.

COMMERCE_BUY: Buyer, Goods, Money, Rate, Recipient, Seller

« Argument clusters »

Il s'agit des cas dans lesquels les deux évènements partagent un même participant, qui correspond au sujet du verbe, et qui est comme ce dernier élidé. Comme pour le head gapping, deux instances de cadre sont annotées sur le même déclencheur et la seconde est repérée d'un marquage : *ARGCLUSTER*. C'est ce qui est représenté en (125).

(125) a. Steven Ross avait **ACHETÉ** une agence de jeunes talents en 1967 et, deux ans plus tard, la société cinématographique Warner - Seven Arts, devenue dans les années 70 Warner Communications .

b. Steven Ross avait **ACHETÉ**$_{\text{ARGCLUSTER}}$ une agence de jeunes talents en 1967 et, deux ans plus tard, la société cinématographique Warner-Seven Arts, devenue dans les années 70 Warner Communications.

COMMERCE_BUY: Buyer, Goods, Money, Rate, Recipient, Seller

Deux éléments permettent de distinguer un argument cluster d'un remplisseur de rôle intégrant une coordination. Tout d'abord, les deux instances de cadre doivent avoir au moins un remplisseur différent pour un des rôles à annoter. Si l'argument cluster porte sur les remplisseurs de rôles qui ne sont pas à annoter, comme dans l'exemple construit (126a), on ne fait qu'une seule instance de cadre. Ensuite, les deux remplisseurs distincts ne doivent pas former de coordination sauf si chaque coordonné a un modifieur de temps ou de lieu différent. Ainsi les exemples construits (126a) et (126b) ne sont-ils annotés qu'avec une instance de cadre, tandis que (125) l'est avec deux.

(126) a. Sam a **ACHETÉ** une fleur le matin à Paris, et l'après-midi à Tokyo.

b. Sam a **ACHETÉ** une chemise et un short.

c. Sam a **ACHETÉ** une glace à Alex et à Claude.

COMMERCE_BUY: Buyer, Goods, Money, Rate, Recipient, Seller

Comparatives réduites

Il s'agit de cas comme (127) et (128) où l'on a affaire à deux instances de cadre partageant un ou plusieurs mêmes remplisseurs. Les deux cadres s'insèrent dans une relation de comparaison qui s'appuie sur celui de leurs rôles dont le remplisseur est différent pour chaque instance. Comme pour les autres types d'ellipse de déclencheur, on annote deux instances de cadre, chacune avec ses propres rôles. La deuxième instance porte un marquage *REDUCEDCOMP*.

(127) a. Toutefois, rien ne permet encore de savoir si l'élimination **TOUCHERA** davantage les sortants que les nouveaux venus.

b. Toutefois, rien ne permet encore de savoir si l'élimination **TOUCHERA**$_{\text{REDUCEDCOMP}}$ davantage les sortants que les nouveaux venus.

(128) a. Sa crise d'identité est profonde et **TOUCHE** tout aussi bien la province québécoise que les populations indiennes.

b. Sa crise d'identité est profonde et **TOUCHE**$_{\text{REDUCEDCOMP}}$ tout aussi bien la province québécoise que les populations indiennes.

FR_CONTINGENCY-OBJECTIVEINFLUENCE: Dependent, Influencer

7.1.2 Annotation des rôles et de leurs remplisseurs

Nous avons vu en 4.2.4.3 que notre annotation des rôles se concentrerait sur les remplisseurs de rôles noyaux, mais que nous annoterions aussi les rôles dont le remplisseur était sous-catégorisé par un des déclencheurs du cadre, et ce bien qu'ils ne soient pas sémantiquement obligatoires. Pour l'annotation des remplisseurs des rôles déterminés comme « à annoter », nous avons suivi un certain nombre de principes, dont les plus importants sont décrits ici.

7.1.2.1 Contenu des remplisseurs de rôles

Pour les remplisseurs de rôles, nous annotons non seulement la tête du remplisseur mais aussi tous ses dépendants (sans distinction entre compléments et modifieurs). Ainsi, dans la phrase (129), tous les dépendants de *les résultats* appartiennent au remplisseur du rôle Content du cadre FR_AWARENESS-CERTAINTY-OPINION.

(129) Le 16 juillet, on **CONNAÎTRA** les résultats de l'offre d'échange faite aux personnes physiques, qui porte sur 57 milliards de francs d'obligations d'Etat, dont la durée de vie ne dépasse pas fin juillet 1994, l'offre se limitant à des OAT 9,5 % 1998 (six ans).

FR_AWARENESS-CERTAINTY-OPINION: Cognizer, Content, Evidence, Topic

À l'inverse du FrameNet de Berkeley, nous avons choisi d'exclure les mots introducteurs sémantiquement vides, comme les prépositions régies et le complémenteur *que*, du contenu des remplisseurs. Ainsi le remplisseur du rôle Content de FR_AWARENESS-CERTAINTY-OPINION en (130a) n'inclut-il pas le complémenteur *que* introducteur. De même le remplisseur du rôle Phenomenon du cadre FR_BECOMING_AWARE n'englobe-t-il pas la préposition *de* qui l'introduit en (130b).

(130) a. [...] M. Tapie se dit **CONVAINCU** qu'il sortira blanchi de l' affaire Toshiba [...]

FR_AWARENESS-CERTAINTY-OPINION: Cognizer, Content, Evidence, Topic

b. [...] il **SE RENDRA COMPTE** de son imprudence [...]

FR_BECOMING_AWARE: Cognizer, Evidence, Instrument, Phenomenon, Topic

Lorsqu'un complémenteur ou une préposition régie n'est pas sémantiquement vide, et qu'elle apporte une information sémantique non fournie par le rôle lui-même, elle est considérée comme faisant partie du contenu du remplisseur. C'est le cas pour (131), où la conjonction *si* appartient bien au remplisseur du rôle Content, puisqu'elle y contribue une information importante.

(131) [...] on **IGNORE** si la banque centrale aurait les moyens psychologiques et financiers d'enrayer une nouvelle chute du rouble.

FR_AWARENESS-CERTAINTY-OPINION: Cognizer, Content, Evidence, Topic

Une préposition sémantiquement pleine mais sous-catégorisée par un prédicat sera donc annotée comme faisant partie du remplisseur de rôle qu'elle introduit si ledit prédicat appartient à notre lexique de déclencheurs. Néanmoins, si cette préposition est elle-même susceptible de déclencher un cadre, ce dernier ne sera pas évoqué par les occurrences sous-catégorisées de cette préposition. Ainsi la préposition *contre* n'évoque-t-elle pas le cadre FR_BEING_IN_FAVOR_OF en (132a) comme elle le ferait en (132b) (exemple construit).

(132) a. Deux tiers des Allemands **SE PRONONCERAIENT** « contre une monnaie européenne commune » [...]
FR_EXPRESSING_SIDE: Content, Polarity, Side, Speaker, Topic

b. Deux tiers des Allemands sont « **CONTRE** une monnaie européenne commune » [...]
FR_BEING_IN_FAVOR_OF: Cognizer, Content, Side

Par cohérence avec les principes qui précèdent, en cas de coordination ou d'énumération avec répétition d'une préposition régie ou d'un complémenteur sémantiquement vide, aucune des pépositions ou des complémenteurs ne fait partie du contenu du remplisseur de rôle. Comme on peut le voir en (133), les conjonctions de coordination, porteuses de sens, y sont en revanche intégrées, de même que la ponctuation des énumérations.

(133) a. Ma thèse est que nous devons **PENSER** qu'il a un objectif et que l' atteindre va dans le sens du meilleur des intérêts du parti communiste soviétique.
FR_AWARENESS-CERTAINTY-OPINION: Cognizer, Content, Evidence, Topic

b. [...] celui qui [...] créa l'agence alliance photo en 1934 est **PARTISAN** du trucage, du photomontage et du métissage des genres qu'il utilise comme un moyen pour explorer de nouveaux domaines.
FR_BEING_IN_FAVOR_OF: Cognizer, Content, Side

7.1.2.2 Traitement des anaphores

Un rôle est parfois réalisé localement par un élément anaphorique tel qu'un pronom ou un déterminant possessif. Nous avons alors comme dans les autres cas annoté cet élément comme remplisseur du rôle en question. De plus, lorsque l'antécédent de ladite anaphore apparaît dans la même phrase, nous l'annotons aussi comme remplisseur du même rôle et distinguons les deux remplisseurs —anaphore et antécédent— au moyen de deux marqueurs appelés *ANAPH* pour l'anaphore et *FULLANTECENDENT* pour l'antécédent. Ainsi en (134), le rôle Goods est réalisé localement par un pronom anaphorique et est aussi utilisé pour annoter son antécédent qui apparaît dans la phrase.

(134) a. Paribas était l'ancienne maison mère de Ciments Français$_{\text{FULLANTECEDENT}}$ avant de **CÉDER** ceux-ci$_{\text{ANAPH.}}$ à leur concurrent italien Italcement.
COMMERCE_SELL: Buyer, Goods, Money, Rate, Seller

Annoter l'anaphore, réalisée localement, nous permet de capturer les régularités syntaxiques de réalisation des divers rôles. Signaler aussi son antécédent présente l'intérêt à la fois sémantique et lexical de permettre de récupérer la tête lexicale qui remplit le rôle. De plus, les annotations résultantes pourront être rassemblées et utilisées dans le cadre de travaux sur la résolution d'anaphores.

7.1.2.3 Localité syntaxique

L'annotation lexicographique BFN a été effectuée sur des phrases choisies car elles exemplifiaient la variété des réalisations possibles pour les différents rôles de chaque cadre. De ce fait,

les remplisseurs de rôles sont généralement locaux aux déclencheurs, c'est-à-dire, à l'exception des sujets, qu'ils appartiennent à la projection maximale dont le déclencheur est la tête. Comme l'expliquent Ruppenhofer *et al.* (2006, 3.2.6, pp. 27-30), il existe néanmoins deux types de situation dans lesquelles des remplisseurs non-locaux sont annotés pour les rôles. Dans le cas de constructions syntaxiques particulières telles que la montée ou le contrôle, ou encore les constructions à verbe support, les arguments du verbe support, à montée ou à contrôle peuvent aussi être annotés comme remplisseurs des rôles du déclencheur. Par ailleurs, lorsqu'un déclencheur se trouve à l'intérieur d'une proposition relative, si le pronom relatif remplit un rôle, il est annoté comme tel et son antécédent l'est aussi. Si ledit pronom n'est pas réalisé, comme cela arrive en anglais, seul l'antécédent sera annoté comme remplisseur du rôle. En pratique néanmoins, les annotations BFN ne suivent pas toujours ces directives et il arrive de rencontrer des remplisseurs non-locaux à leur déclencheur ne rentrant dans aucun de ces cas de figure.

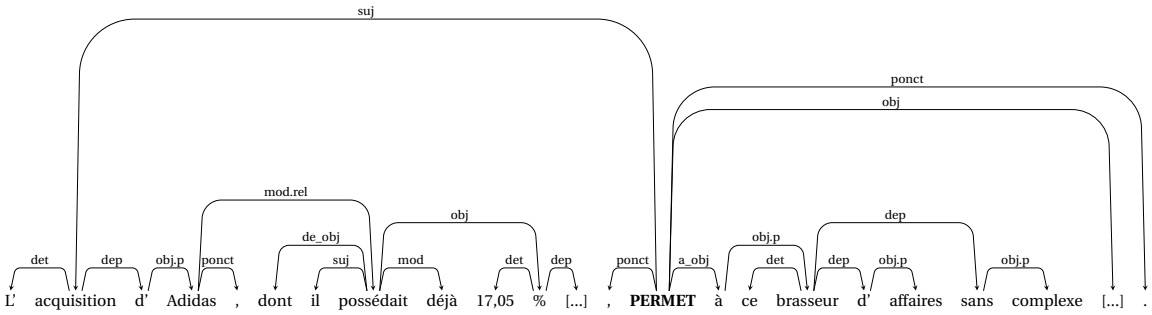
Pour le French FrameNet, nous avons choisi d'annoter les remplisseurs de rôles présents dans la phrase du déclencheur y compris lorsque le rôle n'est pas réalisé localement. Nous avons utilisé un marquage spécial (appelé *NLI* pour *instanciation non locale*) pour distinguer les remplisseurs non locaux des autres, ainsi, leur repérage ne nécessitera pas de calcul syntaxique. Capturer ces remplisseurs non-locaux n'apporte pas d'information sur le linking, la possible réalisation syntaxique des rôles, mais pourra par exemple permettre de travailler sur les types sémantiques des remplisseurs des différents rôles.

Afin de pouvoir facilement déterminer le caractère local ou non d'un remplisseur donné, nous avons défini un domaine de localité précis (en termes de dépendances) pour chaque type de déclencheur de notre lexique. Ainsi pour un déclencheur verbal, sont locaux les dépendants syntaxiques directs (ou via prépositions), ainsi que le nom modifié en cas de participiale épithète : en (135), seuls les rôles Goods et Seller sont donc locaux. Pour un déclencheur adjectival, nous avons considéré comme locaux les dépendants de l'adjectif, le nom modifié en cas d'adjectif épithète, et le sujet en cas d'adjectif attribut du sujet. En revanche, en cas d'adjectif attribut de l'objet, l'objet a été considéré comme non local. Enfin, pour un déclencheur nominal, les dépendants du nom (syntagmes prépositionnels et adjectifs), ainsi que ceux du verbe support s'il existe, sont considérés comme locaux.

(135) a. [...] Mme Lienemann devait faire approuver par le conseil des ministres un projet de loi prévoyant notamment [...] des garanties pour les acquéreurs de logements **VENDUS** par des marchands de biens après travaux [...]
COMMERCE_SELL: Buyer, Goods, Money, Rate, Seller

Lorsqu'un rôle n'a pas de remplisseur local mais que plusieurs segments hors du domaine de localité du déclencheur pourraient le remplir, on annote le candidat-remplisseur le plus local. Il s'agira du remplisseur dont la tête présente le chemin syntaxique le plus court vers le déclencheur, en priorisant les dépendances de compléments sous-catégorisés par rapport à celles des modifieurs. La phrase (136), où *acquisition* évoque le cadre COMMERCE_BUY, présente deux possibles remplisseurs (en gras dans l'exemple) pour le rôle Buyer de ce cadre. Bien que la seconde mention soit linéairement plus éloignée du déclencheur que la première, on peut voir dans l'arbre ci-dessous qu'elle en est syntaxiquement plus proche : la première nécessite en effet de descendre dans la relative modifiant le déclencheur. La phrase est donc annotée comme en (137).

(136) L'ACQUISITION d'Adidas, dont **il** possédait déjà 20,05% du capital, permet à **ce brasseur d'affaires sans complexe** d'ajouter ce joyau des stades à son empire familial.



(137) L'**ACQUISITION** d'Adidas, dont il possédait déjà 20,05% du capital, permet à ce brasseur d'affaires sans complexe d'ajouter ce joyau des stades à son empire familial .

COMMERCE_BUY: Buyer, Goods, Money, Rate, Recipient, Seller

7.1.2.4 Saturation des remplisseurs prédicatifs

Lorsqu'un rôle peut être rempli par une éventualité (évènement ou état), sa réalisation contient un prédicat, ou tête sémantique, et les dépendants sémantiques de cette tête. Dans ces cas-là, le remplisseur peut généralement être réalisé comme une phrase, comme c'est le cas en (138), auquel cas le verbe conjugué, qui est sa tête sémantique, est aussi sa tête syntaxique. Comme on peut le voir figure 7.1, où la tête sémantique du rôle Phenomenon (encadré) apparaît en rose, le remplisseur du rôle est alors le sous-arbre dont le verbe conjugué est la tête.

(138) Nous **CONSTATONS** que les Allemands renouent avec leur grande tradition aéronautique.

FR_BECOMING_AWARE: Cognizer, Evidence, Instrument, Phenomenon, Topic

On peut voir en (139) et figure 7.2 que ce type de rôle peut aussi être rempli par un groupe nominal dont la tête, à la fois sémantique et syntaxique, est un nom prédicatif.

(139) Aucune accumulation de la substance active n'a été **OBSERVÉE** dans le plasma après l'administration de doses multiples tous les 28 jours. UNI

FR_BECOMING_AWARE: Cognizer, Evidence, Instrument, Phenomenon, Topic

Dans ces deux types de cas, le remplisseur est continu et les têtes syntaxique et sémantique concordent. On peut donc annoter le remplisseur simplement en pointant le nœud de sa tête à la fois sémantique et syntaxique. En revanche, lorsque le prédicat qui constituerait la tête sémantique

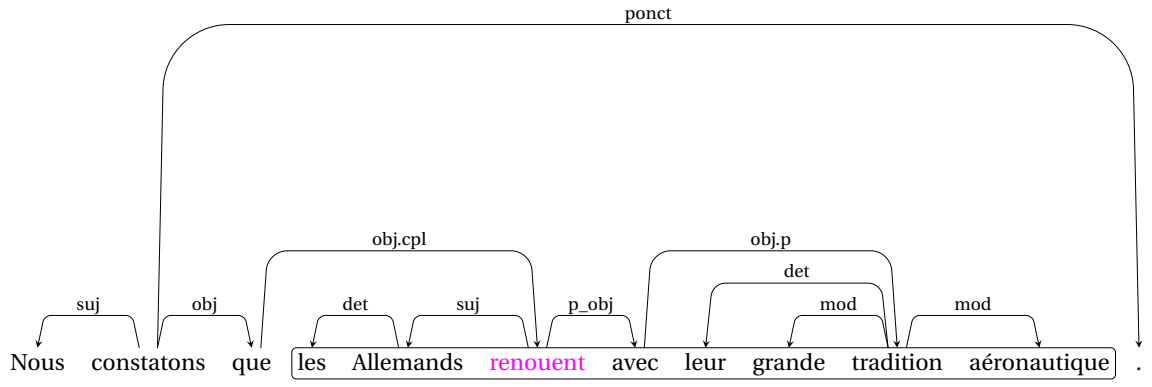


FIGURE 7.1: Le rôle est rempli par une phrase : la tête syntaxique et la tête sémantique concordent

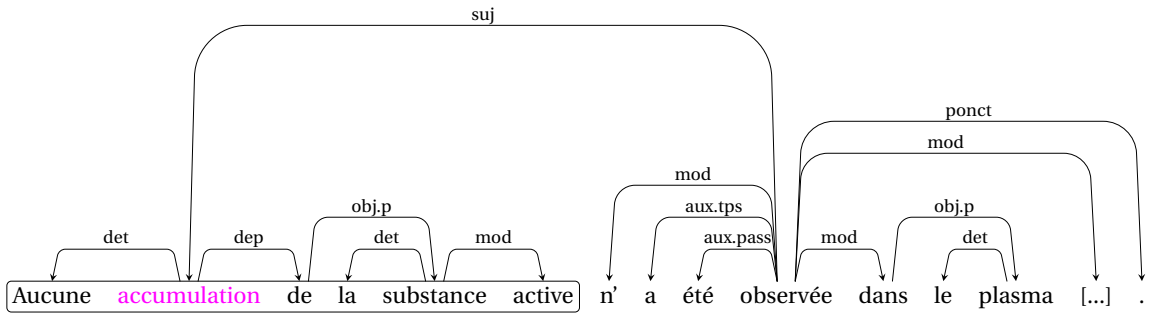


FIGURE 7.2: Le rôle est rempli par un NP prédicatif : la tête syntaxique et la tête sémantique concordent

du remplisseur est un adjectif ou un verbe non conjugué, son sujet ne peut être réalisé comme son dépendant syntaxique. Le remplisseur est alors discontinu, ou du moins n'est plus dominé par le nœud de sa tête sémantique. Nous avons choisi pour ces cas de « saturer » le remplisseur de tous ses arguments. On évite ainsi d'avoir des remplisseurs de type éventualité englobant tous leurs arguments et d'autres où l'argument sujet n'est pas inclus, ce qui permet de viser une certaine uniformité du périmètre sémantique du remplisseur. La saturation est effectuée manuellement au cours de l'annotation : son calcul via la syntaxe est en effet en théorie faisable, mais en pratique sujet à erreurs. Nous verrons section 7.2.5 que le déroulement de l'annotation prévoit un post-traitement qui calcule la tête syntaxique de tous les remplisseurs de rôles —qu'ils soient continus ou discontinus.

Saturation des remplisseurs infinitivaux

Lorsque la tête sémantique du remplisseur est un verbe non conjugué, le sujet de cette tête sémantique ne peut pas être réalisé comme son dépendant syntaxique. C'est typiquement le cas pour un remplisseur infinitival : le sujet de l'infinitif est soit non réalisé, soit réalisé non localement. Si la sémantique du rôle correspond bien à la prédication complète, alors on choisit d'inclure au remplisseur le sujet de l'infinitif.

Pour ce faire, on annote le rôle à la fois sur le sous-arbre dont le prédicat est la tête et sur son sujet, réalisé autre part dans la phrase, comme on peut le voir en (140a) et figure 7.3. Cela nous permet de rendre compte de l'alternance régulière entre (140) et l'exemple construit (140b), où *ils* et *les auteurs de la loi* sont coréférents. Notons bien que cette saturation forcée n'est effectuée que dans les cas où le périmètre sémantique du rôle inclut le sujet, et où ce rôle peut donc aussi être réalisé par une phrase complète. Cela n'est généralement pas le cas du rôle Action dans les cadres comme FR_COGNIZER_AFFECTING.TO_DO ou REMEMBERING_TO_DO, illustrés en (141a) et (141b) respectivement. Dans ces deux cadres, le Cognizer est l'agent de l'Action, et le périmètre du rôle Action est donc l'ensemble de la prédication à l'exception de son sujet.

(140) a. Les auteurs de la loi étaient bien **CONSCIENTS** de se trouver devant une tâche nouvelle et jamais encore tentée — un sentiment qu'on retrouve exprimé dans des termes voisins dans le rapport Delors.

b. Les auteurs de la loi étaient bien **CONSCIENTS** qu'ils se trouvaient devant une tâche nouvelle et jamais encore tentée [...].

FR_AWARENESS-CERTAINTY-OPINION: Cognizer, Content, Evidence, Topic

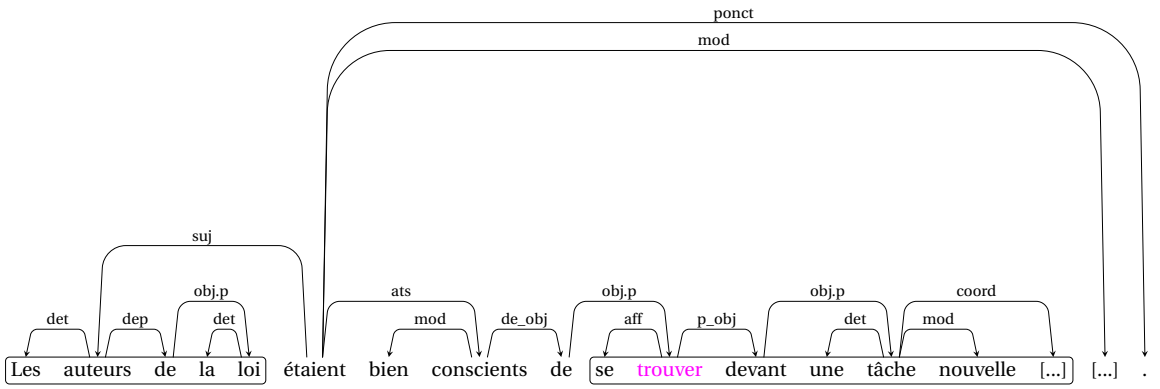


FIGURE 7.3: Le rôle est rempli par une infinitive : on sature le remplisseur à l'aide du sujet de l'infinitive

(141) a. Une découverte d'hydrocarbures, qui reste à apprécier, a **DÉCIDÉ** in extremis Repsol à changer son fusil d'épaule et à rester en Syrie.
FR_COGNIZER_AFFECTING.TO_DO: Action, Cognizer, Medium, Persuader, State_of_affairs

b. Dommage seulement que, dans la tourmente, on ait **OUBLIÉ** de filmer les silences, les visages, les pleurs ... "
REMEMBERING_TO_DO: Action, Cognizer, Salient_entity

FR_AWARENESS-CERTAINTY-OPINION: Cognizer, Content, Evidence, Topic

Dans le cas des infinitives, nous verrons en 7.2.5 que la tête syntaxique du remplisseur discontinu annoté est déterminée en post-traitement à l'annotation, à l'aide d'un ordre de priorité sur les catégories.

Saturation des remplisseurs adjectivaux

Lorsque la tête sémantique d'un remplisseur de rôle (qui sera marquée par un double-soulignement dans les exemples) est un adjectif, il y a plusieurs cas de figures. L'adjectif peut être :

— épithète (*une enfance heureuse*)
— attribut du sujet (*mon enfance fut heureuse*)
— attribut de l'objet (*elle trouve son enfance heureuse*)

Dans tous ces cas, l'argument sémantique premier de l'adjectif (qu'on appelle parfois le « sujet » de l'adjectif) n'est pas un dépendant syntaxique de l'adjectif (*enfance* dans les 3 exemples ci-dessus ne dépend pas syntaxiquement de *heureuse*). Comme pour les infinitifs, on choisit de saturer les remplisseurs, en y incluant les sujets des adjectifs.

Si l'adjectif en question est attribut du sujet (dépendant donc d'un verbe d'état) comme c'est le cas en (142) et figure 7.4, il suffit d'annoter toute la phrase du verbe d'état comme remplisseur du rôle. La divergence syntaxe-sémantique reste dans ces cas-là transparente pour les annotatrices, et une simple règle en post-traitement permet de déterminer que la tête sémantique est l'attribut et non pas la copule.

Par opposition, lorsqu'un attribut de l'objet est la tête sémantique d'un remplisseur de rôle, son premier argument sémantique —ou « sujet »— n'est pas réalisé localement à l'adjectif. Les arbres de l'adjectif et de son sujet peuvent être contigus, comme on peut le voir en (143a) et figure 7.5(a), ou discontinus, comme c'est le cas en (143b) et figure 7.5(b). Dans les deux cas, on intègre le sujet au contenu du remplisseur, et une simple règle permettra d'identifier l'attribut comme tête sémantique du remplisseur de rôle au moment des post-traitements. Dans ce type de cas, comme pour les infinitives, une tête syntaxique est déterminée en post-traitement à l'aide d'un ordre de priorité sur les catégories.

(142) Hachette était incotable à l'ouverture **EN RAISON DE** l'afflux d'ordres de ventes.
CAUSATION: Affected, Actor, Cause, Effect

(143) a. Je **TROUVE** cela dramatique.
FR_AWARENESS-CERTAINTY-OPINION: Cognizer, Content, Evidence, Topic
b. Mais les projets abondent qui vont le **RENDRE** nomade.
CAUSATION: Affected, Actor, Cause, Effect

Remplisseur adjectif ou participe épithète et divergence entre têtes syntaxique et sémantique Il en va encore différemment lorsque la tête sémantique d'un remplisseur de rôle est un épithète, fonction la plus fréquente pour les adjectifs. Le nom que l'adjectif modifie est son gouverneur syntaxique alors qu'il est aussi son argument sémantique. Si l'on tente de saturer le remplisseur en y intégrant le sujet de l'adjectif, alors le contenu du remplisseur sera l'intégralité du syntagme nominal.

Lorsqu'un syntagme nominal est remplisseur, il faut distinguer deux cas. Dans le cas classique, c'est bien l'entité dénotée par le SN qui est sémantiquement le remplisseur. C'est ce qui se passe en (144), représentée syntaxiquement figure 7.6 : la tête sémantique et la tête syntaxique du remplisseur du rôle Cause sont toutes deux le nom *guérilla*.

(144) Sur le plan procédural, l'affaire est au point mort, **GRÂCE À** une interminable guérilla de recours.
CAUSATION: Affected, Actor, Cause, Effect

Ce cas classique est à distinguer des cas où la tête sémantique est une prédication interne au SN, où l'adjectif (ou le participe) épithète prédique sur le nom. En (145), représentée figure 7.7, la cause de l'accroîssement de l'implantation de FO est le fait que sa base militante s'est améliorée, et non pas simplement l'existence de cette base militante. Le rôle Cause est donc sémantiquement *le fait que la base militante de FO s'est améliorée ou est meilleure que les autres*, et non pas *le fait que FO a une base militante, qui d'ailleurs est meilleure qu'avant ou que les autres*. C'est ce que signifie de dire que la tête sémantique du remplisseur est l'adjectif *meilleure*.

(145) FO accroît également son implantation, **GRÂCE À** une meilleure base militante dans le secteur privé.
CAUSATION: Affected, Actor, Cause, Effect

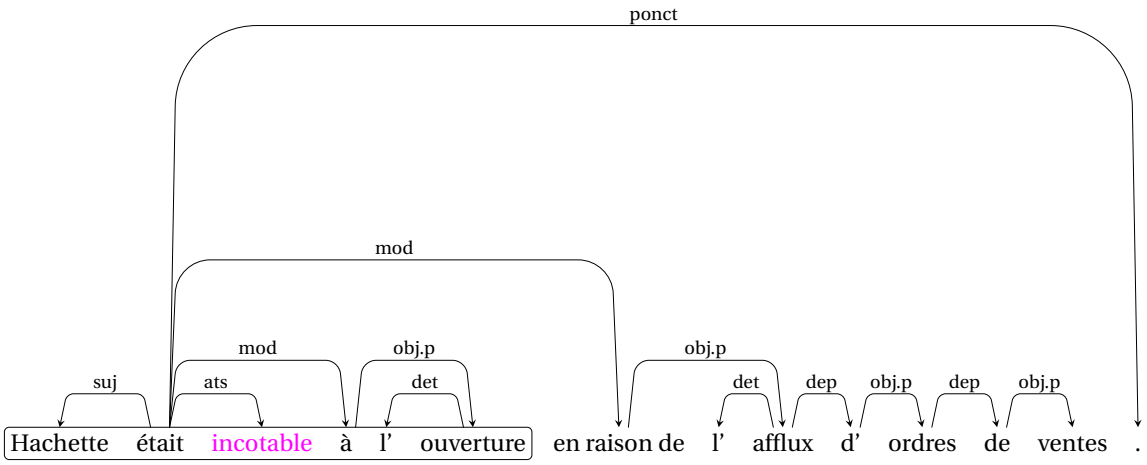
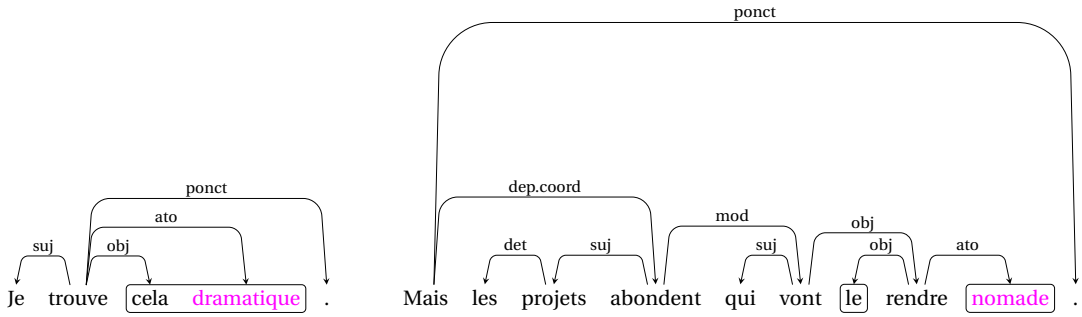


FIGURE 7.4: La tête sémantique du remplisseur est un attribut du sujet : on annote toute la phrase à copule



(a) Remplisseur continu (b) Remplisseur discontinu

FIGURE 7.5: La tête sémantique du remplisseur est un attribut de l'objet : on intègre son sujet dans le remplisseur

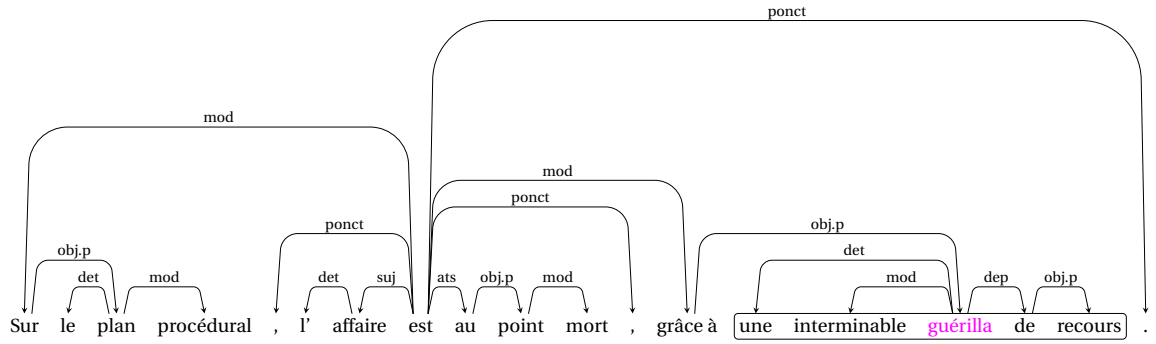


FIGURE 7.6: La tête sémantique du remplisseur est un nom : on annote tout le NP sans marquage spécifique. À contraster avec la figure 7.7

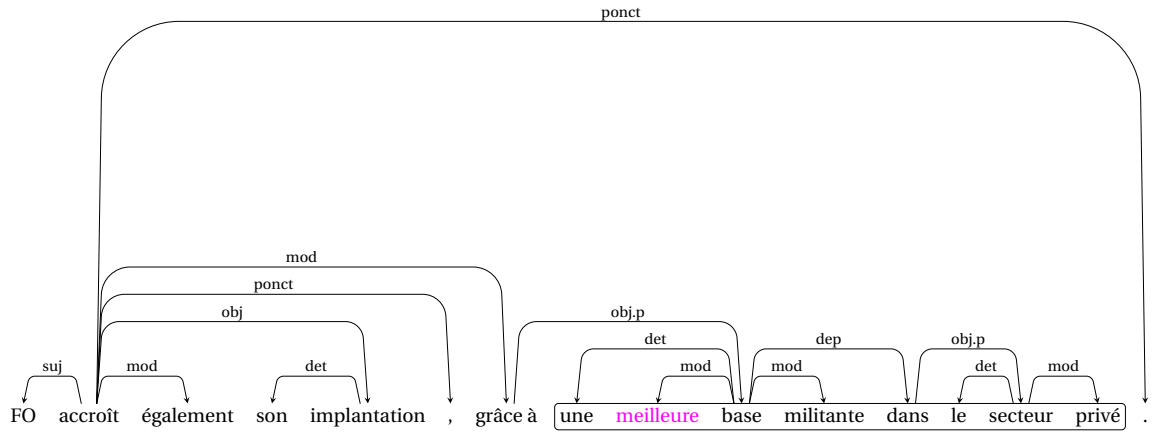


FIGURE 7.7: La tête sémantique du remplisseur est un adjectif épithète : on annote tout le NP en ajoutant un marquage spécial sur l'adjectif. À contraster avec la figure 7.6

Saturer la prédication qui est tête sémantique du remplisseur, c'est-à-dire intégrer au remplisseur le sujet de l'épithète, revient à annoter tout le SN comme remplisseur. Il est alors impossible de faire la différence entre les deux cas de figure que nous avons présentés : (144) où les têtes syntaxiques et sémantiques concordent —la tête sémantique du remplisseur est le nom qui gouverne l'adjectif épithète— et (145), où la tête syntaxique du remplisseur est toujours le nom mais sa tête sémantique est l'épithète qu'il gouverne.

Nous verrons section 7.2.5 (où nous détaillons les post-traitements intervenant à la suite de la phase d'annotation) que par défaut, la tête syntaxique d'un remplisseur est interprétée comme étant sa tête sémantique. Dans les cas de divergence syntaxe-sémantique comme celui-ci, où les têtes syntaxique et sémantique d'un remplisseur ne concordent pas et où rien ne permet de le prédire, nous avons donc pris le parti de signaler la tête sémantique du remplisseur à l'aide d'un marquage spécial (appelé *SEMHEAD*). Cela reste relativement marginal : il y a environ 130 marquages de ce type dans les données du French FrameNet 1.1.

Nous avons aussi utilisé cette technique pour différencier les participes qui sont tête sémantique de leur remplisseur de rôle (comme *écourtée* en (146a)) de ceux qui modifient simplement ce remplisseur (comme *repérés* en (146b)).

(146) a. Quatre séances seulement à la Bourse de Paris au cours d'une semaine écourtée **EN RAISON DES** fêtes du Nouvel An.

b. Des programmes de formation devant **ABOUTIR** à des emplois repérés ont ainsi été montés avec Carrefour ou la SNCF.

CAUSATION: Affected, Actor, Cause, Effect

7.1.2.5 Instanciation nulle

Il arrive qu'un rôle noyau, nécessairement conceptuellement présent, ne soit pas exprimé pour une occurrence d'un déclencheur donné. Nous avons vu en 3.3.3.1 comment étaient annotés les rôles non instanciés dans BFN. Ces derniers reçoivent un marquage spécial leur assignant un parmi trois types d'instanciation nulle :

— l'instanciation nulle définie (DNI), pour laquelle le remplisseur du rôle non réalisé doit être présent dans le contexte linguistique ou discursif ;
— l'instanciation nulle indéfinie (INI), pour les cas dans lesquels le rôle non réalisé reçoit une interprétation stéréotypique (par exemple l'objet omis de *manger* dans *Sam a mangé très vite* est interprété comme *son repas*) ;
— et l'instanciation nulle constructionnelle (CNI), lorsque le rôle manquant a été structurellement omis (par exemple le sujet canonique d'un verbe transitif employé au passif). Au sein de cette classe, aucune distinction n'est faite entre les cas recevant une interprétation anaphorique et ceux qui donnent lieu à une interprétation existentielle.

À l'instar de BFN, nous avons choisi de typer les cas d'instanciation nulle, mais n'avons pas conservé les mêmes catégories, et ce pour plusieurs raisons. Tout d'abord le fait que l'omission d'un remplisseur de rôle soit due à l'utilisation d'une construction particulière est repérable facilement par la syntaxe puisqu'il s'agit de cas réguliers, étudiés et répertoriés. Un marquage spécifique comme le CNI de BFN ne nous semble donc pas nécessaire. Par ailleurs, Ruppenhofer *et al.* (2006) rappellent que les déclencheurs autorisant une INI forment une classe fermée et ont été très étudiés, de même que l'interprétation stéréotypique reçue par l'argument omis. Il ne nous semble donc pas non plus nécessaire de leur donner un marquage spécifique.

Dans le French FrameNet, nous avons donc supprimé la classe CNI. Nous avons de plus conservé la distinction anaphorique vs. existentiel incarnée dans BFN par la différence entre instanciation nulle définie (DNI) et indéfinie (INI) sous une forme un peu différente. Les cas où un rôle reçoit une interprétation existentielle sont annotés *instanciation nulle inconnue* (UNI), marquage qui

rassemble à la fois les cas qui auraient été typés INI dans le BFN et ceux qui y auraient été annotés CNI mais dont le contenu ne saurait être interprété en contexte linguistique ni extra-linguistique. Les cas où l'argument manquant reçoit une interprétation anaphorique, annotés selon les cas comme DNI ou CNI dans le BFN, ont été divisés en deux classes dans le French FrameNet. Les cas où l'argument omis apparaît explicitement dans le contexte linguistique (les quelques phrases précédentes) sont seuls annotés *instanciation nulle définie* (DNI), tandis que ceux dont le contenu est interprétable sans apparaître dans le contexte sont typés *instanciation nulle extra-linguistique* (ENI).

Nous reprenons ci-dessous de manière un peu plus détaillée les caractéristiques de chacun de nos types d'instanciation nulle. Notons au passage que seuls les rôles noyaux non instanciés doivent être marqués d'un de ces trois types : si un rôle à annoter mais non-noyau n'est pas réalisé, il n'est simplement pas annoté. De plus, si aucun des rôles d'un ensemble nucléaire n'est réalisé, seul l'un d'entre eux doit nécessairement être signalé comme relevant d'un type d'instantiation nulle.

Instantiation nulle définie (DNI)

Ce marquage couvre les cas où un remplisseur de rôle n'est pas exprimé dans la phrase du déclencheur mais correspond clairement à un élément explicitement mentionné dans le contexte linguistique, à savoir les phrases précédentes jusqu'au début du document dans lequel se trouve l'occurrence de déclencheur. En (147b), les rôles Buyer, Goods et Seller sont annotés comme DNI puisqu'on les trouve mentionnés dans la phrase précédente présentée en (147a) —le contenu des remplisseurs des rôles y est souligné dans les couleurs correspondantes.

(147) a. [Le groupe Hachette a vendu à la société imprimeries Québecor de Montréal (Canada) l'imprimerie Graficas Monte Alban qu'il possédait au Mexique .]

b. Le prix de **VENTE**, qui n'a pas été révélé, serait inférieur à 5 millions de dollars canadiens. DNI DNI DNI

COMMERCE_SELL: Buyer, Goods, Money, Rate, Seller

Instantiation nulle extra-linguistique (ENI)

Ce marquage couvre les cas où un remplisseur de rôle n'est pas exprimé dans la phrase du déclencheur, ni mentionné explicitement dans le contexte linguistique, mais est néanmoins interprétable grâce à des connaissances extra-linguistiques ou au contexte. Le contexte linguistique présent en (148b) ne contient pas de mention précise des personnes en désaccord mais permet de comprendre qu'il s'agit de personnes impliquées dans la reprise de Perrier par Nestlé. Le rôle Cognizers est donc marqué ENI.

(148) a. Lors d'une rencontre le 30 juin à Bruxelles , les deux hommes ont constaté que les **DIVERGENCES** de vues sont encore trop grandes. ENI

b. Le commissaire européen à la concurrence, sir Leon Brittan, et le président de Nestlé, M. Helmut Maucher, sont convenus, faute d'accord sur le fond, d'un « moratoire » jusqu'au 4 août concernant la reprise de Perrier par l'entreprise suisse, a déclaré mercredi 1 juillet un porte-parole de Nestlé à Vevey.

BE_IN_AGREEMENT_ON_ASSESSMENT: Cognizer1, Cognizer2, Cognizers, Opinion, Topic

L'interprétation du remplisseur doit être relativement précise. Si elle correspond à : *tout élément qui satisfait aux restrictions de sélection du prédicat pour le rôle*, on a affaire à un cas d'UNI (voir ci-dessous) et non pas d'ENI. Ainsi dans *Sam a acheté une voiture*, le vendeur est nécessairement une personne qui possédait une voiture et cherchait à s'en défaire, mais cela ne suffit pas à rendre le rôle Seller ENI.

Instantiation nulle inconnue (UNI)

Ce marquage couvre tous les autres cas, c'est-à-dire ceux dans lesquels le remplisseur du rôle est entièrement sous-spécifié, et tout ce qui peut en être dit est qu'il existe et satisfait aux restrictions de sélection imposées par le prédicat et le contexte. Par extension, ce marquage a aussi été utilisé pour les cas où le remplisseur reçoit une interprétation universelle, comme en (149) où le remplisseur du rôle Cognizer est interprété comme *tout le monde*, ou *tout être doué de conscience*.

(149) Il est encore trop tôt pour **SAVOIR** si au bout de ces deux années, Pablo Correa intégrera le staff technique, puisqu'il paraît décidé à demeurer en France. UNI
FR_AWARENESS-CERTAINTY-OPINION: Cognizer, Content, Evidence, Topic

7.1.2.6 Astuce dans le cas de déclencheurs qui dépendent syntaxiquement d'un de leurs arguments sémantiques

Dans le cas de déclencheurs qui dépendent syntaxiquement d'un des remplisseurs de leurs rôles, on est confronté à un problème : ce remplisseur ne constitue pas un syntagme à lui tout seul. C'est le cas par exemple des adjectifs épithètes, qui prennent pour argument sémantique le nom qu'ils modifient syntaxiquement. On utilise pour ces cas-là une astuce pour accélérer l'annotation, corrigée par les post-traitements automatiques évoqués en 7.2.5.

Adjectifs et participes épithètes

Un adjectif épithète ou un participe épithète dépend syntaxiquement d'un élément (en général un nom) qui correspond à un de ses arguments sémantiques, et donc en général à un de ses remplisseurs. Ainsi en (150), *conducteurs* est argument sémantique de l'adjectif *responsables*. En (151), *mesures* est argument sémantique du participe passé *dites*; et en (152), *une facturation mensuelle* est argument sémantique du participe présent *précisant*.

(150) Un décret crée un brevet d'"animateur pour la formation des conducteurs **RESPONSABLES** d'infractions" [...]
CAUSATION: Affected, Actor, Cause, Effect

(151) Par ailleurs, mais c'est plus secondaire, le "plan" comporte un certain nombre de mesures **DITES** de traitement social, reconduites ou améliorées.
BEING_NAMED: Entity, Name

(152) L'automobiliste reçoit une facturation mensuelle **PRÉCISANT** l'heure, le sens et le jour de ses passages.
FR_STATEMENT-MANNER-NOISE: Addressee, Medium, Message, Speaker, Topic

On a typiquement la même situation pour une préposition ou une conjonction. En (153) et (154), la proposition qui précède la préposition ou la conjonction et celle qui la suit sont toutes deux remplisseurs de rôles du cadre déclenché.

(153) Ils ont été reçus à la boulangerie Leroy **POUR** visiter le fournil et surtout pétrir la pâte afin_de confectionner de délicieux pains au chocolat qu' ils ont dégustés à l'heure du goûter avec un verre de jus de fruit.
FR_MEANS_FOR_PURPOSE: Goal, Means, Time_of_eventuality

(154) Ces affaires se distinguent des autres **PARCE QU'**elles ne sont pas offertes en souscription publique.
CAUSATION: Affected, Actor, Cause, Effect

Pourtant, en syntaxe, la préposition ou conjonction modifie le verbe principal, et la principale sans ce modifieur ne forme pas un constituant : ni *ils ont été reçus à la boulangerie Leroy*, ni *ces*

affaires se distinguent des autres ne forme un constituant. En termes de dépendances, il n'y a pas de nœud qui domine le remplisseur effectif (c'est-à-dire la principale sans le modifieur introduit par la préposition ou la conjonction). On peut voir figure 7.8 qu'il n'y a pas de nœud pour *ils ont été reçus à la boulangerie Leroy* : le nœud dominant la tête *reçus* inclut le modifieur *pour visiter le fournil [...]*. De même figure 7.9, le seul nœud qui domine *ces affaires se distinguent des autres* inclut aussi *parce qu'elles ne sont pas offertes en souscription publiques*.

Pour accélérer l'annotation manuelle dans ce type de cas, on permet de coder comme remplisseur le nœud dominant le nom ou le verbe modifié par la tête sémantique du remplisseur. On peut voir figure 7.10 qu'il s'agit du nœud NC de *mesures* pour la phrase (150) et du nœud V de *distinguent* pour la phrase (154). Un post-traitement automatique corrigera cela après l'annotation : si le déclencheur est contenu dans un remplisseur sans en être la tête syntaxique, alors on interprètera le remplisseur comme étant celui annoté duquel seront retirés le déclencheur et son sous-arbre (sauf les éventuelles marques de ponctuation finales). Ce post-traitement ne s'applique pas aux remplisseurs antécédents d'anaphores que nous avons évoqués en 7.1.2.2.

Incises

On utilise également ce raccourci dans le cas de déclencheurs verbaux qui sont tête d'une incise. En (155), *croient les auteurs* est une incise, et *croient* est rattaché —comme modifieur— au verbe principal (ici le verbe *est*). Le verbe *croient* déclenche FR_AWARENESS-CERTAINTY-OPINION et le rôle Content correspond à toute la phrase sauf *croient les auteurs*.

(155) Si un partenariat avec l'Afrique est nécessaire, **CROIENT** les auteurs, c'est qu'elle est "la terre d'élection d'amitiés renforcées".
FR_AWARENESS-CERTAINTY-OPINION: Cognizer, Content, Evidence, Topic

Comme on peut le voir figure 7.11, on annote par commodité sur le nœud V dominant le verbe principal *est*. Le post-traitement s'appliquera ensuite et retirera l'incise. En effet, annoter précisément le remplisseur de rôle sans l'incise demande forcément de cliquer sur un plus grand nombre de nœuds.

Ce raccourci d'annotation est à manier avec précaution. Il faut en particulier prendre garde que l'application de l'heuristique donne un remplisseur bien formé. Prenons l'exemple de la phrase (156) : *celle-ci est loin d'être une réalité* doit être annoté comme remplisseur antécédent du remplisseur anaphorique *en* du rôle Proposition du cadre EVIDENCE. Pour faire cela, on ne peut pas utiliser le raccourci d'annoter sur le nœud V dominant le verbe principal *est*. Cela couvrirait toute la phrase, et si on enlève (automatiquement) le déclencheur et son sous-arbre, on obtient un remplisseur mal

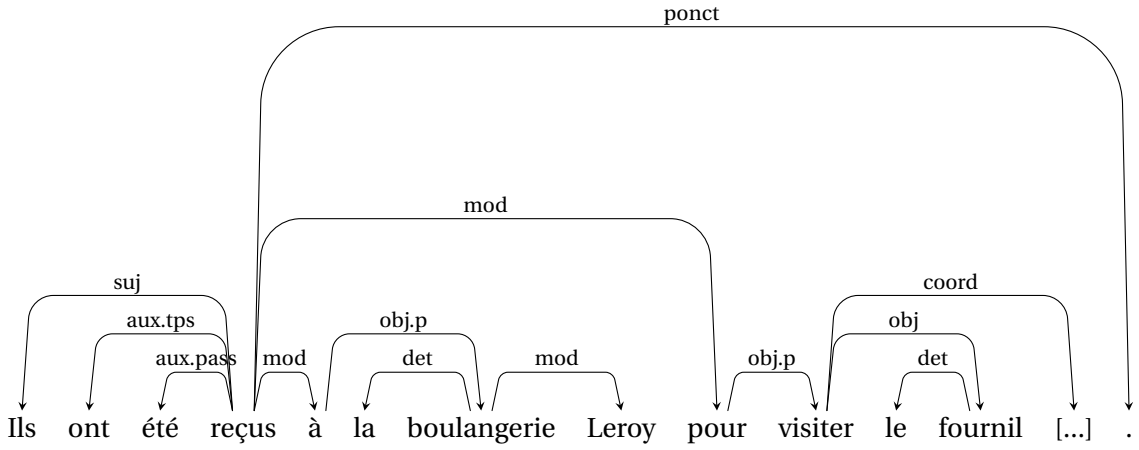


FIGURE 7.8: Représentation syntaxique de la phrase (153)

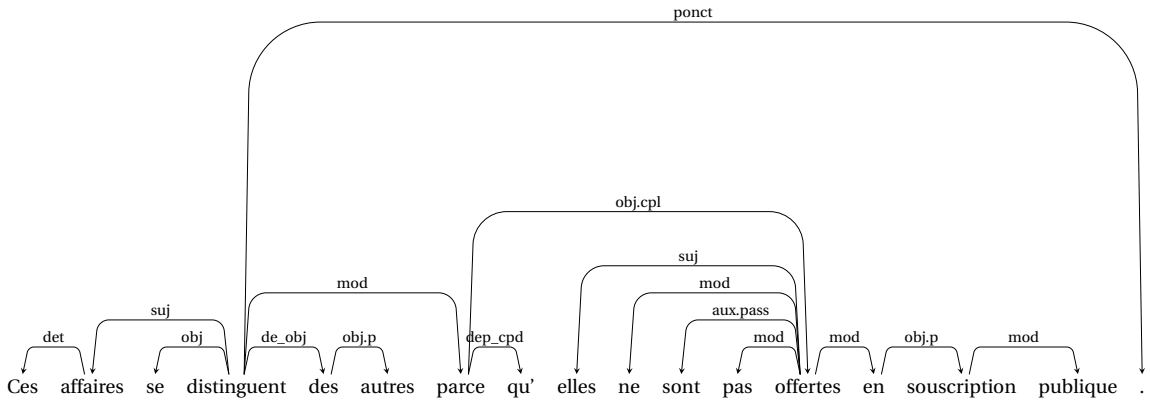


FIGURE 7.9: Représentation syntaxique de la phrase (154)

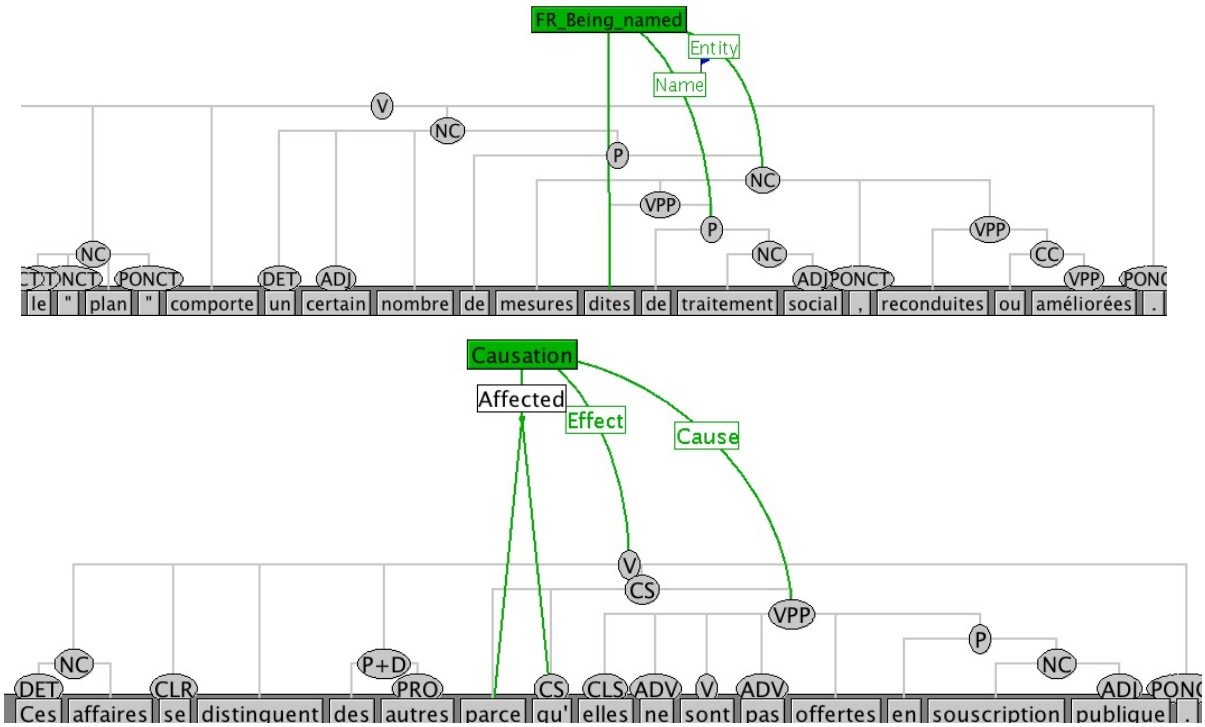


FIGURE 7.10: Astuce pour annoter les phrases (150) et (153)

formé (car incluant le *comme* : le contenu de la Proposition serait alors *celle-ci est loin d'être une réalité , comme*). Il faut alors annoter précisément le rôle Proposition comme cela est fait figure 7.12 pour obtenir exactement *celle-ci est loin d'être une réalité.*

(156) Or celle-ci est loin d'être une réalité$_{\text{ANTEC.}}$, comme en$_{\text{ANAPH.}}$ **TÉMOIGNENT** notamment les difficultés de la négociation sur le GATT.
EVIDENCE: Cognizer, Proposition, Support

7.1.3 Constructions à verbes supports

Dans une construction à verbe support, un nom prédicatif (c'est-à-dire doté d'une structure argumentale propre) est employé comme dépendant d'un verbe fréquent, dont le sens est inhibé, et dont la fonction est essentiellement d'introduire le temps et l'aspect. La tête sémantique de la

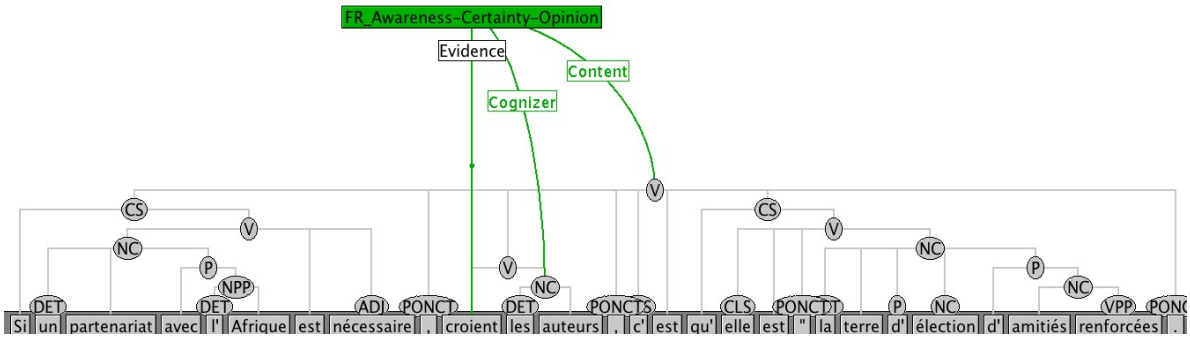


FIGURE 7.11: Astuce pour annoter la phrase (155)

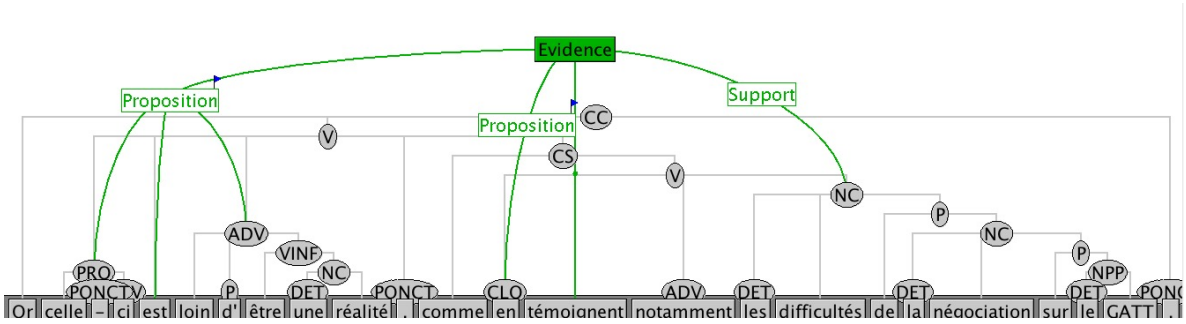


FIGURE 7.12: Cas d'annotation d'une phrase avec incise ne permettant pas d'utiliser le raccourci décrit en 7.1.2.6

construction est alors le nom prédicatif. En général le nom prédicatif est l'objet direct du verbe, comme c'est le cas pour *faire une promenade*, *commettre une agression* ou *suivre un traitement*. Danlos (1992) décrit plusieurs particularités syntaxiques des constructions à verbe support qui permettent de les identifier.

Tout d'abord, le verbe support d'un nom prédicatif permet une structure de type N_0 *Vsup (Prep) Det Npred (Prep*$_1$ N_1*) (Prep*$_2$ N_2*)* où le sujet N_0 correspond au premier argument du nom prédicatif. Ainsi, le sujet de chaque construction à verbe support en (158) correspond-il au premier argument du nom prédicatif présenté au même indice en (157).

(157) a. l'agression de Jean contre Marie
b. la promenade d'Alex
c. le traitement de Sam

(158) a. Jean a commis une agression contre Marie.
b. Alex a fait une promenade.
c. Sam suit un traitement.

Ensuite, pour un nom prédicatif et son verbe support donnés, il existe une relation de paraphrase entre les syntagmes nominaux prédicatifs de type *Det Npred de* N_0 *(Prep*$_1$ N_1*) (Prep*$_2$ N_2*)* et ceux où le nom prédicatif est modifié par une relative de structure *Det Npred (Prep) WH* N_0 *Vsup (Prep*$_1$ N_1*) (Prep*$_2$ N_2*)* (où *WH* est un pronom relatif). Les syntagmes nominaux présentés en (159) paraphrasent ceux qui apparaissent en (157).

(159) a. l'agression que Jean a commise contre Marie
b. la promenade qu'Alex a faite
c. le traitement que Sam suit

Enfin, un verbe support ne peut recevoir pour argument un syntagme nominal prédicatif de forme *Det Npred de N_0 (Prepb N_1) (Prepc N_2)*, c'est-à-dire dans lequel le nom prédicatif est déjà saturé. On peut voir en (160a) que cette construction peut donner lieu à une phrase inacceptable, mais ce n'est pas toujours le cas. Si la phrase obtenue est grammaticale, comme c'est le cas en (160b) et (160c), elle sera plutôt interprété comme mettant en relation deux occurrences du prédicat : une de ces occurrences prend pour argument le sujet du verbe support, et l'autre le premier complément du nom prédicatif. Ainsi (160b) signifie *Sam a fait la (même) promenade que la promenade qu'Alex fait d'habitude*, et (160c) a pour sens *Alex suit le (même) traitement que le traitement que suit Sam*.

(160) a. *Pierre a commis l'agression de Jean contre Marie.
b. ?Sam a fait la promenade d'Alex.
c. ?Alex suit le traitement de Sam.

Les verbes supports peuvent être entièrement inhibés sémantiquement, mais ils peuvent aussi apporter un élément de sens, comme l'aspect. Danlos identifie cinq variantes aspectuelles : l'inchoatif, le duratif, le terminatif, l'itératif et le neutre. Un même nom prédicatif aura un aspect différent selon le verbe support qui l'accompagnera, ainsi *avoir connaissance* est neutre mais *prendre connaissance* est inchoatif. Un verbe support donné n'a pas nécessairement la même contribution aspectuelle pour tous les noms prédicatifs auxquels il peut s'associer : *prendre sa revanche* est neutre alors que *prendre de l'influence* est inchoatif. Un verbe support peut aussi être opérateur causatif, et ainsi introduire un participant supplémentaire responsable de la relation prédicative. C'est le cas de *donner* dans *N_1 donne à N_0 l'espoir de X*, version causative de *N_0 a l'espoir de X*.

Dans le BFN, les verbes supports dont le complément est un déclencheur sont signalés par un marquage spécifique. Ruppenhofer *et al.* (2006) indiquent qu'à terme, ces verbes déclencheront des cadres génériques d'un nouveau type sémantique dédié mais cela ne semble à ce jour pas avoir été acté. Burchardt *et al.* (2009b) indiquent que SALSA annote les verbes supports avec un pseudo-cadre spécifique, dont le seul rôle pointe vers le nom prédicatif supporté.

Pour le French FrameNet, après quelques essais, nous avons finalement renoncé à annoter les verbes supports des noms prédicatifs déclencheurs de cadres, faute d'avoir pu écrire un guide d'annotation suffisamment opérationnel à temps. Ce travail reste à faire pour une prochaine version des données. Signalons cependant que pour les verbes déclencheurs de cadres de nos domaines, certains sens ont été annotés comme verbes support — au lieu de *Other_sense*, lorsque leur caractère de support était hors de doute (se reporter aux annotations des verbes *conclure*, *porter* et *subir*).

7.2 Déroulement du travail d'annotation

7.2.1 Corpus utilisés

L'annotation en cadres et en rôles a été effectuée sur des corpus déjà annotés en syntaxe : le French Treebank, (Abeillé *et al.*, 2003) (ci-après FTB) et le Sequoia Treebank (Candito et Seddah, 2012). Le French Treebank est composé de phrases du journal *Le Monde*, et la version SPMRL 2013 (Seddah *et al.*, 2013) que nous avons utilisée comprend 18 535 phrases avec annotations fonctionnelles. Le Sequoia Treebank utilise le même schéma d'annotation que le FTB. Il est composé de phrases de quatre sources différentes : Europarl français, le journal l'Est Républicain, Wikipédia Fr et des documents de l'Agence Européenne du Médicament. La version 7.0 [2] que nous avons utilisée compte 3 099 phrases. Nous travaillons donc sur un total de 21 634 phrases, soit 624 187 tokens. Les deux corpus sont disponibles en constituants (annotation native) et en version automatiquement convertis en dépendances (Candito *et al.*, 2010). Nous avons opté pour une annotation sur les arbres de dépendances, mais les annotations sémantiques sont facilement projetables sur les arbres

2. https://www.rocq.inria.fr/alpage-wiki/tiki-index.php?page=CorpusSequoia

en constituants. La gestion des composés, qui sont fusionnés en un seul token dans le Sequoia Treebank mais pas dans le FTB, a été uniformisée pour l'ensemble du corpus. Dans les données livrées, les composés sont « expansés », c'est-à-dire apparaissent sur plusieurs tokens.

Annoter sur des corpus annotés en syntaxe a présenté pour nous un double intérêt. D'abord lorsque les remplisseurs de rôles coïncident avec des sous-arbres syntaxiques, l'annotation en est grandement facilitée. Ensuite, l'aspect syntaxico-sémantique du French FrameNet le rend intéressant pour tout travail sur l'interface syntaxe-sémantique, ce qui nous semblait crucial.

Le cycle d'annotation comporte plusieurs étapes : d'abord, la pré-annotation en cadres des déclencheurs potentiels à annoter ; puis, l'annotation en cadres et en rôles de ces déclencheurs potentiels pré-annotés par deux annotatrices ; et enfin, l'adjudication des paires de fichiers annotés. Les fichiers adjudiqués ont ensuite subi un post-traitement. Ces différentes étapes sont décrites dans la présente section.

7.2.2 Pré-annotation automatique

Les annotatrices ont travaillé à partir de déclencheurs potentiels pré-annotés en cadres au sein de phrases du corpus. Ainsi, pour chaque occurrence de déclencheur potentiel, tous les cadres qu'il peut évoquer sont proposés, ainsi qu'un pseudo-cadre appelé *Other_sense* et signifiant que le mot cible ne déclenche aucun cadre de nos domaines. L'utilisation de ce cadre factice permet de faire la différence entre les occurrences de déclencheurs examinées au cours de la phase d'annotation dont on a jugé qu'elles n'évoquaient aucun cadre de nos domaines (annotées *Other_sense*) et les occurrences de déclencheurs qui n'ont jamais été examinées (pas d'annotation de cadre). Cette distinction est nécessaire si l'on souhaite que les données soient utilisables pour l'apprentissage automatique, puisque sans elle, la distribution des sens est faussée.

Nous avons déjà mentionné que nous n'annoterions que les cent premières occurrences de chaque lemme (ou toutes les occurrences pour les lemmes apparaissant moins de cent fois). Plutôt que de travailler lemme à lemme, nous avons décidé d'annoter en groupant les lemmes qui évoquent potentiellement exactement le même ensemble de cadres. Ainsi, *convaincre* et *persuader*, qui évoquent tous deux FR_COGNIZER_AFFECTING.TO_DO et FR_COGNIZER_AFFECTING.VERACITY ont été groupés dans un même fichier pré-annoté, et annotés simultanément. Travailler sur des ensembles de lemmes ayant la même ambiguïté de cadres permet aux annotatrices de ne se former qu'une seule fois à une frontière de cadres donnée, et de traiter l'ambiguïté de manière cohérente pour tous les lemmes concernés.

En plus de ce classement par ensembles de cadres évoquables, les déclencheurs sont prioritairement regroupés selon plusieurs heuristiques. Les verbes pronominaux sont associés aux verbes non pronominaux correspondants (lorsque les deux verbes apparaissent dans le lexique). Ainsi, en cas d'erreur d'annotation (si le fait qu'un déclencheur est pronominal n'a pas été capturé, ou a été capturé à tort —cf. 7.2.2.1), il restera aisé de fournir une annotation cohérente. De même, les mots composés sont groupés avec les mots qui les composent : *croire*, qui peut évoquer FR_AWARENESS-CERTAINTY-OPINION, RELIGIOUS_BELIEF ou FR_BEING_IN_FAVOR_OF a été annoté au sein du même fichier que *donner à croire*, qui évoque EVIDENCE ou FR_ATTEMPT_SUASION.VERACITY. Cela permet d'annoter toutes les occurrences d'un même composé dans le même fichier, et ce indépendamment de potentielles incohérences d'annotation de composés dans le corpus (cf. 7.2.2.2). Notons que ces deux heuristiques amènent à faire cohabiter dans un même fichier pré-annoté des déclencheurs potentiels ne présentant pas la même ambiguïté de cadres. La troisième heuristique consiste à associer les verbes et leurs nominalisations lorsque les deux sont présents dans le lexique (comme *vendre* et *vente* qui évoquent tous deux le cadre COMMERCE_SELL). Annoter ensemble ces lemmes sémantiquement proches permet de gagner encore en cohérence.

Ces groupements automatiques de lemmes permettent de produire un ou plusieurs fichiers pré-annotés, calibrés à cent phrases à annoter. Nous avons déjà dit n'annoter que les cent premières

occurrences de chaque lemme ; le contenu de l'ensemble des fichiers devrait donc être limité à ce nombre d'occurrences pour un même déclencheur. Dans les faits, nous avons coupé à cent occurrences uniquement dans le cas de lemmes apparaissant plus de 110 fois, afin de ne pas laisser quelques occurrences non annotées. Le coupage à cent n'a été donc réalisé que s'il y avait beaucoup plus de cent occurrences. En outre, parmi les occurrences soumises à l'annotation, il pouvait y avoir des faux positifs (en particulier liés aux mots composés) ; de ce fait, même pour des lemmes coupés aux cent premières occurrences, il arrive qu'on obtienne un peu moins de cent annotations au final. Au total, sur le lexique actuel, on obtient 82 lemmes avec plus de 100 annotations, et 27 avec entre 90 et 99 annotations.

7.2.2.1 Pré-annotation des verbes pronominaux

Le clitique réfléchi *se* peut correspondre à différents processus syntaxiques ou lexicaux, ce qui fait l'objet d'une abondante littérature car les distinctions de statut ne sont pas toujours évidentes. On distingue grosso modo les cas suivants :

— le vrai réfléchi, comme en (161a), où *se* réfère au sujet du verbe et a le statut d'actant.
— le vrai réciproque, comme en (161b), où *se* réfère au sujet collectif dont les membres exercent l'un sur l'autre l'action exprimée par le verbe.
— le se-moyen, comme en (161c), dont la caractéristique est que l'agent n'est pas réalisable localement mais est forcément interprété comme existant. Dans l'exemple, on ne peut pas interpréter la phrase sans faire intervenir un agent qui vend ce genre de livre.
— le se-neutre, comme en (161d), où l'agent est « effacé », pas forcément interprété comme existant.
— le verbe dit intrinsèquement pronominal, pour lequel une version non pronominale n'existe pas (par exemple *se suicider*), ou bien pour lequel on ne repère pas en diachronie de relation régulière à une version non pronominale (par exemple s'apercevoir).
— le verbe dit coopératif, comme en (161e), où la version *Xcollectif se V* alterne avec *Y se V à/avec Z*.

(161) a. Alex se regarde.
b. Sam et Alex se parlent.
c. Ce genre de livre, ça se vend dans les gares.
d. Tout à coup, le vase s'est fendu.
e. Sam et Alex se marient.

Cette classification est importante pour l'annotation, car selon les cas, on considère que le déclencheur est le verbe avec ou sans le clitique *se*.

Dans le cas des vrais réfléchis, vrais réciproques et se-moyen, on considère que le *se* marque un changement de diathèse régulier, sans modification du nombre des arguments sémantiques. Le déclencheur à annoter est donc le verbe sans le *se*. Ainsi, (162a) et (162b) sont annotées avec le même cadre.

(162) a. Entre 15 et 24 ans, plus d'une jeune femme sur quatre se **DÉCLARE** sans emploi.
b. « Il faut que l' argent circule », **DÉCLARE** M. Mitterrand lui - même.
FR_STATEMENT-MANNER-NOISE: Addressee, Medium, Message, Speaker, Topic

En revanche, pour les se-neutres et verbes intrinsèquement pronominaux, le déclencheur à annoter inclut le clitique *se*. Les cadres possibles pour le *se+V* ne sont généralement pas les mêmes que les cadres du *V*. Ainsi en (163a), *se rappeler* évoque le cadre FR_MEMORY-REMEMBERING_EXPERIENCE-REMEMBERING_INFORMATION, tandis qu'en (163b), *rappeler* déclenche le cadre EVOKING.

(163) a. M. Wurtz devrait peut-être **SE RAPPELER** qu'il n'y a pas si longtemps, des collègues de son parti, des maires communistes français, envoyaient des bulldozers contre des foyers d' immigrés en France.
FR_MEMORY-REMEMBERING_EXPERIENCE-REMEMBERING_INFORMATION: Cognizer, Mental_content, Salient_entity, Topic

b. Une telle pratique, à la longue, risque fort de **RAPPELER** aux roumains les mauvais souvenirs d'un passé encore tout frais ...
EVOKING: Cognizer, Phenomenon, Stimulus

La distinction entre ces catégories n'étant pas toujours aisée, nous avons utilisé des annotations manuelles produites antérieurement à notre campagne d'annotation. L'annotation des *se+V* du corpus Sequoia a été effectuée dans le cadre du projet Deep-sequoia (Candito *et al.*, 2014b) ; l'annotation des *se+V* du French Treebank a été réalisée ensuite par Marie Candito et Virginie Mouilleron, et adjudiquée par Marie Candito, Benoît Sagot, Virginie Mouilleron et moi-même. Le guide d'annotation développé et utilisé lors de ces deux vagues d'annotation est disponible à l'adresse suivante : `http://www.linguist.univ-paris-diderot.fr/~mcandito/Publications/Guide-annotation-se.pdf`.

7.2.2.2 Pré-annotation des déclencheurs composés

Dans la couche morphologique de notre corpus, l'annotation des composés n'est pas toujours cohérente. Nous avons mentionné que nous travaillions sur la concaténation de deux corpus annotés en morpho-syntaxe : le French Treebank (FTB) et le Sequoia Treebank. Il peut arriver qu'une expression de même sens apparaisse comme composé dans un corpus et expression compositionnelle dans l'autre ; les critères pour déterminer le caractère composé ou non d'une expression ne sont en effet pas les mêmes pour les deux corpus. Pour le Sequoia Treebank, seuls des composés grammaticaux ont été annotés, ce qui fait en particulier qu'aucun composé nominal n'est annoté comme tel. Le FTB inclut donc plus de composés que le Sequoia Treebank. *Homme d'affaires* est par exemple systématiquement annoté comme composé dans le FTB alors qu'il ne l'est jamais dans le Sequoia Treebank. De plus, même au sein d'un même corpus, il arrive de rencontrer des incohérences. Ainsi on peut voir en (164) que bien qu'*au moment de* ait parfois été annoté comme composé, cela n'a pas toujours été le cas.

(164) a. La coopérative n'a repris que cent cinquante salariés sur les deux cent trente que comptait l'entreprise *au_moment_de* la faillite.

b. Autant de qualités récompensées par un prix Pulitzer et par l'accueil du public, en France comme aux Etats-Unis, un an plus tôt, *au moment de* la guerre du Golfe.

Par ailleurs, nous avons dit en 7.1.1.1 que nous annoterions les occurrences de déclencheurs même au sein des entités nommées lorsque le sens de ces dernières est compositionnel. Lorsqu'elles sont annotées en tant que composés, nous souhaitons tout de même annoter les cadres pertinents sur les déclencheurs qui y apparaissent. Ainsi en (165), bien que *chambre de commerce et d'industrie* soit annoté comme composé, on considère bien que *commerce* évoque le cadre COMMERCE_SCENARIO.

(165) Pour le président de la chambre_de_**COMMERCE**_et_d'industrie, Andréas Kyriazis, « le monde industriel et commercial subit encore le poids du passé ».
UNIUNIUNI
COMMERCE_SCENARIO: Buyer, Goods, Money, Participant_1, Participant_2, Participants, Rate, Seller

Ces deux raisons font que nous devons être en mesure d'annoter un cadre sur un déclencheur composant d'un composé, et que les cadres pertinents doivent être pré-annotés sur ces déclencheurs

au sein des composés. La pré-annotation des mots composés est donc conçue pour être tolérante aux incohérences d'annotation de composés dans les treebanks sous-jacents. Ainsi, on pré-annote systématiquement *en effet* avec le cadre EVIDENCE, que la séquence *en+effet* soit ou pas annotée comme un composé dans le treebank. Des heuristiques ont aussi été utilisées pour repérer et pré-annoter les composés du lexique même lorsqu'ils apparaissaient de manière discontinue, comme en (166).

(166) On évita de justesse l'effondrement complet du système bancaire **GRÂCE** notamment **AUX** prêts et aux expéditions d'or en provenance d'Europe.

Cela occasionne bien sûr du bruit, comme on peut le voir en (167) où *grâce à* sera proposé comme déclencheur composé alors qu'il ne l'est pas. Dans ce cas de « faux composé », aucun cadre n'est finalement annoté, pas même *Other_sense*, car il s'agit de bruit dû à la pré-annotation. Ce type de bruit, aisément repérable par les annotatrices, nous semble moins gênant que de manquer de potentielles occurrences du déclencheur à cause d'erreurs d'annotation dans les treebanks.

(167) Ni la plus ou moins apparente « francisation » du montage financier ni les assurances du gouvernement sur l'emploi ne trouvent grâce à leurs yeux.

Lorsqu'on a dans le lexique ASFALDA à la fois un déclencheur composé et un déclencheur qui fait partie des composants de ce composé, on regroupe ces deux déclencheurs dans un même fichier de pré-annotation et on pré-annote à la fois composé et composant à l'aide des cadres pertinents. Ainsi *penser*, qui peut évoquer FR_AWARENESS-CERTAINTY-OPINION ou REMEMBERING_TO_DO a été annoté au sein du même fichier que *donner à penser*, qui évoque EVIDENCE ou FR_ATTEMPT_-SUASION.VERACITY. Dans les phrases où *donner à penser* apparaît, comme c'est le cas de la phrase figure 7.13, tous les cadres mentionnés ont été proposés pour les déclencheurs respectifs de ces cadres.

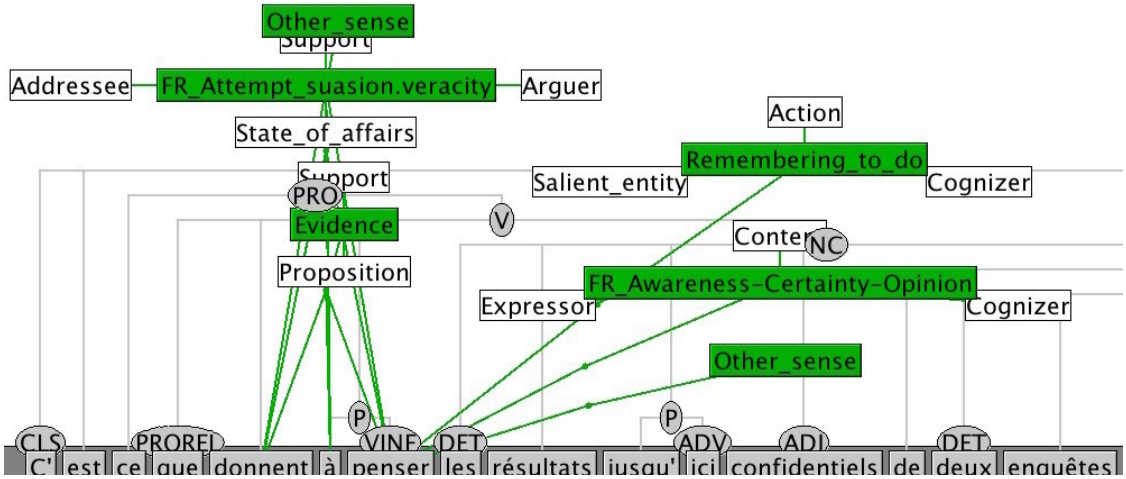


FIGURE 7.13: Pré-annotation d'une phrase où apparait un déclencheur potentiel composé (*donner à penser*) dont un des composants (*penser*) apparaît aussi dans le lexique du French FrameNet

7.2.3 Annotation

Nous venons de voir que les fichiers pré-annotés sur lesquels vont travailler les annotatrices présentaient les occurrences de déclencheurs potentiels au sein de phrases en isolation. Pour l'étape d'annotation décrite dans cette section, ainsi que pour l'adjudication, les annotatrices ont accès à

l'ensemble du corpus délimité en documents [3], et donc au contexte linguistique des phrases qu'elles annotent. Cela est nécessaire pour le choix d'un type d'instanciation nulle (cf. section 7.1.2.5), lorsqu'un rôle noyau n'est pas réalisé.

L'annotation s'est faite par le biais de l'outil graphique SALTO, développé par Burchardt *et al.* (2006a) dans le cadre du projet SALSA de FrameNet allemand (Burchardt *et al.*, 2006b). Le travail d'annotation des occurrences de déclencheurs de chaque fichier a été effectué séparément par deux annotatrices. Pour chaque occurrence de déclencheur potentiel, ces dernières ont utilisé l'interface SALTO pour sélectionner parmi les cadres proposés par la pré-annotation celui qu'évoquait le déclencheur et supprimer les autres. C'est l'étape que l'on peut voir figure 7.14. Si le mot-cible n'évoque aucun des cadres de nos domaines, l'annotatrice ne conserve que le pseudo-cadre appelé *Other_sense*, qui permet de distinguer ces cas de ceux où l'occurrence n'a tout simplement pas été examinée, et passe ensuite à l'occurrence suivante.

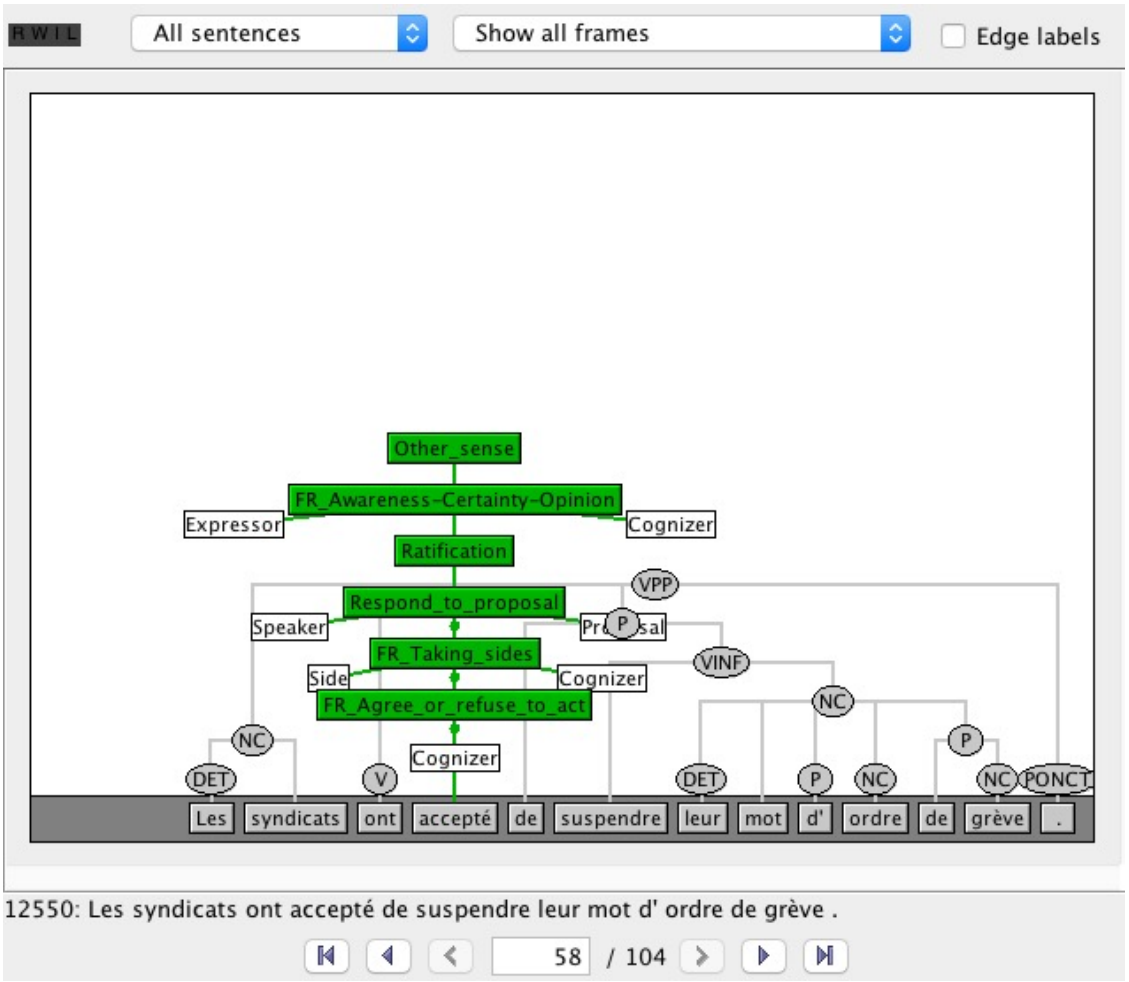


FIGURE 7.14: Sélection du cadre déclenché

Si un cadre de nos domaines a été sélectionné, l'annotatrice annote ensuite le remplisseur de chaque rôle à annoter. Les rôles noyaux du cadre sont les seuls accessibles immédiatement sous SALTO. Les rôles non-noyaux à annoter sont accessibles à partir d'un menu, à la différence des rôles noyaux du cadre qui apparaissent de manière visible autour du cadre (cf. figure 7.15). Cela permet à

3. Pour le Sequoia Treebank, nous avons reconstitué un découpage en documents pour les sous-parties provenant de l'Est Républicain, de la Wikipedia Fr et du corpus EMEA, mais pas pour la partie Europarl, pour laquelle le découpage en tours de parole est perdu.

l'annotatrice de ne pas oublier de leur assigner un type d'instanciation nulle s'ils ne sont pas réalisés et que les rôles en ensemble nucléaire avec eux ne le sont pas non plus.

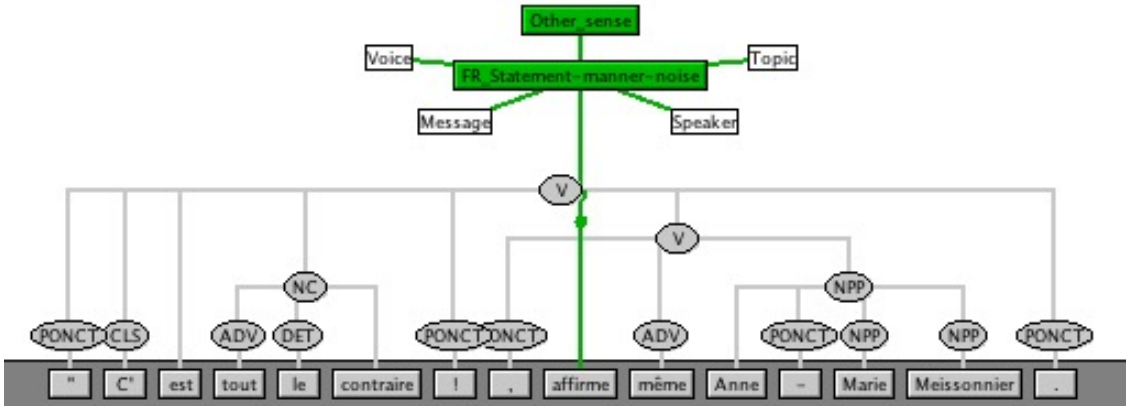


FIGURE 7.15: Rôles noyaux disponibles pour l'annotation

L'outil SALTO est initialement conçu pour visualiser des arbres en constituants. Nous avons converti les arbres de dépendances au format FTBDep de notre corpus vers un format à la Hays-Gaifman (Hays, 1964; Gaifman, 1965), où les nœuds terminaux sont les mots et les nœuds internes correspondent à leurs catégories. Les cadres sont pré-annotés, puis annotés, au niveau des nœuds lexicaux de leurs déclencheurs. Les remplisseurs de rôles, quant à eux, sont annotés de la manière qui paraît la plus pratique ou rapide à l'annotatrice tant qu'elle permet de couvrir exactement l'empan du contenu du remplisseur (c'est-à-dire des mots qui le constituent). Ainsi si le remplisseur d'un rôle est constitué de l'ensemble des mots dominés par une tête, le rôle pourra pointer vers le noeud dominant — ce qui accélère l'annotation. C'est ce que l'on peut observer figure 7.16.

Nous avons pris le parti de ne pas corriger l'annotation syntaxique. Nous l'utilisons pour accélérer l'annotation mais si l'information syntaxique est fausse, il est toujours possible d'annoter au niveau des tokens. De même, le repérage de la non-localité d'un remplisseur de rôle doit s'appuyer sur la « syntaxe correcte », et ne pas tenir compte de potentielles erreurs d'annotation.

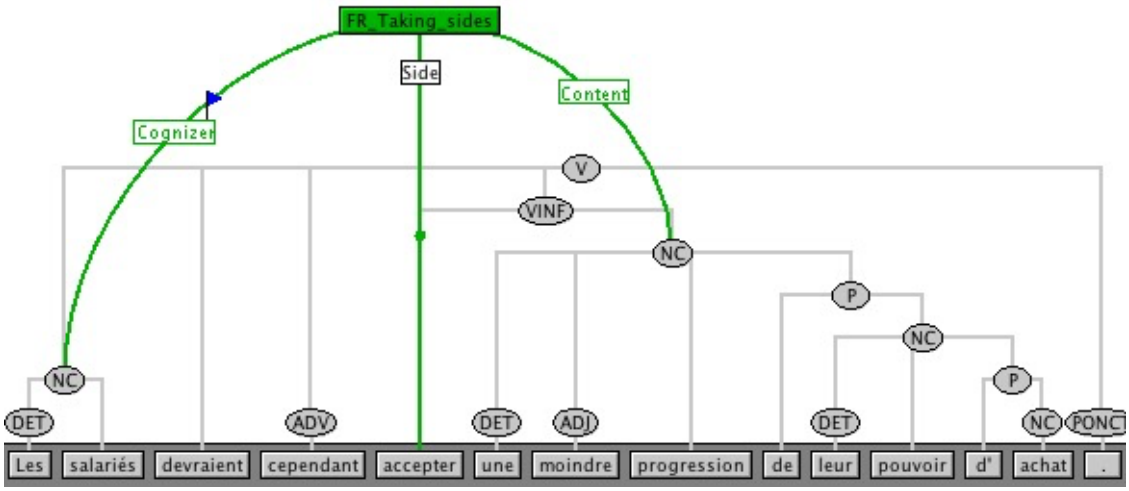


FIGURE 7.16: Annotation des remplisseurs des rôles

Il arrive que l'on ait besoin d'annoter un rôle plusieurs fois, par exemple comme nous l'avons vu section 7.1.2.2 lorsqu'il est rempli par un élément anaphorique local au déclencheur et dont

l'antécédent se trouve aussi dans la phrase. Dans ces cas-là, deux occurrences distinctes du rôle sont annotées, avec chacune un marqueur, ou *flag*, indiquant s'il s'agit de l'anaphore ou de son antécédent (cf. figure 7.17). Des marqueurs sont aussi utilisés dans d'autres cas. Ils permettent de signaler la non-réalisation d'un rôle essentiel en précisant son type d'instanciation nulle, ou encore d'indiquer qu'un rôle est réalisé non-localement au déclencheur. Ce sont aussi ces marqueurs qui sont utilisés — cette fois au niveau des cadres et non plus des rôles — pour signifier l'emploi métaphorique d'un cadre.

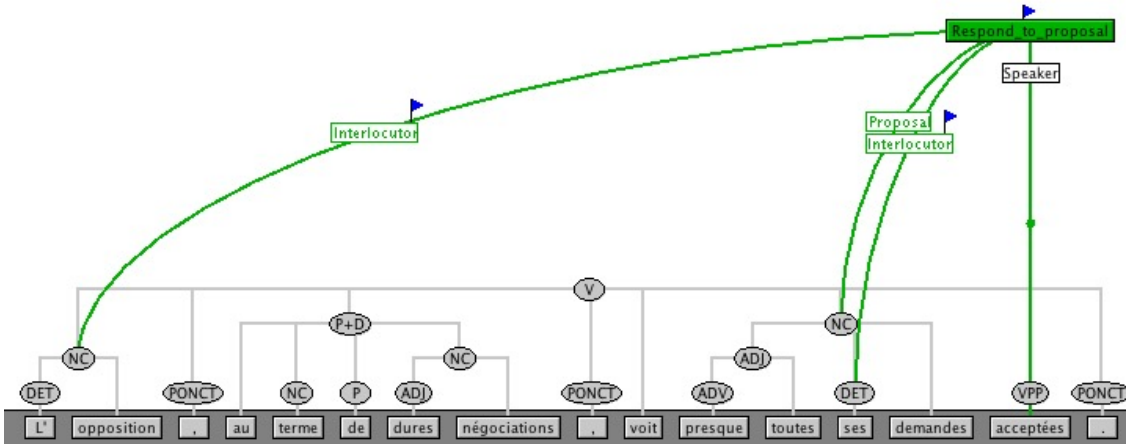


FIGURE 7.17: Annotation d'un rôle rempli par un élément anaphorique

7.2.4 Adjudication

Une fois que chaque annotatrice a terminé un fichier, ce fichier est adjudiqué, toujours sous SALTO, qui repère automatiquement les conflits entre les deux fichiers annotés. Cette adjudication a au départ été effectuée en groupe par les deux personnes ayant annoté le fichier accompagnées d'une ou un des responsables du domaine concerné. Cela a permis de continuer à former les annotatrices, ainsi que de repérer d'éventuels problèmes nécessitant l'ajout de précisions au guide d'annotation. Plus tard, les annotatrices ont pu adjudiquer les fichiers elles-mêmes, en consultant les responsables en cas de doute où de nombre de conflits trop important. Il est aussi arrivé que des fichiers soient adjudiqués uniquement par un responsable du domaine. L'étape d'adjudication est représentée figure 7.18, où les choix communs aux deux annotatrices apparaissent en vert tandis que les décisions prises par une seule des deux apparaissent en bleu pour l'une et jaune pour l'autre.

7.2.5 Post-traitement

Après l'annotation interviennent deux types de post-traitement automatique. Le premier consiste à calculer le contenu correct des remplisseurs. Cette étape est nécessaire après les astuces d'annotation que nous avons dit utiliser. Il s'agit en fait d'extraire le déclencheur et son sous-arbre d'un remplisseur dont la tête domine ce déclencheur. Pour plus de détails, se référer à la section 7.1.2.6.

L'autre type de post-traitement consiste à calculer les têtes sémantique et syntaxique de chaque remplisseur de rôle. Pour ce qui est de la tête syntaxique, cette étape s'appuie sur l'annotation syntaxique sous-jacente, et sur une heuristique dans certains cas. Dans le cas général, on a des remplisseurs de rôles continus et formés par un seul arbre ; la racine de cette arbre est alors la tête syntaxique du remplisseur. Il existe cependant plusieurs situations où cette tête doit être calculée : il s'agit des cas évoqués section 7.1.2.4.

Lorsqu'un remplisseur de rôle est discontinu ou lorsqu'il est continu mais formé de plusieurs sous-arbres syntaxiques non connectés, on considère les nœuds racines de ces sous-arbres, la

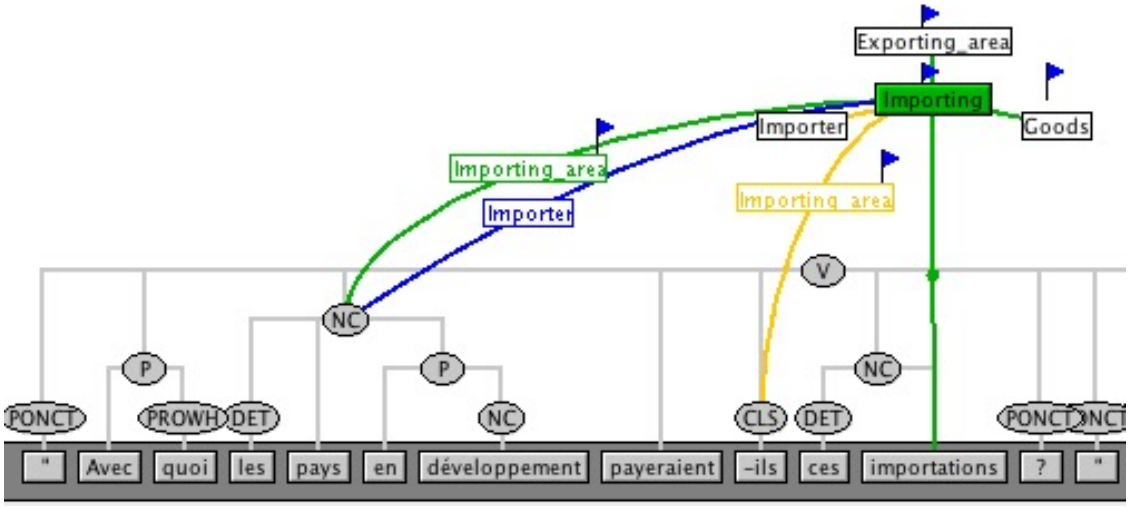


FIGURE 7.18: Adjudication du travail de deux annotatrices

fonction grammaticale qu'ils portent, et leur catégorie. On choisit en priorité l'attribut le plus à gauche s'il existe, et à défaut l'adjectif le plus à gauche, etc. L'ordre de priorité est attribut, adjectif, verbe, nom, adverbe, pronom.

Ainsi, en (168), où le verbe *affirmer* déclenche le cadre FR_STATEMENT-MANNER-NOISE, on considère que le rôle Message est rempli par *les sœurs + avoir été ensorcelées par le curé Urbain Grandier*[4]. Comme on peut le voir figure 7.19, le remplisseur est formé de deux sous-arbres syntaxiques déconnectés, de tête *nom* et *verbe*. D'après notre règle de priorité, c'est le verbe *ensorceler* qui sera choisi comme tête syntaxique de l'ensemble du remplisseur.

(168) En 1634, dans un couvent d'Ursulines à Loudun, les sœurs **AFFIRMÈRENT** avoir été ensorcelées par le curé Urbain Grandier.

De même en (169) où le verbe *trouver* déclenche le cadre FR_AWARENESS-CERTAINTY-OPINION : le rôle Content est rempli par *scandaleux + il n'ait rien fait pour attirer l'attention des dirigeants de la chaîne et s'associe purement et simplement à la mise de la clé sous la porte*. Comme on peut le voir figure 7.20, d'après le schéma d'annotation syntaxique du FTB en dépendances, les têtes des

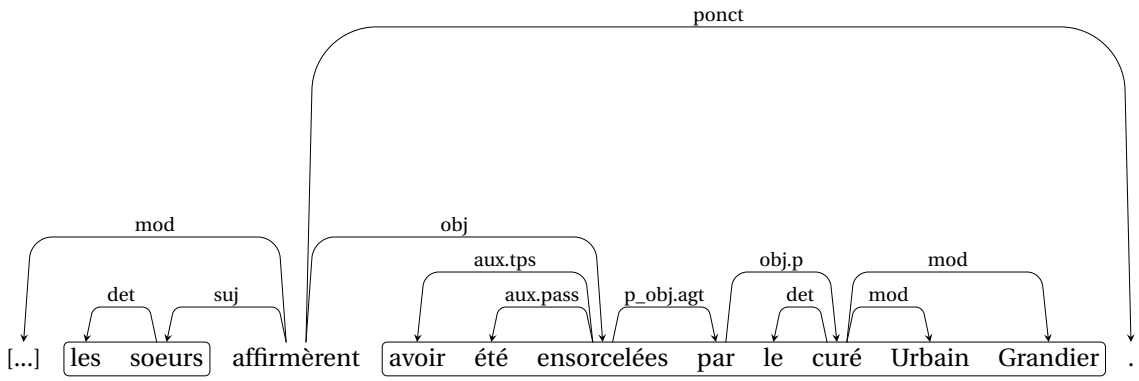


FIGURE 7.19: Représentation syntaxique de la phrase (168)

4. Nous avons expliqué section 7.1.2.4 saturer les remplisseurs infinitivaux en intégrant leur sujet au remplisseur du rôle.

sous-arbres syntaxiques sont *scandaleux* et *fait*. Comme *scandaleux* porte la fonction *attribut de l'objet*, c'est lui qui est choisi comme tête.

(169) Je **TROUVE** scandaleux qu'il n'ait rien fait pour attirer l'attention des dirigeants de la chaîne et s'associe purement et simplement à la mise de la clé sous la porte.

Nous avons mentionné que par défaut, la tête syntaxique serait interprétable dans les données comme étant aussi la tête sémantique. En nous intéressant au cas de la saturation des remplisseurs adjectivaux (section 7.1.2.4), nous avons cependant repéré des cas de divergence.

Nous avons dit saturer les adjectifs ou participes attributs remplissant un rôle en intégrant leur sujet au contenu dudit rôle. Dans le cas des attributs du sujet, la proposition dont le verbe d'état est la tête syntaxique est annotée tout entière comme remplisseur du rôle. Le post-traitement consiste pour ces attributs en l'application d'une règle spécifiant que c'est l'attribut et non pas la copule qui est tête sémantique du remplisseur.

Dans les autres cas où la tête sémantique ne correspond pas à la tête syntaxique, comme le cas spécifique des épithètes décrit en 7.1.2.4 où aucune règle formelle ne permet de déterminer si le nom ou l'adjectif est la tête sémantique du remplisseur, nous avons proposé une annotation manuelle de cette tête sémantique. On obtient après post-traitement que la tête sémantique d'un remplisseur est celle marquée si une telle marque existe, la tête syntaxique sinon.

7.2.6 Documents et outils utilisés tout au long du cycle d'annotation

En sus du guide d'annotation général (cf. annexe B), les annotatrices ont utilisé plusieurs documents et outils. Avant d'entamer le travail sur un nouveau domaine, elles ont pris connaissance de ses différents cadres, en prêtant une attention particulière aux caractéristiques distinctives entre cadres proches que nous avons ajoutées à la définition d'une grande partie de nos cadres (cf. section 5.2.1.2). Cela a permis à la fois une meilleure appréhension générale de la « cartographie » des cadres du domaine et la compréhension précise des spécificités de chaque cadre avant l'annotation de ses déclencheurs.

Lorsqu'un déclencheur s'avérait particulièrement difficile à désambiguïser, les annotatrices ont pu demander aux responsables des domaines pertinents de leur fournir des « guides de polysémie ». Il s'agit de guides informels de désambiguïsation pour un lexème ou groupe de lexèmes précis, avec parfois des règles par défaut puisque nous avions pris le parti de ne pas laisser d'ambiguïté dans les données.

Enfin, tout au long du travail d'annotation et d'adjudication des fichiers, les annotatrices ont pu poser toute nouvelle question dans un espace prévu à cet effet pour chaque domaine. C'est cet espace qui a permis aux différents supports présentés ci-dessus de s'étoffer et d'évoluer tout au long du projet.

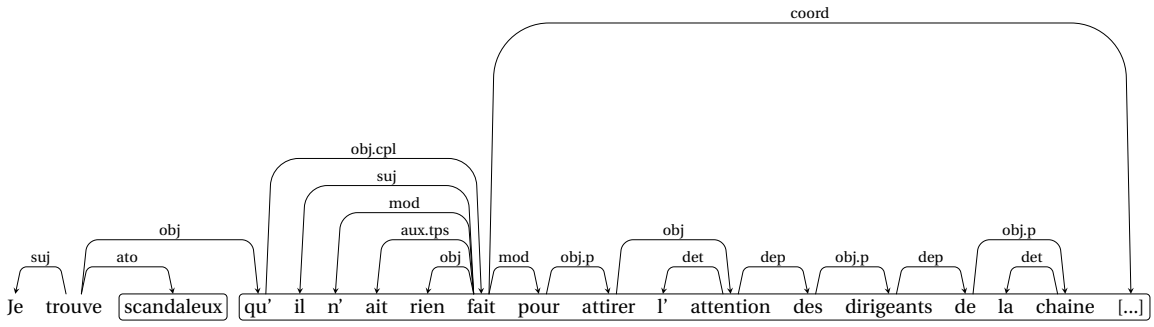


FIGURE 7.20: Représentation syntaxique de la phrase (169)

7.3 Points spécifiques

Les sections qui suivent se penchent chacune sur un point nous ayant demandé une réflexion particulière avant de décider d'une ligne de conduite. Nous nous intéressons d'abord aux constructions à attribut de l'objet, qui fournissent un cas intéressant de divergence syntaxe-sémantique. Nous nous penchons ensuite sur le cas des déclencheurs nominaux, avec en particulier des cas où l'on considèrera un nom à la fois comme déclencheur d'un cadre et comme remplisseur d'un rôle de ce frame.

7.3.1 Constructions à attribut de l'objet

Nous l'avons vu, bien que FrameNet se veuille principalement orienté sémantique, la syntaxe est souvent utilisée, par exemple pour délimiter les cadres et définir le statut nucléaire ou non des rôles des cadres. Il nous semble donc intéressant de nous pencher sur les traitements fournis ou imaginables lorsque syntaxe et sémantique divergent.

Les constructions à attribut de l'objet (ci-après CAO) sont un cas typique de divergence syntaxe-sémantique. Il s'agit de constructions de type N_0 V N_1 X où N_0 est le sujet de V, et N_1 l'objet direct de V et le sujet du prédicat tête de X. Les V de ces constructions présentent la particularité de sous-catégoriser trois arguments syntaxiques (N_0, N_1 et X) et seulement deux arguments sémantiques (N_0 et X). X est quant à lui un prédicat se rapportant à N_1.

Dans la section présente, nous reprenons Djemaa (2014) pour tenter de proposer pour le French FrameNet un traitement linguistiquement motivé de ces constructions. Pour ce faire, tout en répertoriant rapidement quelques critères de classement des constructions à attribut de l'objet, nous précisons les cas sur lesquels nous nous penchons, et en quoi ils présentent un exemple de divergence syntaxe-sémantique. Nous exposons ensuite les différents traitements proposés par BFN en termes de nombre de rôles noyaux fournis pour les constituants, avant de finalement présenter notre choix pour le French FrameNet.

7.3.1.1 Différentes catégories d'attributs

La fonction d'attribut de l'objet peut être occupée par des constituants de catégorie variée.

Nous donnons ci-dessous un exemple de chaque catégorie ainsi qu'un nombre d'occurrences comptées sur nos corpus cibles. Ces nombres d'occurrences ne sont donnés qu'à titre indicatif et pour avoir une idée des proportions de constructions à attributs de l'objet présentes dans les corpus. En effet, n'ayant compté que les constituants annotés *ATO*, l'étiquette du FTB dénotant un attribut de l'objet, nous ne relevons pas les cas de CAO avec changement de diathèse (les CAO dont le verbe principal est au passif sont surfaciquement annotés comme attributs du sujet). Nous sommes par ailleurs tributaires d'éventuelles erreurs d'annotation dans l'un ou l'autre des corpus. Parmi les exemples effectivement annotés ATO dans ces corpus, on trouve :

— des syntagmes adjectivaux (166 occurrences) :

(170) Les analystes trouvent [$_{\text{NP-OBJ}}$les chiffres] [$_{\text{AP-ATO}}$« raisonnablement bons »].

— des infinitives (88 occurrences) :

(171) Les cégétistes auront néanmoins vu [$_{\text{NP-OBJ}}$leur « part de marché »] [$_{\text{VPinf-ATO}}$passer de 42% à 33%].

— des syntagmes prépositionnels (55 occurrences) :

(172) L'ethnologue [...] considère [...] [$_{\text{NP-OBJ}}$l'entreprise africaine] [$_{\text{PP-ATO}}$comme un phénomène culturel].

— des participiales (29 occurrences) :

(173) un montant [$_{\text{NP-OBJ}}$que] les experts étrangers jugent [$_{\text{VPpart-ATO}}$sous-estimé]

— des syntagmes nominaux (27 occurrences) :

(174) [les] services des finances qui jugeaient [$_{\text{NP-ATO}}$pur gaspillage] [$_{\text{NP-OBJ}}$les 15 milliards de francs requis pour la liaison Rhône-Rhin]

On répertorie aussi des cas où l'attribut de l'objet est un syntagme adverbial (exemple (176)) ou une proposition relative (exemple (175)) mais on n'en trouve aucune occurrence dans le FTB.

(175) Je vois [$_{\text{NP-OBJ}}$Sam] [$_{\text{Srel-ATO}}$qui arrive].

(176) Alex trouve [$_{\text{NP-OBJ}}$ça] [$_{\text{Ad-ATO}}$bien].

7.3.1.2 Complément vs. modifieur

On reconnaît généralement deux grands types d'attribut de l'objet, que Guimier (1999) appelle les « compléments attributifs » et « modifieurs attributifs » (respectivement (177a) et (177b)).

(177) a. Sam croit son café chaud.
b. Sam boit son café chaud.

Cette dichotomie s'appuie sur le caractère sous-catégorisé (pour les compléments attributifs) vs. optionnel (pour les modifieurs attributifs) de l'attribut de l'objet dans ces constructions et le fait qu'il constitue un argument sémantique du verbe uniquement dans les cas de construction à complément attributif.

Les deux catégories ont été abondamment décrites dans la littérature : Riegel (1981) parle de verbes essentiellement attributifs et verbes occasionnellement attributifs, puis dans (Riegel, 1991) de construction à double-complémentation et d'attribut amalgamé. Rémi-Giraud (1991) parle d'attributs obligatoires vs. facultatifs et Muller (2000) de verbes opérateurs et verbes à concomitance. D'autres travaux, comme (Blanche-Benveniste, 1991) et (Tobback, 2005) identifient la dichotomie sans en nommer les catégories.

Nous recensons ici les arguments principaux (syntaxiques et sémantiques) cités par Guimier (1999, chap. 1 section 2.1) qui justifient cette distinction.

Plusieurs tests syntaxiques mettent en évidence le caractère sous-catégorisé des compléments attributifs et celui d'ajout des modifieurs attributifs. Le critère le plus clivant est le caractère obligatoire des compléments attributifs vs. optionnel pour les modifieurs attributifs. Ces derniers peuvent en effet être supprimés sans incidence sur la grammaticalité de la phrase et avec stabilité du sens du verbe (exemple (177d)), ce qui n'est pas le cas des compléments attributifs de l'objet. Par exemple en (177c) la suppression de l'attribut donne une phrase grammaticale mais inacceptable. Plus précisément, on peut obtenir une phrase acceptable avec la suppression d'un complément attributif, comme en (181b), mais la suppression modifie le rôle sémantique joué par l'objet direct (nous y revenons infra). Autre test syntaxique permettant de distinguer compléments et modifieurs attributifs : seuls les dépendants des compléments attributifs peuvent être extraits ou cliticisés (exemples (178) et (179), repris des exemples (44) et (45) de Guimier (1999)).

(177) a. Sam croit son café chaud.
b. Sam boit son café chaud.
c. #Sam croit son café.
d. Sam boit son café.

(178) a. Les aubergines, Paul les croyait farcies de chair à saucisse.
b. De quel type de chair Paul les croyait-il farcies ?
c. La chair à saucisse, dont Paul les croyait farcies, était avariée.
d. Paul en croyait les aubergines farcies.

(179) a. Les aubergines, Paul les a mangées farcies de chair à saucisse.
b. *De quel type de chair Paul les a-t-il mangées farcies ?
c. *La chair à saucisse, dont Paul les a mangées farcies, était avariée.
d. *Paul en a mangé les aubergines farcies.

En ce qui concerne la sémantique, Guimier (1999) pointe que le modifieur attributif n'appartient pas à la structure argumentale du verbe principal, tandis que le complément attributif est un argument sémantique du verbe avec lequel il apparaît. Ceci peut être vérifié simplement avec le test de suppression, cette fois utilisé pour dégager des propriétés sémantiques : lorsqu'on supprime un modifieur attributif, la phrase reste non seulement toujours acceptable (177d), mais on a bien une relation d'implication entre la version avec le modifieur attributif et la version sans celui-ci.

(180) a. Sam boit son café chaud.
b. ⇒ Sam boit son café.

Lorsqu'on supprime un complément attributif, on peut obtenir soit une inacceptabilité (177c), soit une phrase certes acceptable mais sémantiquement non impliquée par la version avec attribut (181b), ce qui montre que l'on passe à une version où l'objet direct joue un rôle sémantique différent.

(181) a. Sam croit Paul honnête.
b. ⇏ Sam croit Paul.

Pour la suite de ce travail, nous nous intéresserons uniquement aux constructions à *complément* attributif de l'objet, qui posent le problème le plus crucial de divergence syntaxe-sémantique [5]. Sauf mention explicite, c'est dorénavant à celles-ci que nous ferons référence quand nous parlerons de *constructions à attribut de l'objet* (ou *CAO*).

Nombre d'arguments syntaxiques et sémantiques des CAO

La partie annotation du projet ASFALDA consistant en fait à ajouter une couche d'annotations sémantiques sur un schéma d'annotation existant, les choix quant à la représentation syntaxique des CAO ne nous appartiennent pas.

Le guide d'annotation du FTB annonce l'objectif de « *contribuer à l'émergence d'un standard de découpage en constituants, sans doute un peu grossier mais suffisamment consensuel, et traductible dans différents cadres théoriques* » (Abeillé, 2004, p. 6). Pour les CAO, de la forme $N_0\ V\ N_1\ X$, le FTB analyse la séquence $N_1\ X$ en deux constituants, au contraire de nombreuses analyses de type générativiste qui en considèrent un seul. Cette analyse en « petite proposition » a particulièrement cours en anglais (e.g. Williams (1975), Stowell (1995), Rothstein (1995) et Hoekstra (1992)).

Nous redonnons rapidement dans cette section les arguments qui justifient le choix du FTB d'analyser la phrase (182a) en (182b) plutôt que (182c).

(182) a. Sam trouve cette personne sympathique.
b. Sam trouve [cette personne] [sympathique].
c. Sam trouve [cette personne sympathique].

Guimier (1999) puis Tobback (2005) pointent plusieurs tests classiques de constituance :

— La relativisation de la séquence *[N_1 X]* est agrammaticale alors que l'objet peut parfaitement être relativisé :
(183) a. cette personne que Sam trouve sympathique
b. *cette personne sympathique que Sam trouve

5. Pour les modifieurs, dont nous ne parlons pas plus avant, notons rapidement que l'annotation se fera par deux cadres distincts pour les deux relations de prédication, et que la relation de concomitance (Muller, 2000) qui lie ces deux prédications ne sera pas capturée.

— Il est tout à fait possible de faire porter une question sur l'objet ou sur son attribut, alors qu'il est impossible de questionner la séquence *[N_1 X]* dans son intégralité :

(184) a. Qui est-ce que Sam trouve sympathique ? Cette personne.
b. Comment est-ce que Sam trouve cette personne ? Sympathique.
c. *Qu'est-ce que Sam trouve ? Cette personne sympathique.

— De même, il est tout à fait possible de cliver l'objet ou son attribut, alors que le clivage de la séquence *[N_1 X]* est agrammatical :

(185) a. C'est cette personne que Sam trouve sympathique.
b. C'est sympathique que Sam trouve cette personne.
c. *C'est cette personne sympathique que Sam trouve.

— Il est possible de cliticiser l'objet mais pas vraiment la séquence *[N_1 X]* (Guimier, 1999, exemple (144a))

(186) a. Sam la trouve sympathique.
b. ??Sam trouve cette personne sympathique mais Alex ne le trouve pas.

Certains de ces tests sont nuancés par Tobback (2005), qui pointe que si N_1 et X forment chacun un constituant distinct, cela ne les empêche pas de former ensemble un constituant. L'auteur note aussi qu'il n'y a pas unanimité sur l'inacceptabilité de la pronominalisation de la séquence *[N_1 X]*. Pour ce qui est du clivage, elle rappelle l'argument de Guimier (1999) que certains verbes anglais comme *fear* acceptent le clivage de *[N_1 X]*, et constate qu'à l'inverse le clivage n'est pas toujours très grammatical même pour des séquences dont le statut de constituant est peu controversé. Il nous semble cependant clair que d'après les critères appliqués au sein du FTB pour dégager les constituants, les tests donnent de manière évidente une analyse en deux constituants séparés.

Maintenant que le statut de l'attribut en tant qu'argument syntaxique a été clarifié, il s'agit de voir dans quelle mesure il y a divergence entre syntaxe et sémantique pour les CAO, c'est-à-dire si les trois compléments sous-catégorisés sont sémantiquement sélectionnés par le verbe.

Guimier (1999) montre facilement que N_1 est sélectionné par X : la différence d'acceptabilité entre (187a) et (187b) indique que l'attribut impose des restrictions de sélection à l'objet, qui est donc bien son sujet sémantique.

(187) a. Sam croit la route praticable.
b. #Sam croit le café praticable.

Elle entend ensuite montrer qu'en revanche, N_1 n'est pas sélectionné par le verbe (dans le cas où on a affaire à un complément attributif et non pas à un modifieur). Son argumentation repose sur des paires comme (188).

(188) a. Sam croit la route praticable.
b. #Sam croit la route.

Nous avons déjà relayé en 7.3.1.2 sa preuve selon laquelle dans le cas d'un complément attributif, la version sans l'attribut met en jeu des propriétés sémantiques différentes : la phrase devient agrammaticale ou prend un sens différent, parce que l'objet ne remplit pas le même rôle dans une construction transitive stricte (sans CAO) et dans une CAO. L'inacceptabilité de (188b) ne peut donc pas nous donner de détails sur la structure argumentale de *croire* dans un cas de construction à attribut de l'objet.

Pour tenter de montrer que dans une construction à attribut de l'objet, l'objet direct du verbe principal n'est pas un argument sémantique de ce dernier, nous pointerons donc simplement que nous n'avons pu ni trouver, ni construire d'exemple où la construction à attribut de l'objet serait inacceptable dès lors que l'objet et son attribut sont sémantiquement compatibles.

Pour le cas des verbes de perception, comme nous l'avons déjà mentionné, l'implication observable en (196) est souvent présente, puisque l'objet direct est bien l'entité à l'origine de ce qui est perçu. On pourrait donc penser que l'objet direct est bien un argument sémantique du verbe de perception, contrairement à ce qu'on vient de voir pour les verbes comme *croire*. Encore une fois, des exemples comme (199), où le son perçu n'est pas émis par l'objet *le professeur* mais bien par ses élèves le chahutant, nous montrent que dans les CAO, l'objet direct n'est pas forcément cette entité et n'est donc pas un argument sémantique du verbe de perception [6].

7.3.1.3 Typologie sémantique

On peut regrouper les verbes à compléments attributifs en plusieurs catégories. Cette typologie est très variable selon les auteurs, c'est pourquoi nous n'en tracerons qu'une grossière esquisse.

Verbes pour lesquels une alternance avec une phrase à complétive est possible

Il s'agit de verbes pour lesquels la construction à attribut de l'objet de forme $N_0\ V\ N_1\ A$ peut être paraphrasée par une phrase où le verbe V sera complémenté par une proposition complétive *que* N_1 *copule* A.

(189) a. Sam trouve Alex sympathique.
b. Sam trouve qu'Alex est sympathique.

Parmi les verbes autorisant cette alternance, Guimier (1999) distingue les verbes comme *croire* et les verbes comme *vouloir*, qui se différencient des premiers parce qu'ils n'autorisent ni la cliticisation des compléments de l'attribut, ni l'insertion de la copule *être* devant l'attribut en cas de relativisation de l'objet direct. De plus, la complétive des verbes comme *vouloir* est généralement au subjonctif.

(190) a. Son film, que Sam croit (être) fidèle à la réalité historique.
b. Son film, que Sam veut *(être) fidèle à la réalité historique.

(191) a. Sam croit que son film est fidèle à la réalité historique.
b. Sam veut que son film soit fidèle à la réalité historique.

Cette relation de paraphrase, traditionnellement acceptée (voir par exemple Riegel *et al.* (2004)), peut néanmoins être questionnée. Mais avant de nous poser la question de la réalité d'une équivalence sémantique entre phrase avec CAO et phrase avec complétive, il semble pertinent de mentionner que parmi les verbes autorisant cette alternance, on rencontre des cas où la réalisation d'une version CAO d'une phrase avec complétive est tout bonnement impossible. Cela s'explique par le fait qu'un cas avec complétive est sémantiquement plus générique qu'un cas avec CAO, qui, elle, impose forcément une prédication sur l'actant réalisé par N_1. Dès lors, on peut construire des exemples comme (192) où un verbe qui autorise généralement la construction attributive est complémenté par une complétive, laquelle exprime une prédication sans actant et ne donne donc lieu à aucune version attributive.

(192) Sam trouve qu'il pleut beaucoup à Paris.

Lorsque l'alternance complétive/attribut de l'objet est réalisable, des phrases telles que (189a) et (189b) présentent à première vue une relation de paraphrase. Nous récapitulons ici les exceptions à cette hypothèse trouvées dans la littérature.

Pour Ruwet (1982), l'exemple (193a) ne révèle aucun jugement de l'énonciateur sur la validité de l'opinion mentionnée, alors que (193b) suppose un jugement négatif de l'énonciateur.

6. Guimier (1998) dégage des propriétés sémantiques différentes pour les deux cas, le contenu de la relative étant asserté ce qui n'est pas le cas pour l'infinitive.

(193) a. Paul croit qu'il est malade.
b. Paul se croit malade.

De la même façon, Olsson (1976) note que contrairement à (194a), dans (194b) le verbe est performatif : « *il décrit toujours un acte de communication* » mais « *implique en outre que la personne accomplissant l'acte de communication est également responsable de l'état des choses, déterminé conventionnellement, qu'il "rend public".* » (Borkin, 1974, p. 89)

(194) a. Le juge a déclaré que l'accusé était innocent.
b. Le juge a déclaré l'accusé innocent.

Verbes de perception

Les verbes tels que *voir*, *entendre* et *sentir*, permettent eux-aussi une construction avec complétive *que* N_1 *copule A* comme en (195b), mais le sens de cette dernière ne correspond pas à celui de la version CAO (exemple (195a)) qui serait mieux paraphrasée par une autre construction (exemple (195c)).

(195) a. Sam voit Alex ivre.
b. Sam voit qu'Alex est ivre.
c. Sam voit Alex et/au moment où Alex est ivre.

Guimier (1998) rattache les constructions à attribut de l'objet *adjectival* de ces verbes de perception aux constructions à attribut accessoire de l'objet de Le Goffic (1994). Ces dernières correspondent aux constructions à modifieur attributif vues plus haut — en particulier parce que, comme évoqué en 7.3.1.2, la suppression de l'adjectif attribut n'entraîne ni agrammaticalité, ni changement de sens de l'énoncé restant. En outre, la version CAO implique sémantiquement la version sans attribut, comme on peut le voir en (196).

(196) a. Sam entend Alex ivre.
b. ⇒ Sam entend Alex.

À première vue, cela vaudrait aussi pour les constructions attributives infinitives et relatives :

(197) a. Sam entend Alex arriver.
b. ⇒ Sam entend Alex.
(198) a. Sam entend Alex qui arrive.
b. ⇒ Sam entend Alex.

Néanmoins, des exemples clairs tels que (199) et (200) (Guimier, 1998), où la relation d'implication ne tient pas, nous montrent qu'on a bien affaire à des constructions à *complément* attributif de l'objet.

(199) a. Sam entend le prof se faire chahuter.
b. $\not\Rightarrow$ Sam entend le prof.
(200) a. Sam entend le prof qui se fait chahuter.
b. $\not\Rightarrow$ Sam entend le prof.

Cependant même pour les constructions attributives infinitives et relatives, il n'y a pas équivalence entre CAO et relative : Guimier (1998) pointe un changement de sens entre construction à attribut de l'objet (sens purement perceptif) et avec complétive (sens plus constatif) pour les verbes de perception, et Willems et Defrancq (2000) examinent les différents sens possibles du verbe *voir* suivi d'une complétive ou d'une construction à attribut de l'objet et montrent que les différentes constructions donnent lieu à différentes interprétations.

(201) a. Je vois Jean travailler consciencieusement son latin.
b. Je vois que Jean travaille consciencieusement son latin.

Verbes causatifs

Toujours selon Guimier (1999), ces verbes se distinguent de ceux de la catégorie précédente en ce qu'ils ne peuvent alterner avec une construction complétive.

(202) a. Sam rend Alex malade.
b. *Sam rend que Alex est malade.

De plus, leur objet direct n'est pas nécessairement réalisé (exemple (203)) et certains d'entre eux n'autorisent que les attributs de type adjectival (exemples (204)).

(203) L'absinthe rend fou.

(204) a. *Sam rend Alex en colère.
b. *Sam rend Alex debout.

Notre travail se préoccupant beaucoup des problèmes d'alternance CAO/complétive, nous ne nous intéresserons plus par la suite aux constructions à attribut de l'objet à verbes causatifs.

En ce qui concerne la relation de paraphrase entre complétive et CAO (hors verbes causatifs), il apparaît qu'elle est possible mais pas systématique. Les différents contre-exemples que nous avons vus concernent des cas spécifiques (verbes performatifs et verbes de perception) ou reposent sur des lectures n'étant pas forcément universellement partagées ; nous considèrerons donc que la paraphrase vaut dans le cas général, mais ne tient qu'en première approximation dans certains cas.

7.3.1.4 Traitement dans le FrameNet de Berkeley

Notre objectif est de fournir un traitement linguistiquement motivé des CAO dans le cadre du projet ASFALDA. Nous avons vu jusqu'ici que nous cherchions pour le French FrameNet à partir de la modélisation en cadres et du lexique existant pour l'anglais, et à proposer un équivalent pour le français où les modifications sont justifiées soit par des différences anglais/français, soit par une « correction » d'incohérences manifestes.

Nous devons donc commencer par étudier ce qui a été fait dans le BFN, pour des cas typiques de CAO en anglais (à noter donc que nous passons pour cette section à des données anglaises, ne cadrant pas forcément entièrement avec tout ce que nous avons vu supra sur les données du français). Nous avons cherché les cadres associés à des verbes « typiques » de l'attribut de l'objet en anglais pour observer le traitement qui en serait fait. Par « traitement », nous entendons ici le choix du nombre de rôles noyaux définis pour les cadres regroupant des déclencheurs qui permettent les deux constructions. D'après la caractérisation sémantique des CAO faite plus haut, on s'attend plutôt à une modélisation par des cadres avec deux rôles noyaux. Cependant, nous avons vu que les critères syntaxiques étaient parfois utilisés pour définir et typer les rôles, et Ruppenhofer *et al.* (2006, p. 20) indiquent en particulier qu'un objet direct est censé être représenté comme un rôle noyau. Il y a là une contradiction évidente, et cela se retrouve dans les données BFN.

Dans le BFN version 1.5, nous avons pu recenser trois types de traitement, dont nous parlerons ici plus avant.

— les cadres avec trois rôles noyaux (canoniquement un pour le sujet, un pour l'objet direct et un pour l'attribut)

— les cadres avec deux rôles noyaux (canoniquement un pour le sujet et un pour le deuxième argument sémantique du verbe)

— les cadres équipés de deux jeux de rôles noyaux utilisés distinctement pour annoter les deux types de constructions

On doit cependant préciser qu'étant donnée notre méthodologie, cette liste ne peut pas être exhaustive.

Cadres avec trois rôles noyaux

Le verbe *consider*, qui peut s'employer avec une construction à attribut de l'objet ou avec une complétive [7], est un déclencheur de CATEGORIZATION. Dans ce cadre, une entité sentiente (Cognizer) considère qu'un élément (Item) appartient à une certaine catégorie (Category). Les déclencheurs *consider, view, construe, understand* évoquent tous le cadre CATEGORIZATION.

(205) Whigs **CONSIDERED** his scruples over the oath ridiculous and inconsistent.
« Les Whigs considéraient ses scrupules à propos du serment comme ridicules et inconsistants. »

CATEGORIZATION: Category, Cognizer, Criteria, Item

Cette décision d'annotation pose le problème des verbes qui acceptent les constructions avec complétive aussi bien que les constructions à attribut de l'objet. Les données annotées BFN, annotations full-text comprises, ne comportent aucune occurrence du verbe *consider* avec complétive. Posons-nous donc hypothétiquement la question de l'annotation d'un exemple comme (206). Il apparaît qu'il faudrait choisir entre ne pas annoter ce type d'exemples comme évoquant le cadre CATEGORIZATION et « descendre » dans la complétive pour y annoter les rôles Item et Category. Cela paraît difficile puisque le BFN version 1.5 ne comporte absolument aucun exemple où les rôles seraient annotés au sein d'une complétive (lorsque le déclencheur n'appartient pas à cette complétive).

(206) Whigs considered that his scruples over the oath were ridiculous and inconsistent.
« Les Whigs considéraient que ses scrupules à propos du serment étaient ridicules et inconsistants. »

Ce découpage en rôles pose deux autres problèmes. D'abord, il acte une relation sémantique entre déclencheur et Item qui, nous l'avons vu en section 7.3.1.2, n'existe pas : l'objet du verbe principal d'une construction à attribut de l'objet ne remplit aucune position argumentale de ce verbe. Ensuite, tout cadre évoqué par le remplisseur du rôle Category aurait pour participant le remplisseur du rôle Item. Annoter ces rôles pour le cadre Categorization apparaît comme redondant avec l'annotation de ce potentiel cadre.

Cadres avec deux rôles noyaux

Nous avons étudié les cadres liés aux verbes de connaissance, d'opinion et de croyance, tels que *know*, *think* et *believe*, qui s'emploient aussi bien avec une complétive qu'avec une construction à attribut de l'objet. Le cadre AWARENESS présente des annotations des deux versions de cette alternance. Dans ce cadre, une entité sentiente (Cognizer) a un certain contenu (Content) dans son modèle du monde.

7. La version complétive de *consider* semble cependant bien moins usitée que sa traduction française.

Exemples d'annotation :

(207) Some **BELIEVED** her husband was dead.
« Certains croyaient son mari mort. »

(208) He **KNEW** he had made a bad mistake.
« Il savait qu'il avait fait une grosse erreur. »

(209) I had **THOUGHT** that I was the only murderer in the family.
« Je pensais que j'étais le seul meurtrier de la famille. »

AWARENESS: Cognizer, Content, Expressor, Topic

Les occurrences de constructions à attribut de l'objet sont dans ce cadre toutes annotées avec le même rôle : Content, qui englobe alors l'objet du verbe et son attribut.

(210) You may sincerely **BELIEVE** yourself capable of running a nightclub.
« Tu te crois probablement sincèrement capable de gérer une boîte de nuit. »

(211) She **KNEW** herself to be mortally ill.
« Elle se savait mortellement malade. »

(212) Coffin **THOUGHT** her very clever.
« Coffin la pensait très intelligente. »

AWARENESS: Cognizer, Content, Expressor, Topic

Ce choix d'annoter de la même manière les deux versions de l'alternance semble cohérent puisqu'il fait correspondre un rôle à chaque argument sémantique et permet de rendre compte de la relation de paraphrase entre deux phrases comme les exemples (189a) et (189b).

En nous intéressant aux verbes de perception évoqués supra comme pouvant sous-catégoriser un attribut de l'objet, tels que *see*, *hear* et *feel*, nous avons découvert un cas problématique. Dans le cadre PERCEPTION_EXPERIENCE, évoqué par les verbes susmentionnés, une entité sentiente (Perceiver_passive) perçoit sensoriellement un phénomène (Phenomenon). L'observation des annotations permet de trouver des exemples tels que (213) où, comme pour AWARENESS ci-dessus, un même rôle (Phenomenon) permet d'annoter l'objet et son attribut, mais aussi des exemples comme (214). Dans ces derniers, Depictive, un rôle non-essentiel du cadre qui permet de spécifier l'état du phénomène, est utilisé pour annoter les occurrences d'attribut de l'objet.

(213) a. She had never SEEN him so apoplectic.
« Elle ne l'avait jamais vu si furieux. »

b. Three couples had already been SEEN leaving the hotel with baggage.
« On avait déjà vu trois couples quitter l'hôtel avec leurs bagages. »

(214) a. She could SEE fly across his face a whole sequence of emotions and responses.
« Elle voyait défiler sur son visage tout une séquence d'émotions et de réactions. »

b. I had SEEN him painfully sewing on a shirt-button.
« Je l'avais vu cousant avec difficulté un bouton de chemise. »

PERCEPTION_EXPERIENCE : Perceiver_passive, Phenomenon, Depictive

Nous n'avons pu identifier de constante qui permette de différencier les exemples annotés comme (213) de ceux annotés comme (214). Cette inconsistance apparaît donc comme un problème de cohérence dans la définition des rôles, à moins que l'on n'ait affaire, dans un cas ou dans l'autre, à des erreurs d'annotation.

Cadres avec deux jeux complémentaires de rôles noyaux

Enfin, certains cadres présentent un troisième type de traitement, avec deux jeux de rôles noyaux. ESTIMATING comporte parmi ses déclencheurs le verbe *estimate*, qui comme on peut le voir en (215) et (216) permet aussi bien la construction complétive que la CAO.

Exemples tirés du BFN :

(215) a. It is ESTIMATED that the potential levy off live lamb exports is £750,000 a year.
« On estime que la taxe potentielle sur les exports d'agneaux vivants est de £750 000 par an. »

b. The world 's population, currently ESTIMATED to be about 5,300 million [...]
« La population mondiale, actuellement estimée à 5 300 millions [...] »

Exemples construits correspondant :

(216) a. The potential levy off live lamb exports is ESTIMATED to be £750,000 a year.
« La taxe potentielle sur les exports d'agneaux vivants est estimée à £750 000 par an. »

b. It is currently ESTIMATED that the world's population is about 5,300 million [...]
« On estime actuellement que la population mondiale est de 5 300 millions [...] »

Dans ce cadre, un cogniteur (Cognizer) fournit une valeur approximative (Value) pour un aspect particulier (Feature) d'un élément (Item).La définition indique que l'aspect estimé de l'élément correspond en fait à une question (Question) sur cet élément, et que la valeur de l'aspect correspond à une réponse à la question. Un rôle Estimation permettant une représentation propositionnelle de l'approximation du cogniteur est aussi proposé.

Cela est répercuté au niveau des contraintes sur les rôles. Value exclut Estimation, Feature requiert Item, et Item et Question requièrent Value. Ce système de contraintes fournit donc deux jeux de rôles, ce qui permet d'annoter dans le même cadre mais avec des rôles différents les constructions complétive, comme en (217) et (218) et à attribut de l'objet de même sens (comme en (219) et (220)).

Exemples d'annotation :

(217) Rather belatedly, Crick and Linsley (1983) **ESTIMATED** that a total of 20 'health effects' (cancer deaths plus hereditary effects in the first two generations) would have resulted from the accident.
« Un peu tardivement, Crick et Linsley (1983) estimèrent qu'un total de vingt 'effets sur la santé' (morts de cancer et effets héréditaires sur les deux premières génération) pourraient avoir résulté de l'accident. »

ESTIMATING : Cognizer, Estimation, Feature, Item, Question, Value

(218) This, PW **ESTIMATED**, would cost them between $4.4bn and $5.6bn.
« Cela, a estimé PW, leur coûterait entre 4,4 et 5,6 milliards de dollars. »

ESTIMATING : Cognizer, Estimation, Feature, Item, Question, Value

(219) It was the biggest demonstration since the lifting of martial law in 1988, and was **ESTIMATED** by police to number 10,000.
« Cela a été la plus grande manifestation depuis la levée de la loi martiale en 1988, et la police a estimé son nombre de participants à 10 000. »

ESTIMATING : Cognizer, Estimation, Feature, Item, Question, Value

(220) His fortune, conservatively **ESTIMATED** in excess of a hundred million dollars, had come intact to his beloved daughter, Louise. CNI
« Sa fortune, prudemment estimée à plus de cent millions de dollars, est arrivée intacte à sa fille bien-aimée, Louise. »

ESTIMATING : Cognizer, Estimation, Feature, Item, Question, Value

On ne trouve pas grande quantité de tels cas dans le BFN. Les exemples des cadres de souvenir présentés dans (Djemaa, 2014) ne nous semblent plus si clairs. Le cadre REMEMBERING_EXPERIENCE y était présenté comme proposant deux jeux de rôles distincts, ce qui est le cas mais nous semble devoir être modulé. En effet, aucune annotation de *remember* comme déclenchant REMEMBERING_EXPERIENCE n'étant présente dans les annotations BFN, nous avions appuyé notre analyse de quels rôles seraient utilisés pour quelle construction sur les exemples construits en (221) et (222). Or il nous semble à présent que les versions CAO et complétive de ces exemples présentent une nuance : les exemples avec CAO semblent bien décrire une expérience mais ceux avec complétive semblent plutôt reprendre une information, et seraient probablement à ce titre annotés en REMEMBERING_INFORMATION plutôt qu'en REMEMBERING_EXPERIENCE. À défaut d'annotation pour *remember* en REMEMBERING_INFORMATION comme en REMEMBERING_EXPERIENCE, il est difficile de déterminer si cela est le cas. La différence entre les deux cadres est de toutes manières difficile à clairement établir, raison pour laquelle ils ont pour le French FrameNet été fusionnés.

Exemples construits :

(221) a. I **REMEMBER** them taller.
« Je me les rappelle plus grands. »
b. I REMEMBER that they were taller.
« Je me rappelle qu'ils étaient plus grands. »

(222) a. I REMEMBER them singing on stage.
« Je me les rappelle chantant sur scène. »
b. I REMEMBER that they were singing on stage.
« Je me rappelle qu'ils chantaient sur scène. »

REMEMBERING_EXPERIENCE: Cognizer, Experience, Impression, Salient_entity, State

7.3.1.5 Choix de traitement pour le French FrameNet

Après avoir observé les différents traitements appliqués par le BFN au cas de discordance syntaxe-sémantique présenté par les constructions à attribut de l'objet, nous avons voulu prendre une décision que nous pourrions appliquer uniformément dans tous les cadres du French FrameNet dont des déclencheurs autoriseraient cette construction. Cela évite que le choix ne revienne aux différents annotateurs à chaque occurrence du phénomène, ce qui permettra à la fois de leur faciliter la tâche et d'obtenir une meilleure cohérence des données.

Encoder qu'un même lexème déclenchera deux cadres différents en fonction de la construction dans laquelle il s'insère entraînerait une polysémie systématique et, nous semble-t-il, injustifiée. Cela reviendrait à considérer que le verbe *trouver* possède un sens différent dans les exemples (189a) et (189b) que nous affichons à nouveau ici.

(189) a. Sam trouve Alex sympathique.
b. Sam trouve qu'Alex est sympathique.

De même, même si prévoir un jeu de rôles spécifique pour l'annotation de ce type de constructions nous permettrait d'encoder (189a) et (189b) sous le même cadre, cette solution ne nous paraît pas non plus convenir. En effet, prévoir un rôle essentiel pour l'attribut de l'objet reviendrait à

considérer ce dernier comme un argument sémantique du verbe déclencheur, ce qui n'est, nous l'avons vu plus haut, pas le cas.

Nous projetons donc d'annoter l'objet et son attribut au moyen d'un seul rôle, le même qui sera utilisé pour annoter une complétive de sens équivalent dans le même cadre. Nous rendons ainsi compte de la similarité sémantique approximative entre les versions complétive et à attribut de l'objet d'une même phrase, et conservons un point de vue sémanticiste en prévoyant un rôle noyau par rôle sémantique plutôt que de nous fonder sur la syntaxe. La relation de prédication entre l'objet et son attribut sera, le cas échéant, encodée par un cadre déclenché par cet attribut.

7.3.2 Déclencheurs nominaux : noms référant à une éventualité, un participant, un résultat ou une relation

Nous l'avons vu en 2.2.7, une des raisons qui nous ont poussés à choisir de travailler sur le modèle FrameNet est le fait que la ressource est multi-catégorielle, intégrant à son lexique aussi bien des noms et des adjectifs que des verbes (contrairement à VerbNet par exemple). Nous évoquons ici des points spécifiques aux déclencheurs nominaux.

Dans la plupart des cas, les cadres sont évoqués par des lemmes ayant une structure prédicative dont les arguments correspondent aux participants de l'éventualité décrite par le cadre. C'est le cas pour les déclencheurs verbaux, puisque nous avons vu en 4.2.3.1 que leurs arguments correspondent aux rôles noyaux du cadre qu'ils évoquent. Certains déclencheurs nominaux sont aussi dans cette situation : les noms prédicatifs qui, comme *discovery* en (223), font référence à une éventualité, et dont les arguments sont les participants de cette éventualité.

(223) This calculation was made possible by the recent **DISCOVERY** by nutritionists of the relationship between diet and health.
« Ce calcul a été rendu possible par la récente découverte par des nutritionnistes de la relation entre régime alimentaire et santé. »

BECOMING_AWARE: Cognizer, Means, Instrument, Phenomenon, Topic

BFN comporte aussi des déclencheurs nominaux qui, plutôt que faire référence à l'éventualité décrite par le cadre, font référence à un des participants de cette éventualité. C'est le cas en (224), où en plus d'évoquer le cadre BEING_NAMED, *names* remplit le rôle Name.

(224) He returned to his desk and checked the **NAMES** of the fourteen banks.
« Il retourna à son bureau et vérifia les noms des 14 banques. »

BEING_NAMED: Entity, Name

La possibilité pour les cadres d'être déclenchés par un nom de participant aussi bien que par un nom d'éventualité cadre bien avec les idées de (Fillmore, 1982) sur la sémantique des cadres. En effet, ce dernier décrit un cadre comme un « système de concepts, organisés de manière à ce que pour comprendre n'importe lequel d'entre eux, il faille comprendre toute la structure dans laquelle il s'insère ». Pour lui, « quand un des éléments de la structure est évoqué dans un texte ou dans une conversation, tous les autres sont automatiquement rendus disponibles » (Fillmore, 1982).

Certains noms prédicatifs, aussi appelés *noms relationnels*, mettent en relation deux entités (possiblement événementielles), et quand employés comme tête d'un NP saturé (dont tous les compléments sont remplis) et en emploi référentiel (la distinction entre emploi référentiel et emploi prédicatif d'un nom sera abordée section 7.3.2.1), réfèrent à un des arguments de la relation. C'est typiquement le cas des noms de membres de la famille. Le nom *père* évoque une relation de paternité entre deux individus mais dans *J'ai rencontré le père d'Alex*, les deux membres de la relation sont *Alex* et *le père d'Alex*. Ce type de noms contraste avec les noms comme *similitude*, qui évoquent aussi une relation mais réfèrent directement à cette relation, et en aucun cas à un de ses membres.

La modélisation en cadres des noms relationnels pourrait suivre la syntaxe, on aurait alors un cadre monovalent pour *père*, qui aurait pour seul rôle celui rempli par l'enfant du déclencheur. Ce n'est néanmoins pas ce qui semble le plus naturel. De fait, BFN propose pour les mots comme *père* un cadre KINSHIP fournissant un rôle par participant : un rôle Alter pour la personne à laquelle le déclencheur peut référer, et un rôle Ego pour la personne en relation à qui on présente l'Alter. Ces rôles sont annotés comme en (225).

(225) The **FATHER** of a young cancer victim has written a book which he hopes will offer comfort and help to the parents of children suffering from brain tumours.
« Le père d'une jeune victime de cancer a écrit un livre dont il espère qu'il offrira aide et réconfort aux parents d'enfants souffrant de tumeurs au cerveau. »

KINSHIP : Alter, Ego, Relatives

En plus de sembler plus naturelle, cette pratique permet de rassembler dans le même cadre des noms de relations (et verbes ou adjectifs évoquant cette relation) et des noms de membre de la relation. Le cadre KINSHIP n'en comporte pas d'exemple, mais le cadre CAUSATION peut aussi bien être évoqué par un nom remplissant le rôle Cause, comme *cause* ou *reason*, que par des verbes référant directement au lien entre cause et effet, comme *cause* ou *induce*.

Nous désignons ci-après comme *nom de participant* les noms déclenchant un cadre et référant, en emploi référentiel, à un de ses rôles. Il peut s'agir d'un nom relationnel (comme vu ci-dessus), ou d'une nominalisation, ambiguë avec un nom d'évènement (cf. section 7.3.2.4).

Dans les sections qui suivent, nous commencerons par rappeler la distinction entre emploi référentiel et emploi prédicatif du nom. Nous préciserons ensuite le type d'annotation utilisé pour les noms de participant, dont l'annotation ne sera pas la même selon qu'ils sont en emploi prédicatif ou en emploi référentiel. Enfin, nous nous pencherons plus précisément sur les nominalisations pouvant référer à un participant.

7.3.2.1 Emplois référentiels vs. emplois prédicatifs

Lors de la phase d'annotation en corpus, nous avons dû utiliser une distinction traditionnelle dans l'interprétation sémantique d'un groupe nominal (Riegel *et al.*, 2004, Chapitre VI, section 1.3 : Interprétation et usages communicatifs du groupe nominal) : l'usage référentiel et l'usage prédicatif. Le cas le plus fréquent est l'usage *référentiel*, pour lequel le nom réfère à une entité (réelle ou virtuelle), et l'usage de ce nom précis indique que l'entité en a les propriétés. Dans son usage *prédicatif*, le nom est utilisé uniquement pour prédiquer une propriété sur un autre élément de la phrase : il joue son rôle de prédicat mais ne réfère à aucune entité du monde. Cette distinction est valable pour tous les noms et non pas seulement les noms prédicatifs (dotés d'une structure argumentale). Ainsi, le nom *licorne* apparaît en emploi référentiel en (226), alors que (227) est un emploi prédicatif.

(226) As-tu vu la licorne ?

(227) Ceci est une licorne.

7.3.2.2 Annotation des noms de participants

Les emplois référentiels d'un déclencheur de type *nom de participant* seront annotés comme remplisseurs de rôles, le rôle rempli étant celui du participant auquel réfère le nom. C'est ce que l'on peut observer en (228). Les emplois prédicatifs comme les exemples en (229) sont eux annotés uniquement en tant que déclencheur. Le rôle auquel réfère habituellement le nom déclencheur en emploi référentiel est dans ces cas d'emplois prédicatifs rempli par un autre groupe nominal.

(228) a. [...] déterminer les **CAUSES** exactes de l'incendie.
b. Les **CONSÉQUENCES** de cette tension sont assez désastreuses.
c. George Bush [...] s'est félicité du **RÉSULTAT** de ce voyage.

(229) a. Première **CAUSE** de chômage, les fins de contrat à durée déterminée sont en hausse.
b. Cette mesure est la **CONSÉQUENCE** d'une conjoncture toujours défavorable.
c. Le **RÉSULTAT** de ces dispositions complexes est de ne combler qu'une partie du déficit de la sécu.

CAUSATION: Affected, Actor, Cause, Effect

Pour les emplois référentiels, nous avons de plus choisi de signaler l'occurrence de cadre au moyen d'un marquage : *REFTOPARTICIPANT*. Ce marquage peut paraitre redondant puisque d'après ce que nous venons de voir, il s'applique à toutes les occurrences de rôle dont le remplisseur a pour tête le déclencheur du cadre, mais ce n'est en réalité pas le cas. Il existe des déclencheurs qui incorporent lexicalement un rôle sans néanmoins référer au participant correspondant à ce rôle. En (230), le nom *privatisation* déclenche le cadre COMMERCE_SELL, et incorpore l'information que le rôle Seller est rempli par *l'Etat*. Ce rôle est donc annoté comme rempli par le déclencheur lui-même. Parmi les occurrences de rôles remplis par un déclencheur, le marquage REFTOPARTICIPANT permet donc de distinguer les cas d'incorporation lexicale de ceux où ce déclencheur réfère au participant correspondant au rôle. Il serait bien sûr possible de récupérer les cas REFTOPARTICIPANT en excluant de l'ensemble des cas où un déclencheur est tête d'un rôle ceux pour lesquels une incorporation lexicale a été signalée dans le lexique ; néanmoins, le marquage spécifique nous permet de repérer ces cas immédiatement et sans calcul.

(230) La présentation des modalités de la **PRIVATISATION** de Rhône-Poulenc n'a guère eu d'effet sur le cours du titre. UNI

COMMERCE_SELL: Buyer, Goods, Money, Rate, Seller

7.3.2.3 Liste non exhaustive des emplois prédicatifs d'un nom

Les emplois référentiels sont le cas le plus fréquent. Nous venons de voir que les noms de participants n'étaient pas annotés de la même manière selon qu'ils étaient en emploi prédicatif ou en emploi référentiel. Nous avons donc dû préciser des critères pour distinguer les deux types d'emploi. Faute de pouvoir donner une définition totalement opératoire, nous avons seulement listé dans le guide d'annotation des cas typiques d'emplois prédicatifs. Tous les autres cas sont par défaut considérés comme des emplois référentiels (à la référence réelle ou virtuelle).

Les phrases copulatives avec attribut nominal. Ces dernières sont de la forme *NP_1 copule NP-déclencheur*, où la copule peut être conjuguée avec auxiliaires et modaux et modifiée par des adverbes. Ainsi en (231), *conséquence*, déclencheur de CAUSATION, est en emploi prédicatif. Cette situation ne s'applique qu'aux cas où le déclencheur est directement tête de l'attribut. Dans le cas contraire, on considère que l'emploi est référentiel, comme c'est le cas en (232).

(231) Cette fusion est une **CONSÉQUENCE** directe de la fin de la guerre froide.

(232) En 1981, cette affaire pourrait être une des **CAUSES** de la défaite de Valéry Giscard d'Estaing face au candidat du parti socialiste François Mitterrand.

CAUSATION: Affected, Actor, Cause, Effect

Les phrases copulatives « inversées », aussi appelées *phrases copulatives spécificationnelles*, sont de la forme NP_1 *copule* NP_2. Dans ces phrases, c'est le premier syntagme nominal qui prédique une propriété sur le deuxième qui est, lui, référentiel. Comme on peut le voir en (233), le statut référentiel des NP peut être testé à l'aide d'une clivée, mais pour simplifier, nous avons considéré que si le premier syntagme nominal contenait un déclencheur du domaine Causalité, ce déclencheur était en emploi prédicatif.

(233) a. Ma meilleure amie est Alex.
b. C'est Alex qui est ma meilleure amie.
c. *C'est ma meilleure amie qui est Alex.

(234) Le **RÉSULTAT** est un taux d'épargne élevé, mais qui ne s'accroît plus, et dont la structure se modifie au détriment des placements longs. DNI

CAUSATION: Affected, Actor, Cause, Effect

Les appositions sans déterminant. Lorsque le déclencheur apparaît comme apposition sans déterminant au sein d'un NP, comme c'est le cas en (235), nous considérons qu'il est en emploi prédicatif.

(235) De la revente désastreuse d'une participation dans une compagnie pétrolière au ratage du lancement de l'augmentation de capital de la British Aerospace, en passant par le gonflement des « gros crédits », **CONSÉQUENCE** de la récession..., les faux pas ne manquent pas.

CAUSATION: Affected, Actor, Cause, Effect

Les phrases averbales et constructions de type « connecteur ». Il s'agit de syntagmes ayant pour tête un déclencheur nominal sans déterminant, et suivis d'une ponctuation. Nous considérons ces emplois comme prédicatifs, le déclencheur prédiquant généralement sur l'ensemble de la phrase qui suit.

(236) a. **CONSÉQUENCE** logique de la disparition du centre industriel de Billancourt, le comité d'établissement du CIB et celui du siège social de Renault doivent se dissoudre afin de regrouper leurs actifs au sein d'une nouvelle structure.
b. Seule **CONSÉQUENCE** pour le personnel : une dizaine de cadres iront travailler à Duluth, près d'Atlanta. DNI
c. **RÉSULTAT** : à la mi-mai, Lagos et quelques autres villes étaient le théâtre de violentes émeutes. DNI

CAUSATION: Affected, Actor, Cause, Effect

Les constructions de type *avoir pour N.* Dans ces constructions, de la forme NP_1 *Vavoir pour* $N_{déclencheur}$ NP_2, le déclencheur est en emploi prédicatif.

(237) Cela a souvent eu pour **CONSÉQUENCE** que de bons projets n'ont pas conduit à de véritables opportunités au niveau de l'emploi.

CAUSATION: Affected, Actor, Cause, Effect

7.3.2.4 Nominalisations ambiguës

Certains noms, généralement des déverbaux, présentent une forme ambiguë : leur sens peut être celui d'une éventualité (celle décrite par le verbe correspondant) comme celui d'un des participants

de cette éventualité, ou encore de son résultat. Comme on peut le voir en (238), c'est le cas du déclencheur *achat* qui peut, comme en (238a), référer à l'acte d'acheter un objet mais aussi dans les cas comme (238b), à l'objet acheté, remplissant alors le rôle Goods.

(238) a. Mais les chiffres que vient de publier l'INSEE montrent qu'en novembre les **ACHATS** des ménages en produits manufacturés ont reculé de 5% par rapport à octobre.

b. Selon les statistiques de la direction de l'industrie touristique, le marché des cadeaux-souvenirs et « **ACHATS** à rapporter » (montres, chaussures, vêtements...) a atteint l'an dernier dans l'Hexagone 10,4 milliards de francs, dont 6,4 milliards achetés par les Français en vacances et 4,2 milliards par les touristes étrangers.

COMMERCE_BUY: Buyer, Goods, Money, Rate, Recipient, Seller

Nous avons vu plus haut comment nous traitons les cas de référence à un participant, c'est-à-dire en annotant le constituant dont notre déclencheur est la tête comme remplisseur du rôle pertinent, et en signalant l'occurrence de cadre grâce à un marquage REFTOPARTICIPANT. Pour les cas où une nominalisation présente une ambiguïté de référence à l'éventualité ou à son résultat, ce résultat peut ne pas correspondre à un participant. Ainsi la sous-catégorisation du verbe *reproduire* fait intervenir le reproducteur et l'objet reproduit, mais pas le résultat de la reproduction. Sa nominalisation, *reproduction*, est ambigue : elle peut avoir un sens évènementiel mais aussi référer au résultat de l'évènement de reproduction. Dans un tel cas, aucun rôle ne pourrait être annoté comme rempli par le déclencheur nominal, et il faudrait en toute rigueur définir un cadre spécifique prévoyant un rôle supplémentaire. Nous avons rencontré un cas approchant : celui des nominalisations de verbes de communication, où la nominalisation peut souvent référer au contenu du message et pas seulement à l'acte de langage, comme c'est le cas en (239). Nous avons en pratique ajouté un rôle noyau Message pour ces cas, mais il apparaît a posteriori qu'un statut particulier devrait être créé pour ces cas : en effet, pour la plupart des déclencheurs du frame, le rôle Message ne peut pas être exprimé localement.

(239) [...] les traducteurs ont eu l'intelligence de juxtaposer aux travaux de Kondratieff les **COMMENTAIRES** de ses détracteurs [...]

L'enjeu pour nous en ce qui concerne ces déclencheurs ambigus est de désambiguïser les occurrences entre le sens d'éventualité et le sens de participant (ou résultat) d'une éventualité, afin de pouvoir annoter chaque occurrence comme il convient. Pour ce faire, nous avons fourni à nos annotatrices plusieurs principes et indices censés permettre de décider du statut de l'occurrence du déclencheur, et donc du type d'annotation à effectuer. Nous répertorions ici ces principes et critères.

Principe général. Lorsque les deux lectures sont possibles, et que les critères qui suivent ne permettent pas de distinguer entre les deux, nous avons choisi de privilégier la lecture événementielle et donc l'annotation sans marquage. Ainsi en (240), *cotisation* peut être lu comme la somme d'argent cotisée ou l'acte de cotiser (cette lecture peut être forcée avec *une cotisation par les membres de 17 euros par adhérent*). Nous privilégions alors la lecture événementielle.

(240) Du point de vue financier, l'association fonctionne essentiellement grâce à l'aide d'une subvention attribuée par le siège basé à Valenciennes ainsi qu'une **COTISATION** de 17 euros par an et par adhérents. UNI

COMMERCE_PAY: Buyer, Goods, Money, Rate, Seller

Les indications de durée. Lorsque le nom est ou peut être modifié par une indication de durée, nous considérons avoir affaire à un évènement. C'est ce qui a motivé l'annotation de *cotisation* en (241).

(241) Les déclarations de Mme Nicole Notat [...] en faveur d'une hausse de la contribution sociale généralisée (CSG) et d'un allongement de la durée de **COTISATION** au-delà de trente-sept ans et demi pour les retraites de la « Sécu », montrent que les syndicats finiront peut-être par se lasser de leur dédoublement de personnalité. DNI UNI

COMMERCE_PAY: Buyer, Goods, Money, Rate, Seller

La coordination. On ne coordonne généralement que des noms de même type sémantique, la coordination fournit donc un bon indice quant au statut d'une nominalisation ambiguë. Ainsi en (242a), la coordination d'*achats* avec *reventes*, qui est nécessairement évènementiel, force l'interprétation évènementielle, tandis qu'en (242b), la coordination avec *cadeaux-souvenirs* force l'interprétation d'*achats* comme remplissant le rôle Goods du cadre.

(242) a. Très logiquement, ce sont les fonds d'Etat qui ont alimenté les quatre cinquièmes de ces transactions, en raison des mouvements de hausse et de baisse des cours, générateurs d'un très gros montant d'**ACHATS** et de reventes. UNI UNI

b. Selon les statistiques de la direction de l'industrie touristique, le marché des cadeaux-souvenirs et « **ACHATS** à rapporter » (montres, chaussures, vêtements...) a atteint l'an dernier dans l'Hexagone 10,4 milliards de francs, dont 6,4 milliards achetés par les Français en vacances et 4,2 milliards par les touristes étrangers.

COMMERCE_BUY: Buyer, Goods, Money, Rate, Recipient, Seller

Les verbes supports d'action ou d'évènement. Lorsque le déclencheur ambigu complémente un verbe support d'action, comme *effectuer*, ou d'événement, comme *se dérouler*, nous considérons avoir affaire à un évènement. Parmi ces verbes d'action et d'événement : *faire*, *commettre*, *accomplir*, *mener à bien*, *procéder à*, *avoir lieu*...

(243) La COB avait alors établi que, au moment même où ces **ACHATS** étaient effectués, la banque détenait des informations non publiques. DNI DNI

COMMERCE_BUY: Buyer, Goods, Money, Rate, Recipient, Seller

Les restrictions de sélection du verbe ou du nom enchâssant. Lorsque le déclencheur ambigu complémente un verbe ou un nom imposant des restrictions de sélection sur son complément (c'est-à-dire des contraintes sur le type sémantique de ce complément), on a là un indice pour la désambiguïsation. Ainsi en (244), l'emploi du verbe *payer* complémenté d'un prix et du syntagme dont la tête est le déclencheur *acquisition* indique que ce dernier est ici remplisseur du rôle Goods de COMMERCE_BUY. En effet, on peut *payer somme_d'argent objet* mais pas *payer somme_d'argent évènement*.

(244) Le groupe de Pennsylvanie affirme payer 154 millions de dollars (831 millions de francs) sa nouvelle **ACQUISITION**, prix d' achat et reprise des dettes compris.

COMMERCE_BUY: Buyer, Goods, Money, Rate, Recipient, Seller

Les modifieurs référant aux processus. Lorsqu'un modifieur qui ne peut s'appliquer qu'à un processus et pas à une entité modifie un déclencheur ambigu, nous avons considéré avoir affaire à la version événementielle du déclencheur. Ainsi en (245), *décision* ne peut être qu'événementiel : à la différence du processus décisionnel, le contenu de la décision (ici *pas de renégociation du Traité*) ne saurait être *rapide*.

(245) M. Dumas a rappelé la **DÉCISION** « rapide, claire, ferme », prise par les Onze, la semaine dernière, à Oslo : pas de renégociation du Traité.

FR_DECIDING: Cognizer, Decision, Topic, Medium

Les modifieurs référant à l'agent des processus. Lorsqu'un élément qui modifie le déclencheur ambigu doit porter sur l'agent d'un processus pour que la phrase soit acceptable, on a nécessairement affaire à la version événementielle du déclencheur. Ainsi en (246), l'adjectif *courageuse* indique que l'agent de l'action de prise de décision déclenchée par *décision* (le gouvernement) a fait preuve de courage. Le déclencheur est donc événementiel.

(246) En annonçant le transfert en province de vingt-quatre organismes publics, le gouvernement avait pris, fin 1991, une **DÉCISION** courageuse. UNI

FR_DECIDING: Cognizer, Decision, Topic, Medium

TROISIÈME PARTIE

RÉSULTATS ET SYNTHÈSE

CHAPITRE 8

LE FRENCH FRAMENET

Sommaire

8.1 Évaluation de la qualité des annotations **183**
8.2 Présentation quantitative du corpus obtenu **185**
8.2.1 Extraction d'un lexique quantifié 185
8.2.2 Analyse quantitative 186
8.3 Les macro-rôles **190**

La version 1.2 du French FrameNet est disponible à l'adresse `http://asfalda.linguist.univ-paris-diderot.fr/frameIndex.xml`. Les annotations sémantiques sont librement disponibles ; le corpus intégrant toutes les couches d'annotation (morphologie, syntaxe et cadres) est diffusé à la condition d'obtenir la licence du French Treebank d'Abeillé *et al.* (2003), gratuite pour l'utilisation en recherche. Le présent chapitre s'applique à évaluer la tâche d'annotation, à décrire le lexique extrait des annotations, et à fournir des données quantitatives sur la ressource obtenue.

8.1 Évaluation de la qualité des annotations

Nous avons déjà mentionné en 5.2.2.2 que nous ne fournirions pas de mesure de l'accord inter-annotateurs pour le travail de création et validation du lexique. En effet, le lexique a évolué tout au long du travail : il a été impacté par les changements successifs de périmètre des cadres causés entre autres par le travail sur la nucléarité des rôles de chaque déclencheur ainsi que l'étude de la polysémie potentielle desdits déclencheurs. Même au cours de l'annotation, l'observation systématique des déclencheurs en contexte nous a fait découvrir sans cesse de nouveaux sens potentiels qui nous ont amenés à effectuer des tests de polysémie et parfois à de nouveau changer les frontières de cadres. À notre sens, la mesure d'un accord inter-annotateurs pour la tâche de constitution du lexique n'aurait donc pas vraiment d'intérêt.

En ce qui concerne l'annotation en cadres et en rôles des occurrences de déclencheurs potentiels du corpus à partir de fichiers pré-annotés, elle s'est effectuée en deux temps. Pour la majeure partie des fichiers pré-annotés, nous avons procédé comme décrit en 7.2, c'est-à-dire que chaque fichier a été annoté séparément par deux personnes, et que les conflits potentiels ont été résolus lors d'une adjudication. Par manque de temps et de moyens et pour compléter l'annotation des domaines, nous avons dû travailler différemment pour les derniers fichiers : chacun a été annoté par une seule personne — un responsable d'un des domaines de ses déclencheurs potentiels. Sur 24 717 occurrences de déclencheurs annotés (déclencheurs potentiels n'évoquant aucun cadre de nos domaines compris), 17 667 (environ 71,5%) ont été annotées par deux personnes et adjudiquées, contre 7 050

(environ 28,5%) annotées par une seule personne. L'évaluation de l'accord inter-annotateurs ne porte donc que sur les 71,5% des annotations qui résultent du travail de deux personnes.

Comme le font remarquer Burchardt *et al.* (2009a), les mesures habituellement utilisées pour l'expression de accord inter-annotateurs, comme le kappa, ne peuvent s'appliquer qu'aux tâches de classification pour lesquelles à la fois les classes et les éléments à classer sont fixés. Ce n'est pas le cas pour l'annotation du French FrameNet. Pour la tâche d'assignation des cadres, les éléments à annoter étaient presque fixés (à l'exception des verbes supports), mais les cadres variaient selon le déclencheur potentiel à annoter. Pour l'annotation des remplisseurs de rôles, les éléments auxquels assigner un rôle n'étaient pas fixés (c'est à l'annotateur de déterminer quel est le remplisseur du rôle, et en particulier son empan), et l'ensemble de rôles parmi lesquels choisir dépendait du cadre assigné par l'annotateur. Pour évaluer l'accord inter-annotateurs pour la tâche d'annotation, nous avons donc utilisé une F-mesure [1] inter-annotateurs, à la fois pour le choix des cadres et l'assignation des rôles. Pour cette dernière, nous fournissons une mesure pour la correspondance exacte et une pour la correspondance partielle. Pour la F-mesure s'appliquant à la correspondance exacte, les remplisseurs de rôles considérés comme communs sont les paires de remplisseurs étiquetés du même rôle et recouvrant exactement le même ensemble de tokens. Pour la F-mesure évaluant la correspondance partielle des remplisseurs de rôles, nous accordons aussi des points lorsque le même rôle a été assigné à deux remplisseurs partageant certains tokens : pour calculer le nombre de bonnes réponses, pour chaque remplisseur de rôle de l'annotation 1, nous cherchons dans l'annotation 2 un remplisseur avec à la fois la même étiquette de rôle et le plus haut ratio d'intersection sur union (plusieurs remplisseurs peuvent porter la même étiquette de rôle notamment en cas d'anaphore avec antécédent dans la phrase), et nous ajoutons ce plus haut ratio au total des bonnes réponses.

Nous avons calculé ces F-mesures pour toutes les occurrences de déclencheurs annotées indépendamment par deux annotateurs, c'est-à-dire comme nous l'avons mentionné plus haut environ 71,5% du total des occurrences de cadres du corpus actuel. Les résultats sont présentés table 8.1. La première colonne présente le nombre de déclencheurs pré-annotés proposés pour la double annotation, au total et répartis par domaine, puis par catégorie morpho-syntaxique. Notons ici qu'un déclencheur peut apparaître dans les comptes de plusieurs domaines, soit parce qu'il est associé à des cadres de domaines différents, soit parce qu'un cadre qu'il peut évoquer est lui-même à cheval sur plusieurs domaines. Le faible nombre d'occurrences de déclencheurs du domaine de la communication langagière vient du fait que ce domaine a seulement été partiellement annoté, et contient principalement des déclencheurs ambigus avec des cadres d'autres domaines.

Au total, l'accord est d'environ 86% pour les cadres et 77% pour les rôles lorsque le même cadre a été choisi par les deux annotateurs. L'accord est donc substantiel, ce qui montre que la tâche était bien définie. La ventilation par domaine notionnel montre que l'accord diffère selon les domaines. Il est calculé pour chaque domaine en ne considérant que les déclencheurs potentiels dont au moins un cadre appartient au domaine. L'information fournie par ces mesures porte donc à la fois sur l'ambiguïté entre les domaines et à l'intérieur d'un même domaine. Dans l'ensemble, le domaine présentant le meilleur accord pour l'assignation des cadres est celui des transactions commerciales, tandis que le domaine de la causalité présente le pire accord. Cela pourrait s'expliquer par le fait que l'ambiguïté interne au domaine de la causalité était la plus difficile à gérer. Quant à l'accord sur l'assignation des rôles, la ventilation par catégorie morpho-syntaxique des déclencheurs semble fournir la meilleure explication de la variation de l'accord observable entre les différents domaines : l'accord sur l'assignation de rôles est bien plus bas pour les déclencheurs nominaux que pour les

1. La F-mesure est la moyenne harmonique de la précision et du rappel si l'on considère l'annotation d'un annotateur comme la référence et celle de l'autre comme la prédiction. Intervertir les deux annotations inverserait la précision et le rappel, mais la F-mesure resterait identique.

2. Notons que ces comptes sont valables avant l'assignation de cadres, et donc la désambiguïsation des sens. Un déclencheur potentiel donné est associé à un domaine notionnel si au moins un de ses cadres appartient à ce domaine. Le compte total présenté au début du tableau ne correspond donc pas à la somme des comptes pour les quatre domaines.

	Nombre d'occ. de déclencheurs	% de N	% de V	F-mesure inter-annotateurs		
				Frame	Rôle exact	Rôle partiel
	17 667	36	50	85,9	77,2	81,9
Répartition par domaine notionnel[2]						
Transactions commerciales	3 307	60	40	92,0	73,4	80,4
Causalité	7 691	30	48	79,2	74,2	80,4
Positions cognitives	7 886	28	62	90,6	81,1	86,0
Communication langagière	2 221	23	76	89,6	82,3	87,5
Répartition par catégorie du déclencheur						
V	8 834	-	-	87,6	82,8	87,1
N	6 234	-	-	86,8	68,3	72,5
autre	2 509	-	-	77,7	74,6	82,1

TABLE 8.1: Accord inter-annotateurs pour toutes les occurrences de déclencheurs annotées indépendamment par deux annotateurs ; accord total puis réparti par domaine notionnel et par catégorie morpho-syntaxique du déclencheur. Première colonne : nombre d'occurrences de déclencheurs proposées pour la double annotation. Deux colonnes suivantes : pourcentages de noms et de verbes dans les occurrences de déclencheurs à annoter. Trois dernières colonnes : F-mesure de l'accord inter-annotateurs pour les tâches d'assignation des cadres, assignation des rôles avec correspondance totale, et assignation des rôles avec correspondance partielle. Les chiffres fournis dans ces deux dernières colonnes ne s'appliquent qu'aux cas où les deux annotateurs ont choisi le même cadre pour un déclencheur potentiel.

déclencheurs verbaux. Les domaines avec une forte proportion de verbes (positions cognitives, communication langagière) ont un meilleur score d'accord sur l'assignation des rôles que les deux autres domaines, qui ont bien moins de déclencheurs verbaux. Notons que l'accord obtenu pour les déclencheurs verbaux est comparable à celui rapporté par Burchardt *et al.* (2009a) pour l'allemand au sein du projet SALSA, qui se concentrait principalement sur les déclencheurs verbaux : les auteurs rapportent un accord de 85% pour les cadres et 86% pour les rôles de mêmes cadres, tandis que pour les mêmes mesures uniquement pour les déclencheurs verbaux, nous obtenons respectivement 87,6% et 82,8%. Notre meilleur accord pour les cadres pourrait être dû au fait que nous nous sommes cantonnés à quatre domaines.

8.2 Présentation quantitative du corpus obtenu

8.2.1 Extraction d'un lexique quantifié

À partir des annotations de cadres et rôles que nous venons de décrire, ainsi que des annotations morpho-syntaxiques pré-existantes à notre travail, nous avons pu extraire un lexique quantifié.

Nous appelons *lemmacat* l'association d'un lemme à une catégorie morpho-syntaxique. Notons dès à présent que cette notion est problématique, du fait du découpage en catégories morpho-syntaxiques utilisé pour obtenir nos chiffres. Pour la pré-annotation, nous avons choisi de repérer toutes les occurrences de lemmes sans tenir compte de leur catégorie, pour que les potentielles erreurs d'étiquetage du corpus n'entraînent pas de silence. Dans le lexique quantifié extrait des annotations, les catégories des déclencheurs sont récupérées directement dans le corpus. Cela a eu deux types de conséquences différentes : d'abord, certaines catégories ont dû être filtrées et rassemblées. Ainsi, les déclencheurs annotés comme adverbes interrogatifs ont été considérés avec les autres adverbes : cela évite de considérer comme distincts *c'est pourquoi.adv* et *c'est pourquoi.advwh*.

Ensuite, il arrive qu'une catégorie morpho-syntaxique soit présente dans le lexique entièrement à tort. Cela est particulièrement visible avec la catégorie *déterminant* : dans le lexique de départ (utilisé pour la pré-annotation), on ne trouve aucun déterminant. Le principe de pré-annotation par lemme sans tenir compte de la catégorie fait que 61 occurrences de déterminants ont été pré-annotées. Pour une de ces occurrences, c'est le nom *commerce* qui porte l'étiquette *déterminant* : il s'agit d'une erreur d'étiquetage morpho-syntaxique que notre procédure de pré-annotation nous a permis d'éviter d'oublier. Pour les 60 autres, il s'agissait du déterminant *certain* (l'adjectif *certain* peut évoquer le cadre FR_AWARENESS-CERTAINTY-OPINION) : ces occurrences ayant toutes été annotées *Other_sense*, nous les avons finalement filtrées et retirées des comptes car elles causaient une surestimation du nombre d'*Other_sense*. Elles devront à terme être retirées du corpus. Il existe probablement d'autres cas de ce type, mais notons qu'au vu du lexique extrait, cela semble très marginal.

Nous donnons à l'annexe C un extrait du lexique extrait des annotations. Il contient en particulier les nombres d'occurrences annotées pour chaque sens (couple lemmacat-cadre), quantifiant les nombres de cadres annotés. L'extrait présenté en annexe est au format cadre par cadre, qui permet de visualiser le lexique d'un cadre, mais nous proposons aussi un format lemme par lemme, qui permet de visualiser la polysémie annotée pour un lemme donné [3]. Dans le format lemme par lemme, nous fournissons également la proportion d'occurrences annotées comme « hors domaine » (cadre spécial *Other_sense*). Les lemmes avec un ratio de 0% ont leurs occurrences complètement couvertes par le French FrameNet. Des informations sur les fonctions grammaticales et catégories des remplisseurs de rôles sont également extrayables, mais nous ne l'avons pas encore fait, faute de temps (cf. nos perspectives de travail section 9.3).

8.2.2 Analyse quantitative

Nous présentons ici des données quantitatives sur le French FrameNet, calculées à partir du lexique quantifié extrait des annotations. Le French FrameNet comporte actuellement 24 922 annotations de cadres, dont 8 755 occurrences de déclencheurs annotées *Other_sense*, c'est-à-dire comme ne déclenchant aucun cadre du French FrameNet, et 16 167 occurrences de cadres de nos domaines (64,87 %). Nous désignerons par la suite sous le nom de « vraies occurrences » les occurrences non *Other_sense*. Pour rappel, l'utilisation de ce cadre factice permet de faire la différence entre les occurrences de déclencheurs examinées au cours de la phase d'annotation dont on a jugé qu'elles n'évoquaient aucun cadre de nos domaines (annotées *Other_sense*) et les occurrences de déclencheurs qui n'ont jamais été examinées (pas d'annotation de cadre). Les chiffres que nous venons de mentionner apparaissent dans la ligne « total » de la table 8.2. Pour comparaison, le BFN version 1.5 comptait 173 018 annotations, tous modes d'annotations confondus, et Rehbein *et al.* (2012) rapportaient environ 36 251 instances annotées pour la version 2.0 du projet SALSA de FrameNet de l'allemand. Notons néanmoins que ces projets se sont étendus sur une période beaucoup plus longue.

La table 8.2 présente la taille de notre lexique et le volume d'annotations, au total et par catégorie de déclencheur : dans les deux premières colonnes, elle fournit le nombre de lemmacats distincts avec au moins une annotation de cadre (*Other_sense* inclus, puis exclu) pour chaque catégorie morpho-syntaxique. La différence entre les deux colonnes correspond aux lemmes qui n'ont que des occurrences *Other_sense*. Les deux colonnes suivantes fournissent le volume des annotations : d'abord le nombre d'occurrences de cadres autres qu'*Other_sense* pour les déclencheurs de chaque catégorie ; puis, le nombre d'occurrences d'*Other_sense* annotées pour les déclencheurs de chaque catégorie.

3. L'intégralité du lexique extrait dans ses deux formats est visualisable à partir de la page d'accueil des données sur le site du French FrameNet : `http://asfalda.linguist.univ-paris-diderot.fr/frameIndex.xml`.

La table 8.3 reprend les mêmes nombres mais les présente ventilés par plage de proportion d'*Other_sense* dans les annotations : moins un lemmacat a d'annotations *Other_sense*, plus nos données permettront pour lui une désambiguïsation lexicale efficace. Cette table nous permet de montrer que 490 lemmacats du French FrameNet (dont 225 verbes et 179 noms) ne comportent aucune annotation *Other_sense*. Pour ces 490 lemmacats, qui représentent un total de 7 213 occurrences de cadres, aucune occurrence n'a été marquée comme hors-domaine, et les données sont donc complètes pour une tâche de désambiguïsation lexicale guidée par les données (dans la mesure bien sûr de la restriction de la couverture du corpus utilisé). Nous verrons cela dit plus bas que sur ces 490 lemmacats, seuls 85 présentent une ambiguïté de cadres (405 sont annotés avec un seul cadre).

Intéressons-nous maintenant à la couverture du French FrameNet : nous avons vu table 8.2 qu'il comportait des annotations sur 1 009 lemmacats distincts — 873 si on exclut les *Other_sense*. Dorénavant, sauf mention explicite du contraire, les comptes réalisés seront donnés à l'exclusion des *Other_sense*. On peut voir à la dernière ligne de la table 8.4 que la ressource compte 105 cadres distincts ayant au moins une annotation, et 1 109 sens (c'est-à-dire couples lemmacat-cadre) distincts. Pour comparaison, le BFN version 1.5 comportait 7 711 déclencheurs avec au moins une annotation pour 1 019 cadres distincts. Pour le projet SALSA, Rehbein *et al.* (2012) comptaient 648 lemmes, 1 023 cadres (dont des proto-cadres créés en cas de « trous » du BFN, ne valant que pour un seul déclencheur — sans généralisation) et 1 826 sens distincts.

Les autres lignes de la table 8.4 montrent ces mêmes chiffres, ainsi que le volume d'annotations, répartis selon les domaines d'appartenance des cadres, puis selon la catégorie morpho-syntaxique du déclencheur. Les quatre premières lignes concernent la répartition par domaine.

— La colonne 1 donne le nombre d'occurrences de cadres dont le cadre appartient à tel ou tel domaine. À noter qu'un cadre appartenant à plusieurs domaine est compté dans chacun d'entre eux; cela explique que la somme des comptes pour les domaines ne donne pas le même résultat que les totaux présentés à la dernière ligne, ni que la somme des comptes sur les catégories. On peut remarquer que les domaines comptant le plus d'annotations sont les positions cognitives et la communication verbale.

— La colonne 2 donne le nombre de cadres distincts par domaine. On note que le domaine de la causalité comporte très peu de cadres comparativement au nombre d'annotations : les cadres de ce domaine sont plus généraux que dans les autres domaines.

Catégorie de déclencheur	Nombre de lemmacats distincts (*Other_sense* inclus)	Nombre de lemmacats distincts (*Other_sense* exclu)	Nombre d'occ. de cadres ≠ *Other_sense*	Nombre d'occ. de cadres *Other_sense*
a	53	43	234	199
adv	35	26	407	310
cc	5	2	101	31
conj	25	22	301	70
n	336	296	5 282	2 956
prep	44	35	674	541
v	504	446	9 165	4 581
autre	7	3	3	67
total	1 009	873	16 167	8 755

TABLE 8.2: Taille du lexique et volume d'annotations par catégorie morpho-syntaxique, avec et sans *Other_sense*

Catégorie de déclencheur	Nombre de lemmacats distincts (LD) & nombre d'occurrences de cadres (OC) par proportion de *Other_sense* dans les annotations											
	0%		0 < x < 25%		25 ≤ x < 50%		50 ≤ x < 75%		75 ≤ x < 100%		100%	
	LD	OC	LD	OC	LD	OC	LD	OC	LD	OC	LD	OC
a	25	107	4	90	1	4	8	111	5	66	10	55
adv	17	148	4	230	1	15	1	32	3	270	9	22
cc	2	101	0	0	0	0	0	0	0	0	3	31
conj	15	267	1	16	1	3	2	12	3	67	3	6
n	179	2978	36	1552	18	534	32	1328	31	1306	40	540
prep	24	260	5	306	4	178	1	10	1	108	9	353
v	225	3349	86	4155	49	1829	45	1705	41	2320	58	388
autre	3	3	0	0	0	0	0	0	0	0	4	67
total	490	7213	136	6349	74	2563	89	3198	84	4137	136	1462

TABLE 8.3: Taille du lexique et volume d'annotations par catégorie morpho-syntaxique, ventilée par proportion de *Other_sense* dans les annotations

— La colonne 3 donne le nombre de sens (couples lemmacat-cadre) par domaine. On note que le domaine des transactions commerciales a un lexique assez petit par rapport au nombre d'annotations (2 930 annotations pour 99 sens distinct), donc un nombre moyen d'annotation par sens d'environ 30, contre 12,7 pour le domaine des positions cognitives, où l'on a plus de polysémie.

Les lignes suivantes de la table 8.4 présentent une ventilation par catégorie du déclencheur du cadre annoté. On donne en particulier en colonne 2 les nombres de cadres qui ont au moins une annotation dont le déclencheur est de telle ou telle catégorie (un même cadre est donc compté plusieurs fois). On constate que la plupart des cadres ont au moins un déclencheur verbal (98 cadres sur 105). Les prépositions et conjonctions déclenchent essentiellement des cadres de causalité, ou de but, avec en outre le cadre FR_ATTRIBUTED_INFORMATION (*d'après, selon...*). En colonne 1, on peut voir que plus de la moitié des vraies occurrences de cadres sont déclenchées par des verbes (9 165/16 167), et presque un tiers par des noms. On retrouve cette proportion dans la répartition des sens (colonne 3).

Enfin, la première partie de la table 8.5 présente d'autres chiffres ayant trait à la couverture de la ressource. On y retrouve d'abord des mesures sur le nombre d'occurrences de cadres par lemmacat (première ligne), puis sur le nombre d'occurrences de cadres par sens distinct (deuxième ligne). Ainsi par exemple un quart des lemmacats a plus de 21 occurrences de cadres. La dernière ligne de la table 8.5 fournit des informations sur la polysémie des déclencheurs du French FrameNet. Elle présente des mesures sur le nombre de cadres distincts par lemmacat pour l'ensemble du corpus. Notons d'abord que le nombre de cadres par lemme du French FrameNet est moindre que celui annoncé par Burchardt *et al.* (2009a) (2,33 cadres par lemme) mais cela s'explique aisément vu l'approche lemme à lemme du projet SALSA et le fait que nous nous restreignons à quelques domaines notionnels précis. On peut voir par ailleurs que le troisième quartile donne un seul cadre : au moins les trois quarts des 873 lemmacats ayant au moins une vraie annotation ne déclenchent en fait qu'un seul vrai cadre (plus éventuellement le cadre Other_sense). Pour analyser un peu plus la polysémie du lexique, nous donnons une répartition du lexique supplémentaire table 8.6.

	Nb. d'occ. de cadres	Nb. de cadres distincts	Nb. de sens
Cause	3 895	11	285
Pos. cog.	5 426	44	442
Com. lang.	5 233	47	411
Trans. com.	2 930	19	99
a	234	21	48
adv	407	10	42
cc	101	3	5
conj	301	5	28
n	5 282	85	346
prep	674	10	43
v	9 165	98	594
autre	3	3	3
en tout	16 167	105	1 109

TABLE 8.4: Couverture du French FrameNet : découpage par domaine et par catégorie (*Other_sense* exclu).
Colonne 1 : nombre d'occurrences de cadres dont le cadre appartient à tel domaine / dont le déclencheur est de telle catégorie morpho-syntaxique / en tout.
Colonne 2 : nombre de cadres distincts, dont le cadre appartient à tel domaine / ayant au moins un déclencheur de telle catégorie / en tout.
Colonne 3 : *idem* colonne 1 mais pour les couples cadre-lemmacat.
Dans la ventilation par domaine, un même cadre appartenant à plusieurs domaines est compté pour chacun de ses domaines.

	Moyenne	Écart moyen absolu	1er quartile	Médiane	3e quartile
Nb. d'occ. de cadres par lemmacat	18,52	19,25	2	6	21
Nb. d'occ. de cadres par sens	14,58	15,05	2	5	15
Nb. de cadres distincts par lemmacat	1,27	0,43	1	1	1

TABLE 8.5: Couverture et polysémie de la ressource (*Other_sense* exclu)

	Nombre de lemmacats distincts
un seul sens = *Other_sense*	136
un seul sens ≠ *Other_sense*	405
un seul vrai sens, plus *Other_sense*	294
> 2 sens, plus *Other_sense*	89
> 2 sens dont aucun *Other_sense*	85

TABLE 8.6: Polysémie des lemmacats du French FrameNet : répartition des nombres de sens

8.3 Les macro-rôles

Les relations entre cadres permettent entre autres de définir des relations entre rôles : des relations d'identité peuvent être définies entre les rôles de deux cadres unis par une relation. Cela permet dans une certaine mesure de généraliser sur les rôles, ce qui peut être utile par exemple pour l'apprentissage automatique. Ce type de généralisation est également intéressant pour mettre pleinement à profit le regroupement des cadres en domaines, en obtenant des rôles généraux et centraux aux domaines (comme par exemple les participants recevant ou donnant l'argent dans le domaine des transactions commerciales). Nous avons présenté en 6.7 le travail sur les relations du domaine des positions cognitives. La représentation de ces relations, figures 6.1 et 6.2, permet de voir que tous les cadres du domaine ne sont pas liés, même par l'intermédiaire de plusieurs relations entre cadres. Rien ne permet donc de signaler que le Cognizer de CATEGORIZATION, qui a une position cognitive sur l'appartenance d'un élément à une catégorie, s'apparente au Cognizer de FR_AWARENESS-CERTAINTY-OPINION, qui a une position cognitive sur la certitude d'un contenu. Nous avons décidé de mettre à profit notre approche par domaines notionnels pour généraliser sur les rôles plus directement et plus largement que par le biais de relations cadre à cadre. Nous avons ainsi défini pour chaque domaine un certain nombre de *macro-rôles*, auxquels sont rattachés les rôles pertinents des cadres du domaine. Ces macro-rôles constituent un niveau intermédiaire de granularité, qui permet des généralisations plus larges que ce que permettent les rôles des cadres, mais qui restent bien plus spécifiques que des rôles génériques comme agent, patient ou thème. Ils sont également moins orientés par la syntaxe : par exemple le Cogniteur n'est pas forcément sujet, contrairement au rôle agent traditionnellement attribué aux sujets.

Notons ici qu'un groupe de macro-rôles commun à plusieurs cadres permet d'identifier un domaine. Ainsi les cadres du domaine des positions cognitives ont tous un rôle lié au macro-rôle Cognizer et un rôle lié au macro-rôle Content_or_entity.

Le macro-rôle Cognizer, qui couvre les personnes ayant une position sur un contenu, rassemble par exemple 60 rôles issus des 52 cadres du domaine des positions cognitives. Plusieurs rôles du même cadre peuvent apparaître sous le même macro-rôle quand ces rôles sont en relation d'exclusion. BE_IN_AGREEMENT_ON_ASSESSMENT, par exemple, décrit l'accord (ou désaccord) de deux cogniteurs — aussi exprimable séparément (Cognizer_1 et Cognizer_2) — sur un contenu. Les rôles Cognizer_1 et Cognizer_2 sont en relation d'exclusion avec le rôle Cognizers. Cognizers et Cognizer_1 sont tous deux rattachés au macro-rôle Cognizer. Cognizer_2 est quant à lui rattaché à un macro-rôle Co_cognizer qui rassemble les rôles du même type.

Il arrive aussi marginalement qu'un même rôle soit rattaché à plusieurs macro-rôles. C'est souvent le cas pour les cadres à cheval sur plusieurs domaines, par exemple celui des positions cognitives et celui de la communication langagière. Dans le cadre FR_ATTEMPT_SUASION.VERACITY, un locuteur (Persuader) tente de convaincre un interlocuteur (Addressee) de croire en la réalité d'une situation (State_of_affairs). Il s'agit d'une situation de communication ; le rôle Persuader est rattaché au macro-rôle Speaker, et le rôle Addressee au macro-rôle du même nom. Mais ce cadre appartient aussi au domaine des positions cognitives : le locuteur tente d'influencer l'interlocuteur pour lui faire acquérir une position cognitive sur la situation. L'Addressee est donc aussi rattaché au macro-rôle Cognizer, et le State_of_affairs au macro-rôle True_content (qui couvre les contenus et situations sur la véracité desquels le Cognizer se positionne). Dans de tels cas, un marquage signale la priorité des macro-rôles : FR_ATTEMPT_SUASION.VERACITY a d'abord trait à un acte de communication (nous avons considéré que la communication primait sur la position cognitive), c'est donc le macro-rôle Addressee qui a la priorité sur le macro-rôle Cognizer qui englobe le même rôle du cadre.

Dans certains cas, il nous a paru souhaitable de laisser le choix entre plusieurs niveaux de granularité. Nous avons donc créé deux niveaux de macro-rôles. Les *sur-macro-rôles* peuvent englober des macro-rôles, mais aussi directement des rôles de cadres lorsque ces derniers sont

sous-spécifiés. Nous avons vu au chapitre 6 que nous avions intégré au domaine des positions cognitives des cadres proches de ceux qui correspondent à notre notion de départ, une position d'un cogniteur sur la véracité d'un contenu : dans ces cadres, le cogniteur a une position sur l'opportunité d'effectuer une action, ou sur la légitimité d'un contenu ou d'une situation. En plus du macro-rôle True_content que nous avons évoqué supra, nous avons donc un macro-rôle Action pour les actions qu'un cogniteur estime opportunes, et Legit_content pour les contenus qu'un cogniteur juge légitimes ou bien-fondés. Un sur-macro-rôle Content_or_entity, qui a pour sens *contenu ou entité sur lequel un cogniteur a une position* regroupe ces différents macro-rôles. Il englobe aussi un rôle de cadre : le rôle State_of_affairs du cadre FR_ATTEMPT_SUPPORTED_SUASION, dans lequel un locuteur présente un support à un interlocuteur pour le convaincre qu'un State_of_affairs est vrai ou légitime. Le statut *à vérifier* ou *à légitimer* du rôle State_of_affairs étant sous-spécifié, il nous a semblé préférable d'associer ce rôle au macro-rôle englobant les trois types de contenus sur lesquels un cogniteur peut se positionner.

Le French FrameNet compte en tout 41 macro-rôles, pour un total de 521 paires rôle de cadre/macro-rôle. Quatre des ces macro-rôles sont des sur-macro-rôles, les 37 autres n'englobent que des rôles de cadres. La hiérarchie des macro-rôles, présentée annexe D, est distribuée avec les données.

Nous souhaiterions dans des travaux futurs lier ces macro-rôles aux rôles plus généraux de VerbNet, ce qui permettrait non seulement d'établir une correspondance entre nos rôles et les rôles VerbNet, mais aussi de pouvoir entraîner des système automatiques d'étiquetage en rôles sémantiques sur notre corpus en utilisant des rôles à la granularité bien plus large que ceux de nos cadres et que nos macro-rôles. Nous aborderons cette question un peu plus en détail dans la section consacrée aux perspectives de travail (cf. 9.3).

CHAPITRE 9

CONCLUSION

Sommaire

9.1 Synthèse **193**
9.2 Bilan **195**
9.3 Perspectives **195**

9.1 Synthèse

Le travail présenté dans cette thèse, effectué dans le cadre du projet ASFALDA, a mené à la création d'un corpus du français annoté en sémantique : le French FrameNet [1]. Cette ressource s'appuyant sur le FrameNet de Berkeley (Baker *et al.*, 1998) a été annotée sur deux corpus arborés : le French Treebank d'Abeillé *et al.* (2003) et le Sequoia Treebank de Candito et Seddah (2012). Elle couvre quatre domaines notionnels : la cause, la communication langagière, les positions cognitives et les transactions commerciales. La version 1.2, dernière disponible, comporte 16 167 annotations de cadres de nos domaines et de leurs rôles, portant sur 105 cadres et 873 déclencheurs distincts. Cela se traduit par 1 109 paires déclencheur-cadre distinctes, c'est-à-dire 1 109 sens.
La caractéristique de FrameNet est d'unifier le traitement de plusieurs catégories d'unités lexicales. Ainsi, les 873 déclencheurs ayant au moins une annotation sont surtout des verbes (446) et des noms (296), mais on y trouve également des prépositions, conjonctions, adverbes et adjectifs. La ressource (version 1.0) a d'ores et déjà été utilisée par Michalon *et al.* (2016) pour entraîner un système de repérage automatique de cadres et de rôles sémantiques, ce qui peut être vu comme une analyse sémantique superficielle.

Dans le manuscrit qui s'achève ici, nous avons d'abord replacé la sémantique des cadres dans le contexte des différentes conceptions des rôles sémantiques, et comparé FrameNet à d'autres entreprises de généralisation sur les prédicats et typage des arguments de ces prédicats. Nous avons ainsi pu justifier notre choix de nous appuyer sur le modèle FrameNet dans le cadre du projet ASFALDA. Nous reviendrons sur nos raisons initiales pour choisir le modèle FrameNet à la section 9.2 pour établir si nos attentes correspondaient ou non à la réalité.

Nous avons introduit une nouvelle méthodologie de création de ressource de type FrameNet : le travail domaine par domaine, une méthode cadre à cadre guidée par les données où les cadres et leurs déclencheurs sont sélectionnés pour leur appartenance à des domaines notionnels. À la

1. Consultable à l'adresse suivante : `http://asfalda.linguist.univ-paris-diderot.fr/frameIndex.xml`. Les annotations sémantiques sont librement disponibles. Le corpus intégrant toutes les couches d'annotation (morphologie, syntaxe et cadres) est diffusé à la condition d'obtenir la licence du French Treebank, gratuite pour l'utilisation en recherche.

différence de l'approche cadre à cadre avec annotation lexicographique du FrameNet de Berkeley, notre méthode préserve la distribution des sens et des constructions syntaxiques du corpus annoté, et rend systématiquement compte de la polysémie des lexèmes à l'intérieur de nos domaines. À la différence des approches lemme à lemme comme celle du projet SALSA de FrameNet de l'allemand (Burchardt *et al.*, 2006b) —ainsi que des approches occurrence par occurrence de déclencheur comme l'annotation de plein texte du FrameNet de Berkeley, notre approche permet une couverture en cadres totale pour nos domaines et une couverture lexicale maximale à l'intérieur de nos cadres. Nous avons développé et employé cette méthodologie pour créer un FrameNet du français à partir d'un FrameNet préexistant, mais il nous semble qu'elle serait aussi utilisable pour créer une telle ressource de toutes pièces.

Nous avons aussi reprécisé un certain nombre de notions introduites par le modèle FrameNet en les ancrant dans la syntaxe, sans toutefois en faire le seul critère. Il s'agit plutôt de formaliser certains critères qui semblaient déjà utilisés par le FrameNet de Berkeley au vu des données annotées et des distinctions de cadres opérées, mais qui n'étaient pas explicitées. Ainsi la notion de nucléarité d'un rôle paraissait entièrement sémantique, mais proposait plusieurs propriétés suffisantes au statut de rôle noyau, toutes exprimables en termes syntaxiques uniquement. Nous les avons remplacées par une condition nécessaire et suffisante qui les applique et les renforce. Nous avons ainsi pu formaliser un peu plus les critères de délimitation des cadres proposés par le FrameNet de Berkeley qui dépendaient entre autres de la définition de la nucléarité.

Nous avons présenté le travail de création du French FrameNet en explicitant et justifiant autant que possible la méthodologie employée. Ce faisant, nous avons rassemblé et présenté des tests qui nous ont aidés dans notre travail, comme les tests de polysémie que nous avons souvent utilisés pour la délimitation des cadres. La création du French FrameNet peut être découpée en trois étapes, mais il serait illusoire de considérer ces étapes comme entièrement distinctes. Le travail fourni à chaque étape nous a amenés maintes fois à effectuer des modifications sur les résultats du travail déjà effectué. Ces trois étapes sont :

— la délimitation de la structure du French FrameNet : définition des domaines, choix de cadres préexistants devant appartenir à ces domaines, repérage des potentiels doublons, et vérification de la cohérence des relations liant entre eux les cadres de chaque domaine ;

— la constitution du lexique des cadres sélectionnés, à partir des lexiques fournis par Padó (2007) et Mouton *et al.* (2010), ainsi que de lexiques préexistants pour certains domaines. Cette étape a donné lieu à des études de polysémie sur les déclencheurs potentiels pour le cas échéant créer deux cadres à partir d'un seul, ainsi qu'au repérage des « trous de cadre » c'est-à-dire lorsque des lexèmes qui nous paraissent avoir leur place dans un domaine ne semblent évoquer aucun des cadres préexistants. C'est aussi au cours de cette phase que nous avons parfois été confrontés à des incompatibilités entre les rôles noyaux des déclencheurs que nous souhaitions ajouter à un cadre et ceux proposés par ledit cadre. Ce type de problème a pu nous amener à modifier les cadres choisis à l'étape précédente, ou à en créer de nouveaux spécialement pour certains déclencheurs ;

— l'annotation en cadres et en rôles des occurrences dans nos corpus cibles de chaque lemme du lexique (en coupant aux cent premières occurrences en cas de lemme en comportant plus de cent). Cette étape nous a parfois amenés à découvrir des réalisations inattendues de nos déclencheurs potentiels : avec un sens visiblement différent, ou avec des arguments syntaxiques ne s'adaptant aux rôles d'aucun des cadres jusqu'alors prévus pour le lexème. Nous avons alors dû opérer des changements de lexique (ajouter le déclencheur à un de nos autres cadres) ou parfois créer de nouveaux cadres.

Nous avons aussi présenté l'application des deux premières de ces étapes au domaine des positions cognitives, qui englobe les cadres référant à une situation où un être doué de conscience (appelé *cogniteur*) a un certain degré de croyance en la véracité d'une proposition. Ce domaine a été largement étendu pour inclure les cas où la position du cogniteur porte sur l'opportunité d'effectuer

une action et ceux où elle porte sur la légitimité d'un contenu ou d'une situation, ainsi que par extension de ce dernier cas, certains cadres de jugement.

Enfin, nous avons ajouté au French FrameNet un ensemble de macro-rôles qui permettent de généraliser sur les rôles à un niveau bien plus global que les relations d'identité entre rôles proposées en cas de relation entre cadres. Au nombre de 41, ces macro-rôles constituent un niveau intermédiaire de granularité, qui permet des généralisations plus larges que ce que permettent les rôles des cadres, mais qui restent bien plus spécifiques que des rôles génériques comme agent, patient ou thème.

9.2 Bilan

Nous dressons ici le bilan de la création du French FrameNet. Nous revenons en particulier sur nos raisons initiales pour choisir le modèle FrameNet pour le projet ASFALDA afin de vérifier si nos attentes quant aux avantages présentés par le modèle FrameNet correspondaient à la réalité.

Comme prévu, la ressource obtenue unifie le traitement de plusieurs catégories : nos déclencheurs sont des verbes, des noms, mais aussi des prépositions, conjonctions, adverbes et adjectifs. Cela est particulièrement vrai pour le domaine notionnel de la cause ; c'est d'ailleurs en partie pour son côté « transverse » qu'il fait partie des domaines traités par le French FrameNet.

Comme prévu aussi, nous obtenons à la fois des généralisations sur les lexèmes (les cadres) et sur les rôles sémantiques, à deux niveaux de granularité : avec des rôles spécifiques à un cadre et une quarantaine de « macro-rôles ».

En ce qui concerne l'avantage supposé du modèle FrameNet pour la transposition d'une langue à l'autre du fait de son orientation plus sémantique que les autres ressources, le bilan est plus mitigé. Nous avons constaté beaucoup d'incohérences dans les données BFN. Ces incohérences ne sont à notre sens pas seulement liées aux inévitables imperfections de telles données et inéluctables divergences entre guide d'annotation et annotation effective, mais sans doute également au flou dans la définition de certaines notions. Comme nous l'avons mentionné plus haut, nous avons finalement opté pour l'introduction de critères syntaxiques plus explicites. Cela a occasionné beaucoup plus de modifications de cadres que prévu, essentiellement pour des questions de périmètre des cadres ; nous avons par exemple fusionné de nombreux cadres. Le lien avec la ressource anglaise est cependant préservé, grâce à des relations que nous avons ajoutées entre cadres du French FrameNet et cadres BFN. Il reste que les annotations ont privilégié la sémantique en cas de divergences syntaxe-sémantique, ce qui, grâce aux annotations syntaxiques pré-existantes, donne une ressource intéressante pour l'étude de l'interface syntaxe-sémantique.

Quant à notre stratégie domaine par domaine, elle s'est avérée cruciale pour expliciter les frontières de cadres et limiter ainsi des incohérences et doublons de cadres. Elle présente de plus l'avantage d'aboutir à des données annotées qui respectent la distribution naturelle des phénomènes linguistiques dans les corpus visés, ce qui les rend utilisables comme données d'apprentissage pour un analyseur sémantique —restreint cela dit aux domaines notionnels couverts. La couverture du French FrameNet est réduite à ces quatre domaines, mais on constate que pour environ 500 lemmes, tous leurs sens présents dans nos corpus cibles sont couverts par nos cadres : toutes leurs occurrences (ou pour les lemmes fréquents, leur cent premières occurrences) sont annotées avec un des cadres de nos domaines (aucune annotation *Other_sense*).

9.3 Perspectives

Le travail présenté dans cette thèse peut être poursuivi dans plusieurs directions. Nous proposons ici quelques pistes d'une part pour l'amélioration ou l'extension du French FrameNet, et d'autre part pour son exploitation.

La couverture de la ressource est actuellement limitée à quatre domaines. Si augmenter substantiellement la couverture avec la même méthodologie n'est aujourd'hui pas d'actualité faute de moyens, un travail sur les occurrences actuellement notées comme « hors-domaine » serait plus accessible. Cela pourrait éventuellement être effectué avec des cadres simplifiés, et sans généralisations de lexèmes, à la manière des proto-cadres du projet SALSA. On augmenterait ainsi le nombre de lemmes dont tous les sens sont couverts, et pour lesquels on peut ainsi utiliser les données pour de la désambiguisation lexicale.

Ensuite, plusieurs sortes d'informations pourraient être ajoutées aux annotations actuelles du French FrameNet. Il serait par exemple intéressant de créer un système de traits sur les déclencheurs. Ces traits pourraient permettre de signaler certaines particularités que nous n'avons pas retenues comme critères de délimitation de cadres, mais qu'il serait intéressant —pour quelque raison— de connaître. Nous pensons notamment à un trait qui indiquerait le sens positif ou négatif des déclencheurs d'un cadre, puisque des verbes comme *savoir* et *ignorer* partagent par exemple le même cadre. Une autre piste serait, pour le domaine des positions cognitives, le statut présupposé ou non du rôle correspondant au contenu mental sur lequel porte la croyance du cogniteur. Ainsi, la différence de présupposition entre *savoir* et *penser* par exemple pourrait être capturée. À plus long terme, pour une utilisation de la ressource dans des applications réelles d'analyse sémantique, il nous semble que l'explicitation de la factualité des cadres annotés est cruciale.

Nous avons mentionné section 7.1.1.2, dans le chapitre qui traite de l'annotation, que nous avions utilisé la même stratégie que le FrameNet de Berkeley pour l'annotation des métaphores productives mettant en jeu un déclencheur de nos domaines. Lorsqu'une occurrence de déclencheur évoque un cadre de nos domaines mais que son sens est métaphorique, nous utilisons le cadre correspondant au sens littéral et ses rôles pour l'annotation, et signalons l'occurrence comme métaphorique à l'aide d'un marquage spécial. Cela permettra dans de futurs travaux d'utiliser ce marquage pour repérer les occurrences métaphoriques avant d'annoter leur cadre métaphorique en plus de leur cadre littéral, à la manière du projet SALSA.

Nous avons proposé une hiérarchie de macro-rôles (section 8.3), permettant des généralisations plus larges que les rôles des cadres, tout en restant plus spécifiques que des rôles sémantiques classiques. Il serait à présent intéressant d'établir une correspondance entre nos macro-rôles et une liste de ces « rôles sémantiques traditionnels ». De futurs travaux pourraient en particulier lier nos macro-rôles aux rôles proposés par VerbNet et par Verb∋Net, le VerbNet du français de Pradet (2015); Danlos *et al.* (2016) [2]. Ainsi, il serait possible, d'établir une correspondance entre nos rôles et les rôles Verb∋Net, et sous réserve d'établir une correspondance cadre-à-classe comme celle proposée par le projet SemLink (Palmer, 2009) pour VerbNet, de convertir les annotations du French FrameNet en annotations pour les classes Verb∋Net. Une telle conversion présente un triple intérêt.

— D'abord, il nous semble intéressant de comparer nos macro-rôles, obtenus sur base sémantique, et les rôles Verb∋Net. Rappelons que le modèle VerbNet a une orientation bien plus syntaxique que le modèle FrameNet : nous pourrions donc quantifier les cas où l'orientation syntaxique diverge de l'orientation sémantique. Le cas typique concerne les rôles de verbes converses : ainsi on utilise un seul macro-rôle Money_recipient pour le sujet de *vendre*, l'objet indirect d'*acheter*, et certains objets directs de *rembourser*. À l'inverse, Verb∋Net colle plus à la syntaxe et attribue les rôles d'Agent au sujet de *vendre* et de Source à l'objet indirect d'*acheter* (la construction de *rembourser* qui nous intéresse n'apparaît pour l'instant pas dans Verb∋Net). De même pour les constructions à attribut de l'objet : à l'inverse de Verb∋Net, le French FrameNet n'assigne pas de rôle à l'objet direct, qui n'est pas un argument sémantique du verbe. Il apparaît intéressant de quantifier ces cas de divergences.

— Le deuxième attrait d'une conversion des rôles annotés dans le French FrameNet en rôles

2. Danlos *et al.* (2016) indiquent que Verb∋Net utilise la même liste de rôles que VerbNet à quelques exceptions près : certains rôles prévus pour l'anglais ne peuvent être réalisés en français. Les rôles Verb∋Net constituent en fait un sous-ensemble des rôles VerbNet.

Verb∋Net est le passage à une granularité encore moins fine que nos macro-rôles (modulo les cas de divergences cités au premier point). VerbNet compte 32 rôles dont Verb∋Net semble utiliser 28. La différence ne paraît pas grande avec notre quarantaine de macro-rôles, mais il ne faut pas oublier que l'ensemble de rôles VerbNet se veut applicable à tout verbe (bien que cette applicabilité n'ait jamais été testée par une application à grande échelle) tandis que nos macro-rôles sont très dépendants de nos quatre domaines notionnels, auxquels ils sont restreints. Le passage des annotations du French FrameNet aux rôles VerbNet permettrait donc une bien plus grande généralisation sur les remplisseurs de rôles annotés.

— Enfin, une telle conversion permettrait d'obtenir des exemples en corpus annotés en rôles pour les classes Verb∋Net, ce qui n'existe encore pour aucun VerbNet.

Notons aussi que pour ce qui est de la visualisation des données, certaines améliorations doivent être apportées. Ainsi, actuellement, les macro-rôles ne sont visualisables ni dans les cadres, ni dans les annotations.

Enfin, un projet du nom de *Multilingual FrameNet : A Resource Enabling Cross-Lingual Research for the Natural Language Processing Community*[3], mené par Collin Baker, vient de démarrer. Ce projet prévoit entre autres l'établissement d'une correspondance en cadres et en rôles entre les projets FrameNet déjà développés pour différentes langues, dont le French FrameNet. Les relations spéciales que nous avons créées entre les cadres du French FrameNet et ceux du FrameNet de Berkeley nous permettront de plus aisément prendre part à ce projet.

Pour ce qui est de l'exploitation de la ressource, nos choix d'annotation permettent par exemple au French FrameNet d'être utilisé pour le travail sur la résolution d'anaphores : lorsqu'un remplisseur de rôle est anaphorique et que son antécédent se trouve dans la même phrase, les deux éléments sont annotés comme remplisseur et signalés à l'aide d'un marquage spécial pour distinguer l'élément anaphorique et son antécédent. L'ensemble de ces occurrences, c'est-à-dire plus de 1 800 paires anaphore-antécédent, peut être utilisé comme un corpus d'anaphores.

Le French FrameNet peut être utilisé comme données d'apprentissage d'analyseurs en cadres et rôles (cf. premiers travaux en ce sens de Michalon (2014) et Michalon *et al.* (2016)). En termes de tâches typiques de TAL, l'analyse de type FrameNet correspond à de la désambiguïsation lexicale et de l'étiquetage en rôles sémantiques. Plus précisément sur la partie désambiguisation lexicale, comme certaines occurrences ont simplement été marquées comme « hors domaine », la tâche de désambiguisation lexicale est totalement réalisée pour les lemmes n'ayant aucune annotation de cadre « hors domaine » (c'est-à-dire pour 490 lemmes sur 873). Pour les autres lemmes, la désambiguisation n'est que partielle.

Enfin, le French FrameNet peut être utilisé pour l'étude de l'interface syntaxe-sémantique, ce qui est depuis le début du projet un des buts visés. Les travaux de Candito *et al.* (2014b) et Ribeyre *et al.* (2014) nous ont donné accès à la syntaxe profonde du Sequoia Treebank et du French Treebank, sur lesquels est annoté le French FrameNet. Il s'agit d'une représentation de la syntaxe canonique, c'est-à-dire qui s'abstrait des variations syntaxiques comme les variations de diathèse, de chacune des phrases de ces corpus. Entre cette représentation et les annotations en cadres et en rôles de notre ressource, nous disposons actuellement de tous les outils nécessaires pour extraire un lexique quantifié des réalisations canoniques des rôles pour chaque déclencheur annoté du French FrameNet. Il s'agit pour chaque occurrence de cadre d'extraire la ou les fonctions grammaticales canoniques de chaque rôle du cadre, c'est-à-dire les fonctions sans changements de diathèse[4]. On peut alors extraire un lexique de « linkings », correspondances entre cadres de sous-catégorisation canonique et rôles sémantiques, avec leurs fréquences en corpus. Notre prochaine étape est donc de procéder à cette extraction et d'en étudier le résultat.

3. `https://www.nsf.gov/awardsearch/showAward?AWD_ID=1629989`

4. Pour cette extraction, nous ignorons les réalisations non locales, ainsi que les remplisseurs FULLANTECEDENT. En effet, les remplisseurs de rôles ne sont marqués comme antécédents que lorsque le rôle est aussi réalisé par une anaphore plus locale au déclencheur que l'antécédent.

ANNEXE A

EXEMPLES DE GUIDES DE POLYSÉMIE

Nous avons mentionné au chapitre 5 que pour certains déclencheurs s'avérant difficiles à désambiguïser, les responsables des domaines avaient écrit des guides informels de désambiguïsation à destination des annotatrices. Ces documents, dits *guides de polysémie*, intègrent généralement des définitions, des tests de désambiguïsation et parfois des règles par défaut. Nous présentons ici à titre d'exemple cinq de ces guides de polysémie.

Cadres	**Polysème(s)**	**Guide de désambiguïsation**
FR_ATTEMPT_SUASION.VERACITY, FR_JUSTIFYING	arguer.v	✓ FR_JUSTIFYING : Xhum argue Y[que P / de NP / de ce que P] pour expliquer un effet avéré Z — *Les hommes_politiques observent le silence en ARGUANT de la seule compétence du CSA [...]* — *[...] la Commission des opérations_de_bourse (COB) retardait l'octroi du visa de l'OPA, ARGUANT du flou régnant autour_de la structure du capital de Perrier.* — *Car le groupe anglo_-_néerlandais, ARGUANT de ses besoins futurs, rechigne à garantir la libre utilisation de la conduite pétrolière.* — *[...] qui a arrêté le financement des émissions auquel il contribuait depuis 1974 en ARGUANT qu'elles n'étaient pas adaptées aux exigences du jeune public.* — *Certains d'entre eux ont alors rechigné, ARGUANT que la BFCE était devenue peu_à_peu une banque concurrente.* ✓ FR_ATTEMPT_SUASION.VERACITY : Xhum argue Y[que P] : donne arguments (non exprimés) en faveur du Y : une seule occurrence — *[...] consiste à faire_état de prix relativement stables pendant les derniers mois pour ARGUER qu'il est temps de baisser les taux_d'_intérêt.*
FR_AWARENESS-CERTAINTY-OPINION	considérer.v	On annote les sens : ✓ Être d'avis après réflexion que ; estimer, croire — *Nous avons CONSIDÉRÉ que l'âge de 14 ans est une solution adéquate.* ✓ Juger quelque chose, quelqu'un, lui attribuer telle ou telle qualité — *L'enquête sur sa mort a été classée sans suite, étant CONSIDÉRÉE comme accidentelle.* On n'annote pas : ✓ Observer/envisager quelque chose sous un certain angle ; peser quelque chose, l'apprécier, le prendre en compte — *Il est essentiel de CONSIDÉRER les intérêts de toutes les parties.*

Cadres	Polysème(s)	Guide de désambiguïsation
FR_BECOMING_AWARE, FR_STATEMENT-MANNER-NOISE	constater.v	On rappelle la procédure de désambiguïsation entre cadre cognitif et cadre communicationnel : Dans le cas d'un verbe de position cognitive pour lequel on peut construire des cas clairement communicationnels (et donc associés à un cadre communicationnel). ⇒ on décide que la présence seule de guillemets est insuffisante pour désambiguïser entre cadre communicationnel ou pas. En effet, les guillemets peuvent à eux seuls indiquer au lecteur un discours rapporté : *Alex a vécu "la meilleure expérience de sa vie"* ⇒ on comprend qu'Alex a dit ça, mais la prédication principale (*a vécu*) ne porte pas sur la communication, et d'ailleurs les deux évènements sont disjoints dans le temps : *vécu de l'expérience* et *communication sur le vécu de l'expérience*. Pour désambiguïser entre cadre cognitif et cadre communicationnel : ✓ Voir si présence d'indices fort dans le contexte en faveur d'une communication. ✓ Voir quelle glose est la meilleure (entre une glose avec verbe communicationnel ou bien verbe uniquement cognitif). ✓ Dans le cas d'un cadre cognitif statif, l'aspect peut permettre de désambiguïser. ⇒ L'aspect ponctuel n'est possible que pour la communication. ⇒ En cas de doute, privilégier le cadre cognitif (qui est toujours vrai, même en cas de communication).
FR_MEMORY-REMEMBERING_EXPERIENCE-REMEMBERING_INFORMATION, REMEMBERING_TO_DO	oublier.v se souvenir.v se rappeler.v	REMEMBERING_TO_DO est réservé aux cas où ce qui est réactivé mentalement est une action. ✓ "V + de + Vinf" ⇒ REMEMBERING_TO_DO ✓ "V + de + NP" ⇒ FR_MEMORY-REMEMBERING_EXPERIENCE-REMEMBERING_INFORMATION ✓ "V + que P" ⇒ FR_MEMORY-REMEMBERING_EXPERIENCE-REMEMBERING_INFORMATION, même si le NP ou le que P contient l'idée d'une action
EVOKING, FR_STATEMENT-MANNER-NOISE	rappeler.v	✓ Rappeler quelqu'un à l'ordre : Judgment_direct_address, ✓ Se rappeler à l'attention de : hors champ (on comprend que le sujet est volitionnel dans ce cas, donc ne rentre pas dans EVOKING ni FR_STATEMENT-MANNER-NOISE), ✓ $X_{\text{non volitionnel}}$ rappelle Y à Z_{sentient} ⇒ EVOKING — *Ceci me RAPPELLE mon enfance.* — *Cette odeur RAPPELAIT à Pierre le bonheur perdu.* — *Cela RAPPELLE le film l'arnaque.* — *Le sens de la démarche RAPPELLE de fâcheux précédents.* ✓ X_{sentient} rappelle Y à Z_{sentient} ⇒ normalement ambigu : — Si X_{sentient} n'est pas volitionnel (*Paul me rappelle mon enfance*) ⇒ EVOKING — Si X_{sentient} est volitionnel, on a l'idée de communication volontaire de X_{sentient} ⇒ FR_STATEMENT-MANNER-NOISE — *Vous avez RAPPELÉ, Monsieur le Président, les valeurs auxquelles vous teniez.* — *M. Dumas a RAPPELÉ la décision "rapide, claire, ferme", prise par les_Onze.* — *Les entreprises, RAPPELONS-le, en arrivent maintenant à payer leur découvert à un taux réel.*

Annexe B

Guide d'annotation

ASFALDA French FrameNet - Guide d'annotation

Marie Candito et Marianne Djemaa

Dernière révision le 26 février 2017
https://sites.google.com/site/anrasfalda/

Contents

1 Contexte et vocabulaire **3**

2 Cycle d'annotation **4**
2.1 Pré-annotation automatique 4
2.1.1 Pré-annotation des déclencheurs composés 5
2.1.2 Frame spécial Other_sense 5
2.1.3 Préannotation des verbes pronominaux 5
2.2 Double annotation 6
2.3 Adjudication 7
2.4 Post-traitement automatique 7
2.4.1 Tête syntaxique 7
2.4.2 Tête sémantique 8
2.5 Demande d'évolution du lexique 8
2.6 Accès au contexte de la phrase à annoter 8

3 Annotation d'un frame: principes généraux **8**
3.1 Dans quel cas annoter une instance de frame? 8
3.1.1 Annotation indépendante de la factivité 8
3.1.2 Déclencheur avec sens générique / déclencheur dans entité nommée . 8
3.2 Alternances syntaxiques 9
3.3 Emplois métaphoriques / figurés 10
3.4 Métonymie 11
3.5 Redondance de déclencheurs ou pléonasmes 12

4 Annotation des remplisseurs: principes généraux **12**
4.1 Quels sont les rôles à annoter ? 12
4.2 Utilisation de l'annotation en syntaxe 13
4.2.1 Gestion des erreurs dans l'annotation syntaxique / divergences syntaxe-sémantique 13
4.3 Mots fonctionnels introduisant des remplisseurs 14
4.3.1 Prépositions et complémenteurs vides 14
4.3.2 Coordination / énumération de prépositions régies 15
4.3.3 Coordination de préposition composée 15
4.3.4 Coordination de prépositions qui correspond sémantiquement à une ellipse de prédicat 15
4.4 Localité des remplisseurs 16
4.4.1 NLI: remplisseurs non-locaux 16

4.4.2 Imbrication de remplisseurs . . . 18
4.4.3 DNI : Definite null instantiation . . . 18
4.4.4 ENI : Extra-linguistic null instantiation . . . 18
4.4.5 UNI : Unknown null instantiation . . . 19
4.5 Remplisseurs anaphoriques . . . 19
4.5.1 Relatives . . . 21
4.5.2 Réfléchis . . . 21
4.5.3 Possessifs . . . 21
4.5.4 Cas particuliers: "en" quantitatif / démonstratif "celui" . . . 22
4.5.5 Aspects techniques : 2 instances de remplisseurs, versus un remplisseur "multi-pattes" . . . 22

5 Points spécifiques aux déclencheurs verbaux 23
5.1 Ellipses de verbes . . . 23
5.1.1 Head Gapping : coordination, où le verbe n'est pas répété (mais 2 sujets différents) . . . 23
5.1.2 Argument cluster (i.e. "coordination de séquences de constituants") 24
5.1.3 Comparatives réduites . . . 25

6 Points spécifiques aux déclencheurs nominaux 25
6.1 Nominalisation référant à un évènement versus nominalisation référant à un résultat de l'évènement . . . 26
6.1.1 Critères de distinction entre référence à une éventualité versus référence à un participant ou résultat . . . 26
6.2 Noms relationnels et frames représentant une relation . . . 29
6.3 Usage prédicatif versus usage référentiel d'un déclencheur nominal . . . 30
6.3.1 Liste non exhaustive des types d'emplois prédicatifs (où le nom déclencheur n'est pas en outre remplisseur) . . . 31
6.3.2 Exemples d'emplois référentiels . . . 33
6.4 Noms quantificateurs . . . 33
6.5 Constructions à verbe support . . . 33

7 Annotation des remplisseurs: points divers 33
7.1 Remplisseurs collectifs versus rôle _1 / rôle _2 . . . 33
7.2 Astuce dans le cas de déclencheurs qui dépendent syntaxiquement d'un de leur arguments sémantiques . . . 34
7.2.1 Adjectifs et participes épithètes . . . 34
7.2.2 Incises . . . 35
7.3 Remplisseur adjectival synonyme d'un PP . . . 36

8 Phénomènes de portée 36
8.1 remplisseurs phrastiques/infinitivaux: gestion des verbes introducteurs de complétives/d'infinitives . . . 36
8.1.1 Portée sémantique sur verbe enchâssé versus portée sur verbe introducteur . . . 37
8.1.2 Difficultés de détermination de portée . . . 37
8.1.3 Gestion dans Asfalda : classes de verbes introducteurs . . . 38
8.2 Portée de la négation et des modaux dans un remplisseur . . . 40

9 Saturation des remplisseurs prédicatifs et divergence entre tête syntaxique et tête sémantique **40**
9.1 Saturation des remplisseurs infinitivaux 41
9.2 Saturation des remplisseurs adjectivaux 41
9.2.1 Remplisseur adjectif ou participe épithète et divergence entre tête syntaxique et sémantique 42

10 Bibliographie **43**

1 Contexte et vocabulaire

Le projet ASFALDA (projet ANR-12-CORD-023) vise à produire un "FrameNet du français", sur le modèle de FrameNet (Baker et al., 1998) développé pour l'anglais. Ce guide suppose une certaine familiarité avec les concepts de base de FrameNet, voir en particulier (Fillmore, 2007).

- frame : un frame est une représentation schématique d'une éventualité (au sens évènement ou état) ou d'une entité. Plus précisément, une représentation schématique de la connaissance qu'ont les locuteurs d'une éventualité ou entité, sous-tendant le sens des unités lexicales (Fillmore, 2007). Un frame est constitué d'une définition fixant son périmètre sémantique du frame, d'une liste de **déclencheurs** (des unités lexicales pouvant évoquer le frame), et d'une liste de **rôles**. Dans l'exemple précédent, on considère que

- rôle[1] : les rôles sont des composants nommés d'un frame, et correspondent aux participants et propriétés des situations représentées par le frame. Un rôle a un nom (en anglais), une définition, et est spécifique à un frame. Voir section 4.1 pour une précision du lien entre les rôles et la valence syntaxique des déclencheurs.

- déclencheur[2]: un déclencheur est un mot (simple ou composé) pouvant évoquer un frame donné. Par exemple le verbe *convaincre* peut évoquer le frame FR_Cognizer_affecting.veracity : "un Cognizer est persuadé de la véracité d'un certain Content, soit du fait d'un State_of_affairs soit parce qu'un Persuader humain l'en persuade par la parôle".

- remplisseur (de rôle): il s'agit en contexte d'une expression linguistique qui instancie un rôle donné. Par exemple dans *M. Fitz a convaincu les députés que l'identité du témoin ne devait pas être révélée.*, le verbe *convaincre* déclenche le frame FR_Cognizer_affecting.veracity, et les expressions *"M. Fitz"*, *"les députés"* et *"l'identité du témoin ne devait pas être révélée"* remplissent respectivement les rôles Persuader, Cognizer et Content.

Le FrameNet du français visé comprendra :

- **une structure de frames**: un ensemble de frames, reliés par des relations entre frames. Un frame décrit une éventualité (état ou évènement) prototypique ou une entité. Il est constitué d'une définition fixant son périmètre sémantique du frame, et d'une liste de rôles correspondant

[1] Appelé "frame element" dans la littérature framenet.
[2] Appelé "frame evoking element" dans la littérature framenet.

- **un lexique**: pour chacun des frames, les lemmes (français) pouvant évoquer ces frames.
- des **annotations** de frames et rôles en corpus

Dans une première phase du projet (Candito et al., 2014), les 2 premières parties ont été constituées:

- des domaines notionnels ont été retenus, et les frames relevant de ces domaines ont été définis (en partant des frames existant pour l'anglais). On parlera ci-après des "frames Asfalda".
- le lexique pour les frames Asfalda a été mis au point, domaine par domaine. Le travail a été manuellement réalisé par des "experts" par domaine, en partant d'une souche automatiquement obtenue du lexique anglais.

Le présent guide concerne la tâche subséquente d'annotation de ces frames en corpus. Même si le travail manuel sur le lexique a été fait en ayant accès à un concordancier pour visualiser les occurrences dans le corpus visé, il est prévu que le lexique évolue au cours de la phase d'annotation, mais toujours après validation par les "experts" de chaque domaine.
Le corpus retenu est la concaténation de deux corpus arborés (i.e. ayant fait l'objet d'une annotation syntaxique):

- le Sequoia treebank (Candito and Seddah, 2012) (dans sa version en dépendances, version 7.0, 3099 phrases)
- le French Treebank (Abeillé and Barrier, 2004) (dans sa version en dépendances, telle que livrée à l'occasion de la shared task SPMRL 2013, 18535 phrases (Seddah et al., 2013))

Dans tout le projet, l'**identifiant d'une phrase** est son rang dans la concaténation de ces deux corpus.

2 Cycle d'annotation

Le cycle d'annotation comprend une pré-annotation automatique, une double annotation manuelle, une adjudication, et un post-traitement automatique.

2.1 Pré-annotation automatique

Les occurrences de déclencheur à annoter sont repérées par un pré-annotateur automatique, qui assigne à chaque occurrence tous les frames possibles pour le déclencheur en question. La tâche manuelle restante est de choisir le bon frame et d'annoter ses rôles.
Etant donnée la difficulté à maîtriser le contenu des différents frames, et de certaines frontières entre frames, on a choisi de ne pas annoter linéairement le corpus: des groupes de déclencheurs ont été constitués, et l'outil produit des fichiers pré-annotés pour un groupe de déclencheurs bien précis. Les groupes de déclencheurs sont définis automatiquement, en utilisant plusieurs heuristiques:

- on groupe en priorité les déclencheurs ayant la même ambiguïté de frames (ils sont dans le lexique associés au même ensemble de frames)
- on groupe un verbe pronominal et le verbe non pronominal correspondant

- on groupe un verbe et ses nominalisations éventuelles
- pour un déclencheur composé, si un de ses composants est également un déclencheur, il est ajouté au groupe.

Pour un groupe de déclencheurs ainsi automatiquement constitué, le pré-annotateur produit un ou plusieurs fichiers pré-annotés, en limitant la taille à 100 phrases. Pour un déclencheur ayant plus de 100 occurrences, on n'annotera que le premier fichier produit (extension "PRIMO", les autres fichiers auront une extension "EXTRA", à ne pas annoter).

Important: lorsque l'on travaille sur un fichier pré-annoté, il faut avoir en tête le groupe de déclencheurs "cibles" de ce fichier (cf. tableur), qui sont ceux automatiquement pré-annotés dans ce fichier. Seuls ces déclencheurs doivent recevoir un frame lors de l'annotation du fichier. Il se peut que des déclencheurs synonymes apparaissent dans les phrases du fichier, mais ne soient pas pré-annotés s'il ne sont pas un des déclencheurs cibles du fichier. Ces synonymes, s'ils sont bien dans le lexique courant, donneront lieu à leurs propres fichiers pré-annotés ($\longrightarrow$ les occurrences du même frame mais pour d'autres déclencheurs sont proposées dans un autre fichier).

2.1.1 Pré-annotation des déclencheurs composés

La pré-annotation des mots composés est conçue pour être tolérante aux incohérences d'annotation de composés dans les treebanks sous-jacents. Ainsi, on pré-annote systématiquement "en effet" avec le frame Evidence, que la séquence "en"+"effet" soit ou pas annotée comme un composé dans le treebank.
En outre, des heuristiques déclenchent la pré-annotation d'occurrences potentielles de composés présentant des discontinuités : la pré-annotation permet des discontinuités avant certains composants de composés (à , de, où ...)
Cela permet par exemple de capturer le déclencheur "au moment où", dans des occurrences comme:

(1) ... ce qu' on pourrait attendre du parti au pouvoir en France , **au moment** précis **où** la révolution de 1989 ...

(2) Cela **au moment** même **où** la " révolution " des pays de l' est ...

Cela occasionne certainement du bruit ($\longrightarrow$ il faut dans ces cas simplement enlever les frames proposés, cf. section 2.2), sans doute un peu de silence.

2.1.2 Frame spécial Other_sense

On utilise un "frame" spécial: Other_sense afin de pouvoir faire la différence entre

- les occurrences de déclencheur examinées lors de l'annotation manuelle, mais pour lesquelles il a été jugé que l'occurrence ne correspondait à aucun frame Asfalda $\longrightarrow$ (l'occurrence est alors annotée avec le frame Other_sense)
- et les occurrences de déclencheurs sur lesquelles aucun annotateur n'est passé (et donc aucun frame annoté)

Les déclencheurs préannotés le sont aussi systématiquement avec ce frame Other_sense.

2.1.3 Préannotation des verbes pronominaux

Le clitique réfléchi *se* peut correspondre à différents processus syntaxiques ou lexicaux, ce qui fait l'objet d'une abondante littérature car les distinctions de statut sont souvent difficiles à faire. On distingue schématiquement:

- le vrai réfléchi (*Jean se parle*)
- le vrai réciproque (*Paul et Marie se parlent*)
- le se-moyen (*Ce genre de livre, ça se vend dans les gares.*), dont la caractéristique est que l'agent n'est pas réalisable localement, mais est forcément interprété comme existant (dans l'exemple, on ne peut pas interpréter la phrase sans faire intervenir un agent qui vend ce genre de livre)
- le se-neutre (*Tout d'un coup, le vase s'est fendu*), ici au contraire l'agent n'est pas forcément existant
- le verbe dit intrinsèquement pronominal, i.e. soit pour lequel une version non pronominale n'existe pas (ex: *se suicider*) ou bien pour lequel on ne repère pas en diachronie de relation régulière à une version non pronominale (ex: *s'apercevoir*)
- les verbes dits coopératifs (*Paul et Virginie se marient*), où la version Xcollectif se V alterne avec "X se V à/avec Y".

Cette catégorisation est importante pour l'annotation, car selon les cas, on considère que le déclencheur est le verbe avec ou sans le clitique *se*.

- dans les cas vrai réfléchi, vrai réciproque, se-moyen, on considère que le *se* marque un changement de diathèse régulier, sans modification du nombre des arguments sémantiques. Le déclencheur à annoter est le verbe sans le *se*.
- Alors que pour les cas se-neutre et intrinsèquement pronominaux, le déclencheur à annoter inclut le clitique *se*. Les frames possibles pour le se+V, recensés dans le lexique, ne sont en général pas les mêmes que les frames du V (voir aussi la section 3.2).

Parce que la distinction entre ces différentes catégories de *se* est souvent difficile à faire, elle a fait l'objet d'une annotation manuelle en amont du projet ASFALDA. La pré-annotation automatique utilise la catégorisation des occurrences de *se* pour préannoter les frames pertinents pour selon les cas le se+V ou le V seul.

2.2 Double annotation

Pour un fichier pré-annoté donné, on vise de le faire annoter par 2 personnes indépendamment, puis de faire une adjudication des conflits entre les 2 résultats, via l'outil Salto (Burchardt et al., 2006) qui repère automatiquement les différences d'annotation.
La pré-annotation génère des tableaux recensant des fichiers à annoter, accessibles en ligne. Un tableau est assigné à un ensemble d'annotateurs. On produit les fichiers de façon à traiter:

- d'abord les déclencheurs associés à un seul frame,
- puis les déclencheurs associés à plusieurs frames mais tous du même domaine

- enfin les déclencheurs ambigus entre frames de différents domaines

Les annotateurs choisissent leur fichier à annoter n'ayant pas déjà 2 annotateurs signalés et notent tout de suite la date de début de travail sur ce fichier, dans la colonne prévue à cet effet (pour éviter d'annoter un fichier déjà pris par 2 annotateurs). Un annotateur AAA travaille ensuite sur un fichier XXX à annoter en réalisant des sauvegardes sous le svn, sous $ASFALDASVN/annotation/out/AAA/XXX.salto.xml. Lorsqu'il a terminé un fichier, il le signale dans le tableau, dans la colonne prévue.
Pour une occurrence de déclencheur (simple ou composé, et si composé, continu ou discontinu) pré-annotée comme évoquant un ou plusieurs frames, 3 possibilités:

1. Il s'agit d'une erreur de pré-annotation, i.e. il ne s'agit pas d'une occurrence de déclencheur cible
   - il peut s'agir d'une séquence pré-annotée comme déclencheur composé mais ne correspondant pas au composé en contexte (par exemple dans *"les assurances du gouvernement sur l' emploi ne trouvent pas grâce à leurs yeux"*, *"grâce à"* sera préannoté à tort
   - ou inversement, un composant de composé peut avoir été pré-annoté avec ses propres frames: par exemple *rapport* dans *par rapport à*
   - $\longrightarrow$ dans ces deux cas, retirer tous les frames proposés (y compris Other_sense)
2. Il s'agit bien d'une occurrence de ce déclencheur qui déclenche bien un des frames proposés $\longrightarrow$ garder uniquement ce frame et annoter ses rôles
3. Il s'agit bien d'une occurrence du déclencheur mais ce dernier n'évoque aucun des frames proposés $\longrightarrow$ enlever tous les frames proposés sauf Other_sense

Important: Avant de commencer l'annotation d'un fichier, lisez bien:

- la documentation des frames pré-annotés dans le fichier, et les indications de frontières entre frames
- les éventuels guides de polysémie pour les triggers du fichier: des guides de polysémie ont été écrits pour une bonne partie des déclencheurs ambigus (cf. fichier en ligne Polysémie). Si un déclencheur n'a pas un tel guide, et que son annotation se révèle trop difficile, suspendez l'annotation et provoquez une discussion sur le wiki pour mettre au point un guide de polysémie pour ce déclencheur.

2.3 Adjudication

L'adjudication est faite via Salto. Voir le guide technique d'adjudication sous Salto.

2.4 Post-traitement automatique

Pour accélérer l'annotation des remplisseurs, en particulier permettre de n'avoir à pointer qu'un seul noeud syntaxique pour un remplisseur, on utilise un post-traitement automatique pour supprimer d'un remplisseur certaines parties. Le traitement n'est pas détaillé ici car il nécessite d'avoir introduit d'autres notions. Reportez vous à la section 7.2.

2.4.1 Tête syntaxique

Le post-traitement se charge également de calculer la tête syntaxique des remplisseurs de rôles. On utilise l'annotation syntaxique sous-jacente pour cela, plus une règle de priorité dans le cas de remplisseurs discontinus formés de plusieurs sous-arbres syntaxiques non connectés (on choisit l'attribut le plus à gauche, et à défaut l'adjectif, verbe, nom adverbe ou pronom le plus à gauche, dans cet ordre de priorité).
Ainsi par exemple dans:

(3) Paul **dit** être parti à 23h

On considère que le rôle Message est rempli par "Paul" + "être parti à 23h" (cf. section 9). Le remplisseur est formé de 2 sous-arbres syntaxiques déconnectés, de tête nom et verbe $\longrightarrow$ On choisit verbe comme tête syntaxique du tout.

2.4.2 Tête sémantique

Par défaut, la tête syntaxique sera interprétable dans les données comme étant la tête sémantique. On a cependant repéré des cas de divergence, avec annotation manuelle de la tête sémantique (section 9.2.1). On obtient après post-traitement que la tête sémantique d'un remplisseur est celle marquée si une telle marque existe, et sinon est la tête syntaxique.

2.5 Demande d'évolution du lexique

Le lexique a été construit en général en étudiant les premières occurrences des déclencheurs (utilisation du concordancier sur le corpus), mais il se peut que l'annotation effective révèle des problèmes de lexique (déclencheur associé à tort à un frame, ou bien frame manquant pour un déclencheur donné).
Si pour un déclencheur pré-annoté, vous voudriez annoter avec un frame Asfalda existant, mais qui n'a pas été prévu pour ce déclencheur $\longrightarrow$ provoquez une discussion à ce sujet dans le wiki. S'il s'avère après discussion qu'il s'agit d'un oubli, le frame sera ajouté à ce déclencheur dans le lexique, mais vous devrez manuellement ajouter le frame dans le fichier en cours d'annotation, avec tous les rôles adéquats.

2.6 Accès au contexte de la phrase à annoter

Les phrases à annoter dans les fichiers proposés sont regroupées par déclencheur ou groupe de déclencheurs et ne sont donc pas du texte continu.
Si la phrase est insuffisante pour une bonne compréhension, allez voir le texte dans le fichier $ASFALDASVN/corpus/sequoia+ftb.id_raw_articlefrontiers, en utilisant les id de phrases pour retrouver la phrase et son contexte. Le fichier contient des marques de découpage du corpus en articles, ce qui permet de mieux cerner le contexte d'une phrase.

3 Annotation d'un frame: principes généraux

3.1 Dans quel cas annoter une instance de frame?

3.1.1 Annotation indépendante de la factivité

On annote indépendamment de la réalisation effective de l'éventualité évoquée : on annote une instance de frame dès lors que l'idée du frame est évoquée, même dans le cas d'une

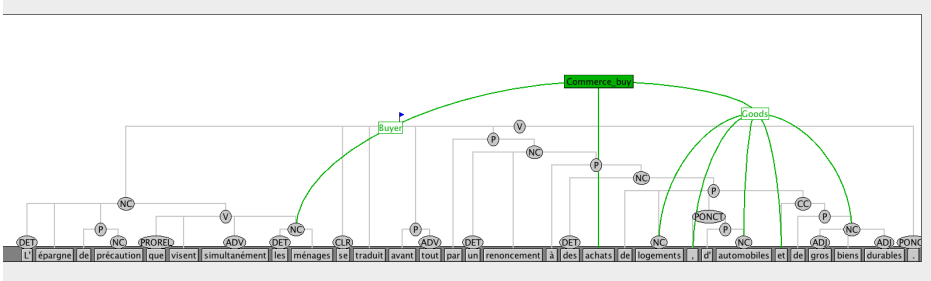


Figure 1:

éventualité hypothétique, future, niée etc... Il faut donc annoter indifféremment selon que l'on a du passé/présent/futur, du conditionnel, ou des prédicats sous négation.
Exemple : dans l'exemple suivant, les achats ne sont pas réalisés (cf. "renoncement") mais l'idée d'achat est bien évoquée (rem: les ménages sont NLI (section 4.4.1); à noter aussi l'annotation d'un remplisseur coordonné avec prépositions régies exclues du remplisseur, cf. section 4.3.2)

3.1.2 Déclencheur avec sens générique / déclencheur dans entité nommée

Pour décider si on annote une instance d'un frame X sur une occurrence de déclencheur Y, il faut se demander si Y dans ce contexte évoque bien l'idée de X.
A noter, qu'il peut s'agir d'une évocation générique: dans l'exemple ci-dessous, on a un acte d'achat générique, mais qui évoque bien l'idée d'un Commerce_buy.

(4) Un **achat** est toujours source de stress.

Ce raisonnement amène à annoter des instances de frame y compris au sein d'entités nommées, pour peu qu'elles aient un sens au moins partiellement compositionnel: dans *"le ministre du commerce et de l'industrie"*, *"commerce"* évoque bien l'idée de Commercial_scenario.

3.2 Alternances syntaxiques

On distingue les changements de diathèse qui ne modifient pas le nombre d'arguments sémantiques du verbe et les autres. En effet, un principe FrameNet fondamental est que pour être regroupés sous le même frame, des déclencheurs doivent avoir le même nombre et type de rôles sémantiques (ou le même nombre de paquets de rôles, lorsqu'il y a des paquets de rôles en exclusion mutuelle). Les alternances qui n'enlèvent ni n'ajoutent d'argument sémantique sont considérées comme des variantes dans la correspondance entre arguments sémantiques et arguments syntaxiques, et annotées avec le même frame.
C'est le cas par exemple pour le passif, alternance la plus fréquente. Il est important de noter que l'on parle ici de l'interprétabilité de l'existence d'un argument, et pas de sa réalisation effective. On traite donc de la même manière une phrase à l'actif, au passif avec agent et au passif sans agent. Le cas sans agent est traité comme les autres cas d'omission (cf. les flags de "null instantiation" section 4.4).

(5) [La France]$_{\text{Seller}}$ a **vendu** [les frégates]$_{\text{Goods}}$ à [Taïwan]$_{\text{Buyer}}$.

(6) [Les frégates]$_{\text{Goods}}$ ont été **vendues** à [Taïwan]$_{\text{Buyer}}$ (par la [France]$_{\text{Seller}}$).

Le clitique réfléchi se a des fonctions diverses, certaines relevant de changements de diathèse. On rappelle qu'à ce stade de l'annotation, la désambiguisation du se a déjà été faite, et les frames pertinents sont proposés en fonction des cas (cf. section 2.1.3).
Pour le se-moyen, le réfléchi, le réciproque, on utilise les mêmes frames pour la version avec ou sans se. La caractéristique du se-moyen est en effet que l'actant sujet de la version transitive n'est pas réalisable dans la version se-moyen, mais est interprétable. Dans l'exemple ci-dessous, il faut bien interpréter un Cognizer anticipant la spéculation, que l'on marque comme UNI (unknown null instantiation).

(7) 12969. [La spéculation]$_{\text{Phenomenon}}$, il est vrai , ne s' **anticipe** pas. Cognizer=UNI

(8) 12969'. *La spéculation , il est vrai , ne s' **anticipe** pas par les agents.

Pour le réfléchi ou réciproque, on considère schématiquement que le sujet du verbe remplit deux rôles. Le clitique se est considéré comme un marqueur grammatical et ne porte pas de rôle.

(9) 4688. [Le marché]$_{\text{Cognizer+Persuader}}$ commençait à se **convaincre** d' [une tendance haussière durable]$_{\text{Content}}$.

Les verbes "intrinsèquement pronominaux" (soit que la forme sans se n'existe pas (se suicider), ou ait un sens très différent de la forme pronominal (s'apercevoir de qqch versus apercevoir qqch)) constituent des entrées de lexique se+V distinctes du V, avec des frames en général différents. On a la même situation dans le cas du se-neutre (en effet, la caractéristique du se-neutre est que l'actant sujet de la version transitive n'est pas sémantiquement interprétable dans la version neutre, on a bien un actant en moins et donc un frame différent[3]
D'autres alternances moins productives ne modifient pas le nombre d'arguments (comme l'alternance locative) mais ne sont pas pertinentes pour les domaines notionnels annotés.
Le causatif est syntaxiquement représenté dans les treebanks que nous utilisons comme une alternance syntaxique introduisant un actant (l'argument causateur). Le verbe *faire* est traité comme un auxiliaire. Par exemple:

(10) analyse syntaxique de 1774 modifié: [Les inspecteurs de l'ONU pour le désarmement de l'Irak]$_{\text{sujet}}$ ont [fait]$_{\text{aux_caus}}$ <u>détruire</u> [les stocks d'armement]$_{\text{objet}}$ [par des casques bleus]$_{\text{par_obj}}$.

Etant donnée la régularité et productivité de la construction causative, il serait aberrant de définir pour tout frame un équivalent frame-causatif, avec un argument causateur supplémentaire. Pour la représentation sémantique, on choisit de traiter le causatif de manière compositionnelle, avec le verbe faire déclenchant le frame Causation. Le rôle Effect est réalisé par l'infinitif introduit par faire et ses dépendants autres que le sujet. Le sujet remplit le rôle Cause ou Actor.

(11) annotation sémantique de 1774 modifié: [Les inspecteurs de l'ONU pour le désarmement]$_{\text{Actor}}$ de l'Irak ont **fait** [détruire les stocks d'armement par des casques bleus]$_{\text{Effect}}$.

Du fait de la divergence entre analyse syntaxique et sémantique, on peut avoir fréquemment discontinuité du remplisseur du rôle Effect, en particulier avec des clitiques sur le verbe *faire*:

[3]La relation sémantique entre la version transitive et la version neutre est capturée par une relation entre frames, la relation is_inchoative_of.

(12) 2252: [Il]$_{\text{Actor}}$ [le]$_{\text{Effect}}$ **fera** [condamner à un franc de dommages et intérêts pour diffamation]$_{\text{Effect}}$.

3.3 Emplois métaphoriques / figurés

En cas d'emploi métaphorique ou figuré, on annote uniquement la lecture littérale, mais en utilisant un flag "METAPHOR" sur l'instance de frame. Dans l'exemple suivant, il ne s'agit pas à proprement parler d'un achat de biens ou de services, on utilise un flag "METAPHOR" sur l'instance de frame Commerce_buy.

(13) l' **achat** d' une clientèle politique majoritairement liée à l' Occident.

Il est cependant parfois délicat de trancher entre emploi métaphorique ou considérer qu'il s'agit d'un autre sens du déclencheur. Pour trancher, on reprend d'une part les critères de (Ruppenhofer et al., 2006, p. 90) pour distinguer polysémie et métaphore : dans une métaphore, le déclencheur doit être substituable par ses synonymes / il doit y avoir correspondance parfaite (bijection) entre les rôles du frame littéral, et les rôles de l'emploi métaphorique.

(14) Pour eux , l' économie britannique , en renonçant au " grand large " , **acquitte** décidément un bien lourd tribut à l' Europe .

On ajoute le critère qu'à l'inverse, si il y sens distinct, celui-ci doit pouvoir s'utiliser ds d'autres contextes. Dans l'exemple suivant, pour "revendre", on constate qu'il n'est pas substituable par synonyme, donc il n'y pas métaphore. Mais le verbe n'admet pas les constructions syntaxiques classiques (#Elle revend des idées), donc pas de sens distinct pour "revendre" mais plutôt une construction figée (hors champ) "avoir des X à revendre").

(15) 15050 Cette fille d' immigrés italiens installés à Brighton a des idées à revendre.

3.4 Métonymie

Les noms de lieu peuvent classiquement être employés avec un sens correspondant aux personnes habitant dans le lieu ou bien appartenant à l'institution représentant le lieu. On les annote alors comme remplisseur d'autres types sémantiques :

(16) Au total , [le Japon]$_{\text{Exporter}}$ a **exporté** , en 1991 , 9 500 véhicules vers l' ex- URSS .

(17) l'objectif étant ... d' **inciter** [ces pays]$_{\text{Agent}}$ à passer par la porte insulaire réunionnaise.

Attention cependant, on n'applique cela qu'aux cas où le nom de lieu est dans une position commutant effectivement avec un nom d'animé. On ne considèrera pas qu'il y a métonymie dans le cas de noms de lieu introduits par une préposition locative, même si c'est ce PP qui donne clairement l'indication sur la sémantique d'un remplisseur animé. Dans l'exemple suivant, le complément en *chez* introduit bien le rôle Agent du frame Reason (ici l'animé ayant les doutes).

(18) ce réseau **suscitait** plus de doutes que de certitudes chez [les dirigeants d' Elf]$_{\text{Agent}}$.

mais dans les deux exemples suivants, le PP souligné donne l'indication de qui remplit le rôle Agent (qui a l'espoir / qui a l'opposition), mais ne commutent pas avec un animé, donc

on considère que le rôle Agent est non instancié (ici à annoter comme "extra-linguistic null instantiation", cf. section 4.4).

(19) les espoirs que **suscite** dans les Etats membres une future union économique et monétaire.

(20) Ce projet ... a **suscité** une telle opposition , de part et d' autre de la frontière.

Autre exemple de ce type :

(21) "Dans tel pays, les **importations** de blabla" ⟶ rien ne permet de dire que "tel pays" est l'Importer, seulement l'Importing_area: on annote "tel pays" en Importing_area mais pas Importer.

3.5 Redondance de déclencheurs ou pléonasmes

On a parfois "redondance de déclencheurs". Par exemple dans :

(22) Les mécanismes déflationnistes souvent évoqués ces derniers temps sont -ils les seuls **à l'origine de** ce **résultat** ?

On a deux déclencheurs du frame Causation, référant à la même instance de relation de cause à effet:

- *à l'origine de* déclenche Causation, Cause="les seuls", Effect="ce résultat"
- *résultat* déclenche Causation, avec Effect="ce résultat", cf. usage référentiel de "résultat", cf. section 6.3)

Malgré la redondance, on annote bien l'exemple (22) avec deux instances du frame Causation, non reliées entre elles (le fait de repérer quelles instances de frames sont "coréférentes" dépasse le cadre de notre annotation).

(NB: l'annotation du déclencheur "à l'origine de" et du déclencheur "résultat" ne se fait pas forcément dans le même fichier pré-annoté : il ne faut pas ajouter de frame sur un déclencheur qui n'est pas préannoté).

Autre exemple similaire :

(23) 1699: J' en suis satisfait pour deux **raisons** , tout d' abord **parce que** la Commission relie la Charte des droits fondamentaux à la conférence intergouvernementale.

NB: les deux instances de frame peuvent, selon les déclencheurs, ne pas avoir le même niveau de granularité (mais être plutôt en relation de subsomption). Par exemple:

- *"susciter une réaction"* : susciter ⟶ frame Reason + réaction ⟶ frame plus spécifique Response
- *"dus à des motifs personnels"* : motif ⟶ frame Reason + dû ⟶ frame Causation

4 Annotation des remplisseurs: principes généraux

4.1 Quels sont les rôles à annoter ?

Le lien entre les rôles et la valence syntaxique des déclencheurs n'est pas complètement précisé dans la littérature framenet. Nous prenons le parti de restreindre les rôles défi-

nis dans les frames et annotés en corpus: seront définis et annotés uniquement des rôles remplissables par

- des éléments sous-catégorisés par les déclencheurs du frame (cela correspond à la majorité des cas)
- dans le cas de déclencheurs adjectivaux, prépositionnels ou conjonctions, on considère en outre le rôle qui dans l'emploi standard (canonique) du déclencheur, est réalisé par le gouverneur syntaxique du déclencheur. Par exemple pour un adjectif, on peut considérer l'emploi épithète comme emploi canonique, et le nom modifié par l'adjectif est typiquement un remplisseur de rôle. Par exemple le frame FR_Awareness-certainty-opinion a en gros deux rôles principaux: un Cognizer a dans son modèle cognitif un certain Content. L'adjectif *conscient* peut déclencher ce frame. Il sous-catégorise un complément en *de*, remplissant le rôle Content, et le rôle Cognizer est réalisé typiquement par le nom modifié lorsque l'adjectif est employé comme épithète, comme dans *les partenaires sociaux conscients des sacrifices à accomplir*.
- dans certains cas le déclencheur est lui-même remplisseur de rôle (plus précisément est la tête du constituant remplissant le rôle). C'est typiquement le cas de noms relationnels comme par exemple *conséquence*, qui évoque une relation de cause à effet, représentée par le frame Causation, avec un rôle Effect et un rôle Cause. Dans *Les assurances doivent chiffrer les conséquences du sinistre.*, on considère que *conséquences* déclenche le frame Causation, avec *le sinistre* comme rôle Cause, et le rôle Effect est tout le NP dont la tête est *conséquences*. Voir la section 6 pour plus de détails sur ces cas.

Pour chaque frame a été définie en amont la liste des rôles à annoter (cette liste apparaît en colonne 6 du fichier décrivant la sous-structure asfalda : $ASFALDASVN/framenet/fndata-asfalda/substructure_asfalda.txt), et pour chaque rôle s'ils sont Core ou pas:

- Les rôles Core sont sémantiquement obligatoires étant donnée la sémantique du frame et sont soit sous-catégorisés par au moins un déclencheur du frame, soit réalisables comme le gouverneur syntaxique du déclencheur, soit exprimés par le déclencheur lui-même (cf. les 3 cas supra).
- Les autres rôles à annoter sont les rôles sous-catégorisés par au moins un déclencheur du frame, mais pas sémantiquement obligatoires. Par exemple dans *Paul a crié au dentiste "j'ai mal"*, l'interlocuteur est réalisable par un complément sous-catégorisé (un PP en *à*), mais qui n'est pas sémantiquement obligatoire: lorsqu'il est absent comme dans *Paul a crié "j'ai mal"*, c'est soit qu'il est omis mais interprétable en contexte, soit qu'il est inexistant.

Sous Salto, apparaissent par défaut uniquement les rôles Core à annoter. Les autres rôles déclarés comme "à annoter", mais non Core, sont accessibles par clic droit "Add element".

4.2 Utilisation de l'annotation en syntaxe

L'annotation des frames et rôles se fait sur des arbres en dépendances syntaxiques, qui ont été annotés manuellement indépendamment du projet ASFALDA.
Vous avez toute liberté pour indiquer le contenu exact d'un remplisseur (i.e. l'ensemble des mots-feuilles constituant le remplisseur). La syntaxe peut permettre d'accélérer les choses, en particulier, dans la majorité des cas, un remplisseur est constitué de tous les

mots dominés directement ou indirectement par une tête. Pointer vers le noeud dominant accélère l'annotation.
Attention à bien prendre dans un remplisseur pas seulement la tête, mais également tous les compléments et modifieurs (en particulier les relatives).

4.2.1 Gestion des erreurs dans l'annotation syntaxique / divergences syntaxe-sémantique

Il se peut cependant qu'il y ait des erreurs d'annotation syntaxiques, ou plus rarement, qu'il y ait des divergences syntaxe-sémantique telles que l'on ne trouve pas de noeuds dominant exactement tous les mots du remplisseur, et seulement ceux-ci. C'est le cas pour un remplisseur réalisé via deux compléments syntaxiques, dans le cas d'un attribut de l'objet par exemple. C'est le cas également pour un remplisseur discontinu.

Dans ces cas, on peut sous Salto rattacher un rôle à plusieurs noeuds (annotation "multi-pattes, voir la section 4.5.5 également), quitte à rattacher directement au niveau des tokens si on ne trouve pas de noeud syntaxique adéquat.

Attention du coup, cela veut dire que l'on a toujours plusieurs façons d'annoter le même ensemble de mots (pour former un remplisseur) : on peut utiliser les noeuds syntaxiques, les noeuds feuilles tokens et un mélange des deux. On prend la convention d'annoter au plus haut possible : i.e. d'annoter le plus possible en utilisant des noeuds syntaxiques. Mais les différences possibilités sont de toutes façons neutralisées lors du post-traitement automatique des annotations.

(Voir également infra section 7.2: Cas de déclencheurs qui dépendent syntaxiquement d'un de leur arguments sémantiques)

On se sert effectivement de l'analyse syntaxique pour annoter (en particulier en prenant un noeud NC pour annoter tout un NP) MAIS ceci est modulo les erreurs d'analyse syntaxique: si possible il faut annoter en fonction de l'analyse syntaxique correcte (que l'on pense correcte) qui peut diverger de celle qui est proposée.

NB: On ne note pas les erreurs repérées d'analyse syntaxique : c'est hors de portée du projet d'annotation sémantique.

4.3 Mots fonctionnels introduisant des remplisseurs

4.3.1 Prépositions et complémenteurs vides

Contrairement au FrameNet anglais, on choisit de ne pas inclure dans les remplisseurs les prépositions régies (sémantiquement vides) ni les complémenteurs sémantiquement vides:

(24) J'ai **vendu** ma voiture à **[**Paul**]**$_{\text{Buyer}}$
$\longrightarrow$ le Buyer est "Paul" et pas "à Paul"

On exclut le "à" introduisant un A_OBJ, le "de" introduisant un DE_OBJ, le "par" introduisant un complément d'agent, le "que" introduisant une complétive. On exclut le "de" introduisant une source (*je suis parti de Paris*), le "à" introduisant une destination (*je vais à Paris*).

Attention : ceci vaut aussi pour les contractions *au / aux* de catégorie P+D, pourvu que le à soit un a-obj, et idem pour les P+D *du / des* dans le cas de DE_OBJ (ou de complément de nom, auquel cas la fonction n'est pas spécifiée). Celà amène malheureusement à des remplisseurs auxquels manque le déterminant:

(25) l' **acquisition** du **[**yacht**]**$_{\text{Goods}}$ par Deviers

Plus généralement, on n'inclut la préposition que si elle ajoute une information qui n'est pas encodée dans le nom du rôle, comme par exemple "à propos de" pour le rôle Topic:

(26) 4334 Une regrettable erreur nous a fait **écrire** à propos de **[**Wadi , film d' Amos Gitaï [...]**]**$_{\text{Topic}}$, que **[**Haïfa se trouvait dans les territoires occupés**]**$_{\text{Message}}$

Attention : on garde le complémenteur "si" à l'intérieur du remplisseur.

(27) 3109 Rien ne permet encore de **savoir** **[si** l' élimination touchera davantage les sortants que les nouveaux venus**]**$_{\text{Content}}$

versus

(28) 3109 transformé: Rien ne permettait alors de **savoir** **que** **[**l' élimination toucherait davantage les sortants que les nouveaux venus**]**$_{\text{Content}}$

4.3.2 Coordination / énumération de prépositions régies

En cas de coordination avec répétition de la préposition régie, on n'inclut dans le remplisseur aucune des prepositions, mais on inclut bien les conjonctions de coordination et virgules (ce qui donne un rôle "multi-pattes"):

(29) 6816 L'acquisition par la firme d' **[**un courtier et**]**$_{\text{Goods}}$ d'**[**un contrepartiste**]**$_{\text{Goods}}$:

Idem pour le complémenteur que : *Paul dit que blabla et que blibli* $\longrightarrow$ rôle Message = blabla + et + blibli.

Lors de coordinations avec plus de 2 coordonnés, on prend la virgule comme coordonnant. Par exemple dans "*je vends aux grands, aux petits et aux moyens*", le Buyer est "grands , petits et moyens".

Idem également dans le cas d'une énumération, i.e. sans coordonnant : "*je vends aux grands, aux petits, aux moyens*" : on ne prend pas les prépositions régies dans le remplisseur, mais on prend les virgules : "grands , petits , moyens".

4.3.3 Coordination de préposition composée

(30) La balle **au-dessus de1** la table et de2 la chaise
$\longrightarrow$ le déclencheur retenu est "au-dessus de1" (frame Locative_relation), le rôle Ground est constitué de "la table" "et" "la chaise".

4.3.4 Coordination de prépositions qui correspond sémantiquement à une ellipse de prédicat

(31) J'ai **acheté** **[**une chemise**]**$_{\text{Goods}}$ à **[**Paul et**]**$_{\text{Beneficiary}}$ à **[**Pierre**]**$_{\text{Beneficiary}}$.
$\longrightarrow$ même si on comprend qu'il y a 2 achats (a priori 2 chemises distinctes), étant

donné que les 2 actes pointeraient vers le même remplisseur pour le rôle Goods ("une chemise", même si cela correspondrait sémantiquement à 2 instances distinctes de chemises, mais une seule formulation), on n'utilise pas l'annotation "ARG CLUSTER" pour ce cas. On aurait un arg cluster dans le cas où les 2 actes auraient 2 jeux de 2 ou plus remplisseurs distincts :

(32) J'ai **acheté** une chemise à Paul, et une chemise à Pierre ⟶ arg cluster (voir la section 5.1.2).

4.4 Localité des remplisseurs

Pour une occurrence de déclencheur donnée, le cas standard le plus typique est le cas où les rôles, en particulier les rôles Core, sont exprimés dans la même phrase, qui plus est "localement" au déclencheur, i.e. dans le domaine de localité syntaxique du déclencheur (voir infra pour sa définition).

On peut cependant avoir des cas où un rôle n'est pas exprimé localement au déclencheur, soit qu'il n'est pas du tout exprimé dans la phrase courante, soit qu'il est exprimé dans une position non locale. On note ces deux cas par des "flag": soit sur le rôle non exprimé, flag de type "null instantiation" (dans le même esprit que les flags CNI / DNI / INI du FrameNet anglais), soit un flag "NLI" sur un rôle exprimé dans la phrase du déclencheur, mais pas localement à celui-ci.

Techniquement sous Salto un "flag" sur le rôle s'obtient par clic droit sur le rôle (les flags pré-définis apparaissent).

Attention: on ne met ces flags que sur les rôle "Core" (ceux qui sont affichés par défaut sous Salto).

NB: dans tous les cas, on annote toujours le remplisseur le plus local possible (cette annotation étant éventuellement doublée, dans le cas d'un remplisseur anaphorique, de l'annotation de l'antécédent, si présent dans la phrase, cf. infra section 4.5).

4.4.1 NLI: remplisseurs non-locaux

Le flag à utiliser pour les remplisseurs non locaux est NLI (Non-local instantiation).

Pour un déclencheur **verbal**, les remplisseurs locaux sont les dépendants syntaxiques directs (ou indirects via prépositions) du verbe/du nom. On ajoute aussi le nom modifié en cas de participiale épithète.
Donc dans ces 3 exemples, tous les remplisseurs sont locaux, et donc sans flag:

(33) Paul a **vendu** une chemise à Anna

(34) La **vente** d'Adidas par Tapie à Pentland

(35) les voitures **achetées** par Paul ne durent pas ⟶ "les voitures" = rôle Goods, local

Pour un déclencheur adjectival, on considère comme local :

- les dépendants (prépositionnels ou autres) de l'adjectif (*Paul est conscient du problème*)

- le nom modifié en cas d'adjectif épithète (*les gens conscients du problème*)
- le sujet en cas d'adjectif attribut du sujet (*les observateurs sont sceptiques*)

On considère NLI l'objet direct pour tous les cas d'attribut de l'objet:

- attribut de l'objet adjectival (*le plan laisse les observateurs sceptiques*)
- attribut de l'objet prépositionnel introduit par comme (*ils considèrent ce départ comme probable* : "ce départ" NLI par rapport à "probable")
- attribut de l'objet infinitive (*elle entend Paul parler à Pierre* : "Paul" NLI par rapport à "parler")

Pour un déclencheur nominal, on considère comme local les dépendants du nom (sous forme de SP ou d'adjectifs).
On aura un NLI aussi par exemple:

- pour le sujet des infinitifs (constructions avec verbe à contrôle, verbe à montée...)

  (36) Jean veut vendre sa moto
  $\longrightarrow$ le Seller "Jean" est NLI

- pour l'objet des infinitifs dans le cas d'un tough movement

  (37) un livre difficile à vendre $\longrightarrow$ le rôle "Goods"=le NP de tête "livre" est NLI par rapport à "vendre")

- dans le cas d'extraction longue distance, et donc potentiellement en cas de relatives, interrogatives, clitique "en", clitique "dont" (à juger selon les occurrences)

  (38) le livre [que]$_{\text{Goods/ANAPH+NLI}}$ Paul veut **vendre**

  (39) Le collectionneur [en]$_{\text{Goods/NLI}}$ a fait récemment

NB: un remplisseur avec flag FULLANT sera par définition non local, donc pas besoin de remettre explicitement un flag NLI en plus de FULLANT.
En dehors de ces cas de non-localité syntaxiquement régulières, on a aussi des cas de mentions hors domaine de localité, mais qui ne présentent pas de régularité. Exemples :

(40) 20089 [...] [il]$_{\text{Interlocuteur}_1}$ n' avait pas **discuté** de la baisse du dollar contre le yen lors de sa rencontre avec [le secrétaire américain au Trésor]$_{\text{Interlocuteur}_2}$ [...]
$\longrightarrow$ le deuxième interlocuteur de la discussion est clairement le deuxième participant de la rencontre.

(41) 21231: La terrible pression de [la grande distribution]$_{\text{Buyer/NLI}}$ pour **acheter** le plus bas possible asphyxiait [les fabricants]$_{\text{Seller/NLI}}$.

NB: lorsqu'un rôle n'est pas exprimé localement au déclencheur, on peut avoir des cas où plusieurs segments sont possibles pour être remplisseur non local (NLI). On choisit dans ce cas le remplisseur le plus local possible, au sens syntaxique. C'est une notion difficile à expliquer précisément (la définition formelle est simple, mais difficile à appliquer manuellement). On doit juger de la longueur approximative du chemin syntaxique entre 2 têtes, et également du type de chemin : on considère toujours comme plus local des dépendances de compléments sous-catégorisés par rapport à des modifieurs. Exemple :

(42) le Crédit lyonnais souscrira directement au capital de la Bf.G. , par ailleurs , un apport fusion à la Bf.G de Crédit lyonnais Deutschland pour 190 millions de deutschemarks ainsi que des **acquisitions** directes de titres à hauteur de 160 millions de deutschemarks permettront à l' établissement de M. Jean - Yves Haberer de détenir 26 % du capital de la Bf.G pour environ 3 milliards de francs .
⟶ les 2 mentions en orange sont concurrentes pour être le rôle Buyer/NLI du déclencheur "acquisitions"
⟶ la seconde mention est plus proche syntaxiquement du déclencheur

(43) L' **acquisition** d' Adidas , dont il possédait déjà 20 , 05 % du capital , permet à ce brasseur d' affaires sans complexe d' ajouter ce joyau des stades à son empire familial .
⟶ la seconde mention, bien que plus éloignée en termes de nombre de mots, est plus locale syntaxiquement (on préfère remonter à un dépendant du verbe "permettre", dont le déclencheur est le sujet, plutôt que de descendre dans la relative modifiant le déclencheur.)

4.4.2 Imbrication de remplisseurs

Un autre cas de non localité est celui d'un remplisseur apparaissant au sein d'un autre remplisseur. Ces cas sont assez rares, mais surviennent en particulier lorsqu'un des rôles peut être exprimé localement ou bien comme sujet d'un autre rôle. On annote alors une imbrication de remplisseurs.
Exemple:

(44) 2081 Il serait utile que je fasse quelques brefs commentaires sur les **raisons** de **[**[nos]$_{\text{Agent/NLI}}$ recommandations**]**$_{\text{Action}}$.

Une expression locale alternative de l'agent pour l'exemple ci-dessus serait:

(45) Il serait utile que je fasse quelques brefs commentaires sur les **raisons** pour **[**nous**]**$_{\text{Agent}}$ de **[**faire ces recommandations**]**$_{\text{Action}}$.

4.4.3 DNI : Definite null instantiation

A utiliser si le remplisseur n'est pas exprimé dans la phrase du déclencheur, mais correspond clairement à un élément mentionné dans les phrases précédentes (contexte linguistique, voir le fichier $ASFALDASVN/corpus/sequoia+ftb.id_raw pour avoir accès aux phrases de contexte).

⟶ le rôle ne pointe sur aucun élément de la phrase, mais porte un flag DNI

4.4.4 ENI : Extra-linguistic null instantiation

A utiliser si le remplisseur n'est pas exprimé dans la phrase du déclencheur, ne correspond pas à un élément mentionné clairement et explicitement dans les phrases du contexte, mais est tout de même interprétable, en utilisant des connaissances extra-linguistiques et/ou le contexte).

⟶ le rôle ne pointe sur aucun élément de la phrase, mais porte un flag ENI

(46) 443: D' ores et déjà , il faut envisager la phase de négociation avec les riverains concernant la procédure , qui peut prendre d' ailleurs des formes diverses , pour l' **acquisition** des emprises nécessaires au passage des réseaux d' assainissement .
⟶ d'après les phrases précédentes, on comprend que l'acquisition serait faite par la mairie, mais pas de mention précise "la mairie" dans le contexte ⟶ ENI

(47) 3825: D' autre part , la grande conférence internationale qui se tiendra à Washington pour **discuter** de l' aide à apporter aux nouveaux Etats indépendants pourrait avoir lieu vers le 22 janvier .
⟶ les interloculeurs de la discussion seront les personnes qui participeront à la conférence mentionnée dans la phrase. Elles ne sont pas citées donc pas DNI, mais clairement identifiées donc ENI.

(48) Dans ces conditions , loin de s' améliorer , la situation se dégrade encore plus vite et ceux qui ont **acheté** cherchent à **revendre** sans y arriver .
⟶ on comprend qu'il s'agit d'achat / vente de logement (et pas de n'importe quoi), mais pas de mention précise préalable ⟶ ENI

(49) Les constructeurs (automobiles) concentrent leurs **ventes**
⟶ les Goods sont des voitures, donc ENI ⟶ les Buyer sont ... n'importe qui pouvant acheter une voiture, on ne sait rien de plus, donc UNI

4.4.5 UNI : Unknown null instantiation

A utiliser pour tous les autres cas : i.e. cas où le remplisseur est totalement sous-spécifié, et tout ce que l'on sait c'est qu'il satisfait aux restrictions de sélection imposées par le prédicat et le contexte. Exemples :

(50) J'étais en train de manger quand Paul a appelé
⟶ tout ce que l'on sait c'est que "je" mange qqch de comestible / un repas ⟶ UNI

(51) 3161 les argentins **achètent** déjà des tirelires parce qu' ils ne jetteront plus les pièces de monnaie
⟶ le Seller est totalement sous-spécifié (à part qu'il faut que ce soit qqun qui ait des tirelires à vendre)⟶ UNI

(52) 19201: ...les crédits affectés dans le budget 1990 de [Paris]$_{\text{Buyer/NLI}}$ aux **acquisitions** [foncières]$_{\text{Goods/ELLIPTIC}}$ ont subi une coupe ...
(pour le rôle Goods : voir section 7.3, pour le rôle Buyer : "Paris" est une mention métonymique référant à l'organisation la ville de Paris)

NB: les flags UNI, ENI et DNI ne sont utilisables uniquement s'il n'y a pas de remplisseur exprimé dans la phrase. Si on a une expression linguistique pour un remplisseur, pas de flag UNI / ENI / DNI, même si l'expression est vague ou anaphorique. Exemple:

(53) la **cession** d'Adidas à [un tiers]$_{\text{Buyer}}$ ⟶ on ne connait pas l'identité de ce tiers, mais on ne met pas de flag UNI

4.5 Remplisseurs anaphoriques

Dans le cas d'un remplisseur anaphorique, d'un côté on voudrait annoter l'anaphore comme remplisseur. En effet cela donne des informations relatives à la distribution des fonctions grammaticales pour remplir tel ou tel rôle. D'un autre côté, d'un point de vue sémantique, on voudrait récupérer l'antécédent, pour avoir des informations sur la classe sémantique du

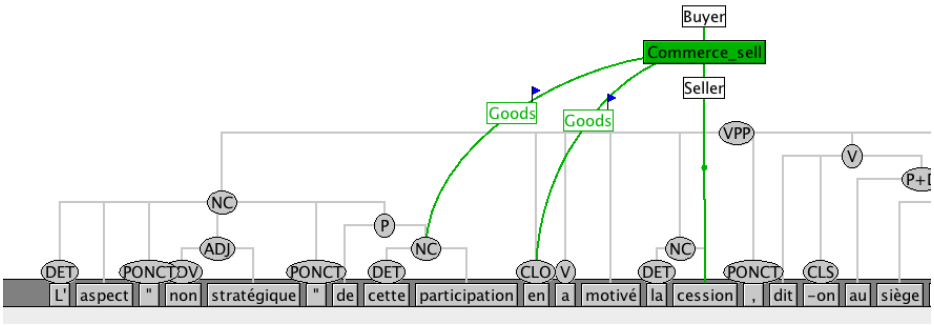


Figure 2: Exemple d'annotation d'un remplisseur anaphorique (en outre non local) et de son antécédent.

remplisseur.

Pour cette raison, on annote à la fois l'élément anaphorique (pronom, clitique, possessif (anaphore du possesseur)) et l'antécédent sémantiquement plein le plus proche (comme fait également dans le projet FrameNet original).

Mais on distingue les 2 remplisseurs en utilisant le flag "FULLANT" pour l'antécédent sémantiquement plein le plus proche, et "ANAPH" pour le remplisseur anaphorique.

NB: On n'utilise le flag "ANAPH" que s'il y a annotation d'un "FULLANT".

Exemple de remplisseur de type clitique :

(54) On pensait que les plus - values compenseraient l' insuffisance du rendement de [ces actifs]$_{\text{Goods/FULLANT}}$ par rapport aux charges financières assumées pour [les]$_{\text{Goods/ANAPH}}$ **acheter** .

NB: un remplisseur marqué FULLANT est par définition non local, donc inutile d'ajouter le flag NLI. En revanche, un remplisseur marqué ANAPH peut être en outre non local : on doit alors cumuler les 2 flags ANAPH et NLI, comme par exemple le clitique "en" dans l'exemple de la figure 2.

NB: on n'utilise le doublage ANAPH / FULLANT que dans le cas d'un élément anaphorique qui n'est pas un nom (pronom, clitique, déterminant possessif, formulations comme "ce dernier"). On ne l'utilise pas dans le cas de deux NP définis qui réfèrent à la même entité. Ainsi dans l'exemple ci-dessous, il y a deux expressions qui réfèrent à l'acheteur : "le groupe français" et "Thomson" mais la mention la plus locale étant nominale ("le groupe français") et pas pronominale, on l'annote sans annoter par ailleurs Thomson.

(55) Thomson a calmé ses ardeurs, en effet le **rachat** de LTV par [le groupe français]$_{\text{Buyer}}$ a été mal perçu.

NB: Par ailleurs, on n'utilise le doublage ANAPH / FULLANT que si l'on dispose dans la phrase d'un antécédent comportant une tête lexicale sémantiquement pleine (donc pas un pronom). Dans l'exemple suivant, le remplisseur "qui" a pour antécédent "ce", qui a lui-même pour antécédent "La thrombine active également les facteurs V et VIII" ⟶ c'est cette dernière qui est annotée comme FULLANT (le "ce" étant lui-même pronominal).

(56) [La thrombine active également les facteurs V et VIII]$_{\text{Cause/FULLANT}}$, ce [qui]$_{\text{Cause/ANAPH}}$ **entraîne** une plus grande production de thrombine.

4.5.1 Relatives

Ce traitement s'applique en particulier dans le cas d'un déclencheur inclut dans une relative.

(57) [les gens]$_{\text{Buyer/FULLANT}}$ [qui]$_{\text{Buyer/ANAPH}}$ ont **acheté** du vin.

(58) ceux [qui]$_{\text{Buyer}}$ ont **acheté** du vin ⟶ "qui"=Buyer, pas d'antécédent sémantiquement plein, donc pas d'annotation de "ceux" et pas de flag

(59) des consommateurs , qui ... veulent vérifier avant d' **acheter** et en avoir pour leur argent ⟶ le Buyer le plus local est "qui" (⟶flag ANAPH) + "des consommateurs" =Buyer/FULLANT

(60) [le type]$_{\text{Buyer/FULLANT}}$ à [qui]$_{\text{Buyer/ANAPH}}$ Paul a **vendu** sa moto.

(61) Et les parents vont soupirer en pensant aux [piles]$_{\text{Goods/FULLANT}}$ [qu']$_{\text{Goods/ANAPH}}$ il faudra inlassablement **acheter**.

NB: pour le FULLANT d'un pronom relatif, c'est en général le nom modifié par la relative. **POUR ALLER PLUS VITE**, on note comme remplisseur FULLANT le noeud NC dominant le nom modifié : ce noeud NC inclut lui-même la relative.

(62) 12824: ... pour lui permettre d' acheter [les francs [qu']$_{\text{Goods/ANAPH}}$ on lui **vend**]$_{\text{Goods/FULLANT}}$ libéralement. ⟶ le Goods/FULLANT est noté sur le noeud NC dominant "francs", et couvre donc "les francs qu'on lui vend libéralement".

ATTENTION: si on a un syntagme relatif (comme "prep+qui") et que la préposition est sémantiquement pleine (et pas régie), alors l'antécédent du pronom relatif ne correspond pas au remplisseur dans son ensemble. Par exemple :

(63) le champ à travers lequel Paul a **couru** ⟶ "à travers lequel" = le rôle Path, "champ" est l'antécédent de "lequel", mais n'est pas le remplisseur entier ⟶ dans ce cas on annote uniquement "à travers lequel"=Path, sans flag.

4.5.2 Réfléchis

Le "se" n'est jamais utilisé comme remplisseur anaphorique : même pour les occurrences du clitique *se* qui sont considérées comme des occurrences du "vrai réfléchi" ou "vrai réciproque" (cf. section 2.1.3, on considère le "se" comme un marqueur a-sémantique, et c'est le sujet du verbe réfléchi ou réciproque qui remplit deux rôles à la fois. Par exemple :

(64) Et tous se **disaient** que , tout bien réfléchi , une âme , cela n' a pas de prix . ⟶ le sujet "tous" remplit à la fois le rôle Speaker et l'Addressee.

4.5.3 Possessifs

Les déterminants possessifs sont anaphoriques d'un point de vue sémantique, et donc peuvent constituer la mention la plus locale d'un remplisseur d'un déclencheur nominal. Le possessif correspond à la pronominalisation du complément en "de" d'un nom, et est sémantiquement ambigu quand au rôle joué: dans *l'acquisition de XXX*, XXX peut être l'acheteur ou les biens vendus. Exemples:

(65) Des intermédiaires ont facilité l' **acquisition** des [frégates de Taïwan]Goods

(66) Concernant [les frégates de Taïwan]$_{\text{Goods/FULLANT}}$, il est à noter que des intermédiaires ont facilité [leur]$_{\text{Goods/ANAPH}}$ **acquisition**.

(67) Concernant [les frégates de Taïwan]$_{\text{Goods/FULLANT}}$, il est à noter que des intermédiaires [en]$_{\text{Goods/ANAPH}}$ ont facilité l' **acquisition**.

(68) [Des consommateurs]$_{\text{Buyer/FULLANT}}$ qui retardent [leur]$_{\text{Buyer/ANAPH}}$ **achat** $\longrightarrow$ il s'agit d'un achat fait par les consommateurs. La mention la plus locale est "leur"=Buyer/ANAPH, ensuite "qui", mais on remonte jusqu'au premier antécédent sémantiquement plein : "des consommateurs"=Buyer/FULLANT.

ATTENTION: le déterminant doit être un possessif. Dans le cas d'un déterminant démonstratif, c'est tout le NP qui est une description définie anaphorique, faisant référence à un évènement préalablement évoqué dans le contexte. Par exemple :

(69) La **vente** d'Adidas par Tapie a été rapide. Cette **cession** intervient à un moment critique. $\longrightarrow$ pour le Commerce_sell évoqué par "cession", on aura des DNI sur les rôles, cf. ils sont remplis dans la phrase précédente, avec le déclencheur "vente".

4.5.4 Cas particuliers: "en" quantitatif / démonstratif "celui"

Le pronom "celui" est en général anaphorique avec une classe de référents, mais pas avec le référent précis lui-même. Dans l'exemple suivant, "celui" reprend la classe sémantique "sac" mais ne renvoie pas à l'entité dénotée par le NP "deux sacs". :

(70) J'avais repéré deux sacs, j'ai **acheté** finalement celui avec une bandoulière.

Dans ce cas, on note quand même un double remplisseur, avec flag ANAPH / FULLANT, mais pour le FULLANT, on pointe sur le ou les tokens de la classe (ici le token "sacs"), et pas sur le NC englobant tout le NP ("deux sacs").
Important: La double annotation ANAPH / FULLANT n'est à utiliser que si c'est l'anaphore est la tête du remplisseur. Donc par exemple dans *"Paul a vendu sa voiture"*, on a Goods="sa voiture", et on ne note pas l'antécédent de "sa", qui n'est qu'une partie du remplisseur.

Avec le "en" quantitatif (*Paul en a vendu trois* $\longrightarrow$ Goods="en"+"trois" / *Paul en vend une autre* $\longrightarrow$ Goods="en"+"une autre"), c'est le *"en"* qui fournit la classe sémantique du remplisseur, via sa relation anaphorique avec un nom du contexte. Dans ce cas, on utilise le même type d'annotation qu'avec *"celui"*, le FULLANT est le nom et pas tout le NP. Donc par exemple :

(71) Une législation peut en **amener** une autre plus restrictive. $\longrightarrow$ Cause="une législation", Effect/ANAPH="en"+"une autre", Effect/FULLANT="législation".

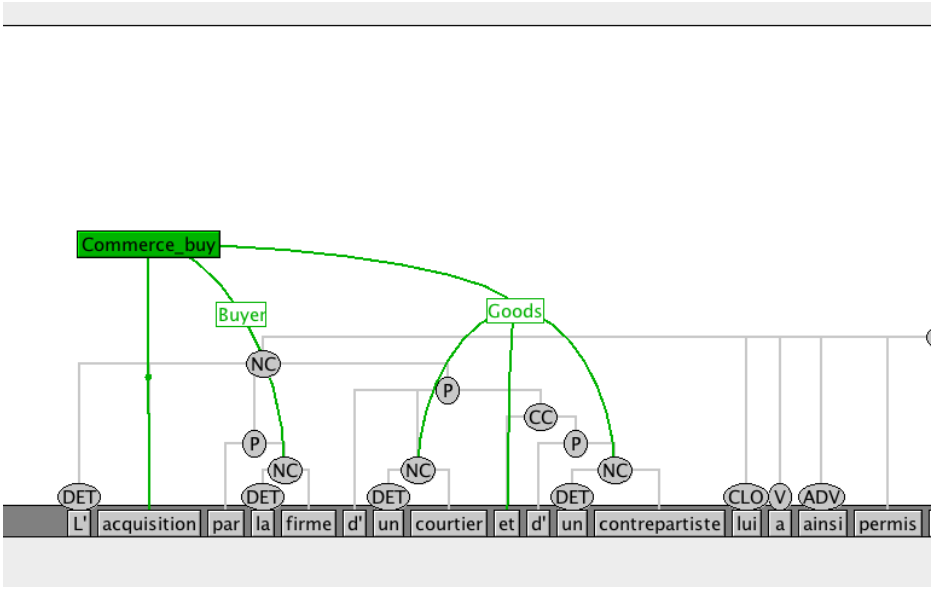


Figure 3: Exemple d'annotation "multi-pattes": une instance de rôle (ici Goods) contient plusieurs "pattes".

4.5.5 Aspects techniques : 2 instances de remplisseurs, versus un remplisseur "multi-pattes"

La solution ci-dessus dans le cas d'un remplisseur anaphorique, doublé d'un remplisseur "FULLANT", est à distinguer techniquement d'un remplisseur contenant plusieurs noeuds (dit "multi-pattes"):

- Pour le cas ANAPH / FULLANT : après avoir créé le premier remplisseur (par exemple celui avec flag ANAPH, il faut pour créer le second faire un clic droit sur le FRAME, puis Add element : répéter le même rôle. Par exemple dans la figure 2, on a bien deux instances de rôles Goods.
- En revanche, pour un remplisseur constitué de plusieurs noeuds : il faut sous salto annoter le rôle sur un premier noeud, puis faire un clic sur le rectangle donnant le nom du rôle et en maintenant le clic aller jusqu'au prochain noeud à inclure dans le remplisseur. Ce qui donne graphiquement par exemple dans la figure 3 pour le rôle Goods : on a un seul lien Goods, divisé en 3 pattes.

5 Points spécifiques aux déclencheurs verbaux

5.1 Ellipses de verbes

Certaines coordinations mettent en jeu une ellipse du prédicat, et la partie elliptique peut correspondre sémantiquement à une autre instance d'éventualité. On choisit dans ces cas d'annoter deux instances du même frame, sur la même occurrence de déclencheur, et de marquer par un flag la deuxième instance de frame[4]. On se limite aux 3 cas suivants:

[4]flags accessibles via clic droit sur l'instance de frame "Add flags".

5.1.1 Head Gapping : coordination, où le verbe n'est pas répété (mais 2 sujets différents)

(72) Exemple construit: Paul **achète** une chemise et Pauline des chaussures
⟶ il y a bien 2 actes d'achat, avec 2 jeux de remplisseurs différents.

(73) 6604: ... dont 6,4 milliards **achetés** par les Français en vacances et 4,2 milliards par les touristes étrangers.
⟶ on a 2 actes d'achat : les 6,4 milliards achetés par les français, et les 4,2 milliards achetés par les touristes étrangers
⟶ on crée deux instances du même frame Commerce_buy, sur le même token "achetés", avec chacun les remplisseurs adéquats.

En outre, on marque l'instance correspondant à l'ellipse (la deuxième : l'achat par les touristes) au moyen d'un flag "HEADGAP" sur l'instance de frame.

5.1.2 Argument cluster (i.e. "coordination de séquences de constituants")

Dans ce cas, le deuxième conjoint de la coordination ne contient ni le verbe (non répété), ni le sujet (non répété).
De la même manière, on utilise deux instances de frames, sur la même occurrence de déclencheur, avec un flag ARGCLUSTER sur la deuxième instance.

(74) Exemple construit: Paul achète une chemise à Marie et des chaussures à Pierre.

(75) 10047: Steven Ross avait **acheté** une agence de jeunes talents en 1967 et , deux ans plus tard , la société cinématographique Warner - Seven Arts.
⟶ une instance de Commerce_buy avec Buyer="Steven Ross", Goods="une agence de jeunes talents"
⟶ plus une autre instance, flag ARGCLUSTER, avec Buyer="Steven Ross", Goods="la société cinématographique Warner - Seven Arts"

Important: attention à distinguer arg cluster et remplisseur coordonné. Pour choisir d'annoter via arg cluster et donc avec plusieurs instances de frames, il faut que les deux instances de frame aient au moins un remplisseur différent, pour un des rôles à annoter: si le arg cluster porte sur des éléments qui ne sont pas à annoter (par exemple des modifieurs temporels), alors on ne fait qu'une seule instance de frame :

(76) Exemple construit: Paul a **acheté** [une fleur]$_{\text{Goods}}$ le lundi à Paris, et le mardi à Tokyo.
⟶ une seule instance de Commerce_buy

Il faut en outre que les 2 remplisseurs distincts ne forment pas une coordination, sauf si on a des modifieurs de temps ou de lieu différents pour les différents coordonnés:

(77) Paul **achète** [une chemise et un maillot]$_{\text{Goods}}$.
⟶ la coordination porte sur un seul remplisseur (Goods) ⟶ on ne fait qu'une seule instance de Commerce_buy, avec un remplisseur coordonné.

(78) Paul **achète** une chemise à [Marie et]$_{\text{Beneficiary}}$ à [Pierre]$_{\text{Beneficiary}}$ ⟶ idem la coordination porte sur un seul remplisseur ⟶ on ne fait qu'une seule instance de Commerce_buy, avec remplisseur coordonné.

(79) 10141: ... les **rachats** successifs de la CFAO puis de Conforama et de Le Printemps ....
⟶ on ne fait pas trois instances de Commerce_buy pour le déclencheur "rachats", mais un seul, avec Goods="la CFAO puis Conforama et Le Printemps".

(80) 16928: - Accor a **acquis** , depuis octobre 1989 , non pas 1 % , comme nous l' avons écrit [...] , MAIS bien 2 % du capital du club méditerranée.
⟶ ici "mais" est le coordonnant du NP coordonné "non pas 1 % , comme nous l' avons écrit [...] , MAIS bien 2 % du capital du club méditerranée".
⟶ tout ce NP coordonné est le remplisseur Goods, pour le déclencheur "acquis".

Mais si les 2 coordonnées correspondent à des lieux / des temps explicitement différents, alors on fait un arg cluster :

(81) Exemple construit: Paul a **versé** 3 millions en 1992 et 4 millions en 1993.
⟶ arg cluster remplisseur 1 = "3 millions", et remplisseur 2 = "4 millions"

A fortiori, lorsque les 2 remplisseurs ne forment pas de coordination, comme par exemple avec la préposition "contre" :

(82) 12688: Pour l' heure , l' Etat a **versé** , en 1992 , 13 , 7 milliards de francs de prestations contre 12 , 1 milliards en 1991 .
⟶ instance 1 : (en 1992) l'état a versé 13,7 milliards
⟶ instance 2 : (en 1991) l'état a versé 12,1 milliards

5.1.3 Comparatives réduites

(83) Exemple construit: Paul **félicite** plus Pierre que Nathan.
⟶ ambiguïté à résoudre en contexte entre *"Paul félicite plus Pierre que Paul ne félicite Nathan"* et *"Paul félicite plus Pierre que Nathan ne félicite Pierre"*.

Dans les 2 cas on fait 2 instances du frame évoqué par "félicite", la deuxième instance avec flag "REDUCEDCOMP".

(84) 9361: Chipie , [...] , reconnaît être plus **acheté** par les filles que les garçons . [5]

On annote pour ce dernier exemple deux instances de Commerce_buy:

- une avec Goods="Chipie, [...]" / Buyer="les filles"
- l'autre, avec flag "REDUCEDCOMP" avec Goods="Chipie, [...]" / Buyer="les garçons"

NB: la situation est différente lorsque l'on annote un déclencheur présent dans le deuxième terme de la comparative et pas le premier (ds X adv_comparatif que Y : qd le déclencheur est dans Y). Dans ce cas, le Y a tendance à être elliptique : attention, on n'utilise pas des morceaux du premier terme de la comparative pour annoter le second terme elliptique :

(85) 5653: ... faire passer la pilule des réformes nécessaires ... que beaucoup d' Italiens ... risquent ... de trouver plus amère que **prévu** ⟶ le déclencheur "prévu" est dans le deuxième terme de la comparative ⟶ on note le frame Expectation avec rôle Phenomenon = ENI

[5] On interprète que ce qui est acheté c'est "Chipie, [...]", avec métonymie entre le nom de marque et les produits produits par la marque.

6 Points spécifiques aux déclencheurs nominaux

Voir à ce sujet également la page "focus: déclencheurs nominaux"
`https://sites.google.com/site/wikiasfalinks/home/annotation-en-corpus-via-salto/dÃ]clencheurs-nominaux`
$\longrightarrow$ on annote non seulement les occurrences de sens processus, mais également les occurrences de sens correspondant à un participant du processus et/ou au résultat du processus

6.1 Nominalisation référant à un évènement versus nominalisation référant à un résultat de l'évènement

Certains noms correspondant à des nominalisations de verbes sont ambigus, dans la mesure où ils peuvent référer à l'éventualité décrite par le verbe ou bien référer à un "participant" de cette éventualité, et/ou au résultat de l'éventualité. Par exemple pour le nom "achat":

(86) l'achat a duré 3 heures / l'achat a été annulé $\longrightarrow$ "achat" réfère bien au processus d'achat, qui serait évocable avec le verbe "acheter" $\longrightarrow$ achat déclenche le frame Commerce_buy

(87) j'ai montré mes achats à Vanessa
$\longrightarrow \neq$ processus, on parle des objets achetés $\longrightarrow$ les objets achetés sont considérés comme un "participant" de l'évènement d'achat, ce qui se traduit par le fait que le frame Commerce_buy a un rôle Goods pour y référer.

Pour le nom "reproduction", on peut avoir le sens évènementiel "reproduire une oeuvre" ou bien le sens du résultat de l'évènement de reproduction :

(88) La reproduction du tableau a été effectuée avec soin, par un faussaire professionnel

(89) La reproduction du tableau a été placée à la place habituelle du tableau.
$\longrightarrow$ "reproduction" réfère ici à l'objet physique, qui est le résultat d'un processus de reproduction.

Dans toute la suite on parle des occurrences comme (87) et (89) comme les occurrences REFTOPARTICIPANT (on ne distingue pas participant ou résultat). Dans un emploi de type REFTOPARTICIPANT, est tout de même évoquée l'éventualité sous-jacente, même si le nom réfère à un participant ou résultat, et pas à l'éventualité.

Pour cette raison, on choisit d'annoter ces éventualités sous-jacentes, s'il s'agit d'éventualités évoquant des frames de nos domaines. Par définition, pour ce type d'emploi, un des rôles du frame sera rempli par le constituant dont le déclencheur nominal est la tête.

Donc par exemple pour (87) : "achat" déclenche Commerce_buy, et on note le rôle Goods = "mes achats" (noeud NC dominant "achat"). On se retrouve alors avec un remplisseur qui contient le déclencheur.

Rem: en théorie, dans certains cas de référence au résultat d'une éventualité (comme (89) ci-dessus), parfois le résultat ne correspond par à un participant : la sous-cat du verbe "reproduire" ne fait pas intervenir le résultat de la reproduction, mais uniquement le reproducteur et l'objet reproduit. Dans ces cas-là, il manquerait un rôle à annoter. MAIS pour l'instant, pour les frames qui nous occupent, on a toujours trouvé un rôle déjà prévu,

auquel réfère la nominalisation de type REFTOPARTICIPANT.

6.1.1 Critères de distinction entre référence à une éventualité versus référence à un participant ou résultat

Reste à décider en contexte si une nominalisation évoque l'éventualité ou bien un participant/résultat de l'éventualité.

Comme principe général, on pose que lorsque les 2 lectures sont possibles, on privilégie la lecture événementielle, et donc avec annotation sans remplisseur qui contienne le déclencheur. Par exemple :

(90) 412: Du point de vue financier , l' association fonctionne essentiellement grâce à l' aide d' une subvention attribuée par le siège basé à Valenciennes ainsi qu' une **cotisation** de 17 euros par an et par adhérents . ⟶ pour **cotisation**, les 2 lectures semblent possibles (la somme d'argent, ou bien l'action de cotiser, par exemple, on peut forcer cette deuxième lecture avec "une cotisation par les adhérents de 17 euros") ⟶ on privilégie en cas d'ambiguïté, la lecture évènementielle.

(91) 14647: Une Europe dont les **importations** en provenance de pays en voie de développement (hors OCDE pour simplifier) augmentent inexorablement : 9 , 2 % du total en 1984 , 13 , 7 % en 1992 .
⟶ d'un côté "en provenance de pays" fait plutôt opter pour importations=les biens importés
⟶ d'un autre côté c'est plus les actes d'importation qui augmentent que les biens importés (#les biens augmentent ne veut pas dire grand chose) (cf. exemple supra avec "réduire")
⟶ on privilégie l'annotation processus

Indice: Indication de durée : Une consigne générale est que lorsque l'on a / on peut modifier le nom avec une durée, ou bien que l'on a une information de durée, alors il s'agit forcément de l'évènement:

(92) 4786 Enfin , le CERC rappelle qu' un allongement de la durée de **cotisation** [...] avantagerait relativement les non- cadres , dont la durée de cotisation est actuellement de quarante ans .
⟶ COTISATION interprété comme l'évènement (cf. durée de cotisation).

Indice: coordination: Un autre indice est la coordination du nom. En effet, on ne coordonne en général que des noms de même type sémantique, donc par exemple :

- "achats et autres babioles" ⟶ "achats" réfère aux biens achetés (⟶ REFTOPARTICIPANT)
- "achats et autres activités" ⟶ "achats" réfère au processus d'achat (⟶ pas REFTOPARTICIPANT)

(93) 6320 ... une nouvelle convention collective qui prévoit notamment des garanties de reclassification et de **rémunération**
⟶ comme "reclassification" n'est pas ambigu (réfère forcément au processus), il permet de désambiguiser "rémunération"

Indice: Verbe support d'action ou d'évènement : si le N est un complément des verbes comme "effectuer", "faire", "commettre", "accomplir", "mener à bien", "mener à son terme", "procéder à" alors il s'agit d'un évènement. Donc par exemple :

(94) 3555 Près_de 50_000 **importations** de ce type auraient été effectuées en 1990.
⟶ importations =forcément un évt, donc ça ne peut référer qu'à l'acte d'importer et pas aux biens importés

Indice: restrictions de sélection du verbe ou nom enchâssant : si le N est un complément d'un verbe ou nom posant des restrictions de sélection sur ce complément (i.e. des contraintes sur le type sémantique du complément) alors la restriction de sélection peut parfois désambiguiser. En effet, le sens "processus" et le sens "participant" n'ont en général pas le même type sémantique, cf. en général le sens "participant" réfère à un objet (par exemple les biens achetés dans le cas de "achat"). Idem si le N est complément d'un N posant lui-même des restrictions de sélection

(95) 15209: ... le règlement des **achats** par les détaillants slovaques constitue la principale angoisse de Pavel Maly .
⟶ on ne peut pas régler un acte d'achat / on règle une personne, un bien ou une somme ⟶ donc dans cet exemple "achats" réfère aux Goods.

(96) 16223 Le groupe de Pennsylvanie affirme payer 154 millions de dollars (831 millions de francs) sa nouvelle **acquisition** , prix d' achat et reprise des dettes compris .
⟶ on paie un bien un certain prix, mais on ne peut pas "payer un acte d'achat un certain prix" ⟶ donc ici frame sur "acquisition" avec REFTOP (et frame également sur "achat")

Indice: modification du nom par un modifieur référant au processus

(97) M. Dumas a rappelé la **décision** " rapide , claire , ferme " , prise par les_Onze , la semaine dernière , à Oslo : pas de renégociation du Traité .
⟶ même si la partie après les ":" réfère au rôle Decision, ici la modification par "rapide" du nom "décision" indique qu'il s'agit du processus (c'est le processus de décision qui est rapide, ça ne peut pas être l'absence de renégociation du traité qui est rapide)
⟶ pas REFTOPARTICIPANT

Indice: modification du nom par un modifieur référant à l'agent du processus

(98) le gouvernement avait pris , fin 1991 , une **décision** courageuse
⟶ la modification par "courageuse" est un indice qu'il s'agit du fait de décider : "courageuse" s'interprète comme "en décidant le décideur a fait preuve de courage, on privilégie dans ce cas la lecture évènementielle ⟶ pas REFTOPARTICIPANT

En l'absence de ce type d'indices (qui sont rares...), il faut se fier à son interprétation, en privilégiant l'annotation sans REFTOPARTICIPANT en cas d'hésitation. Ci-dessous quelques exemples :

(99) 9386 De plus , on constate à l' heure actuelle un changement de comportement des consommateurs , qui , d' abord , réduisent leurs **achats** , et ....
⟶ on ne peut pas réduire les biens (*les consommateurs réduisent leurs sacs) mais on réduit (un peu mieux) l'acte d'achat (les consommateurs réduisent leurs actes d'achat) ⟶ frame Commerce_buy

(100) Depuis dix - huit mois , il dirige la production et le personnel de Lafarge Zemment , la dernière **acquisition** du groupe Lafarge Coppée .
⟶ L'apposition indique que Lafarge Zemment=acquisition , donc il s'agit forcément de l'objet acquis, pas du processus.

Dans le cas d'un nom évènementiel avec déterminant démonstratif, on a souvent affaire à une coréférence évènementielle (et donc pas à un cas REFTOPARTICIPANT), sans que ce soit systématique.

(101) 3795 Cette **acquisition** permet surtout à Grand Metropolitan de se doter d' un réseau commercial en Amérique latine ...
⟶, "cette acquisition" réfère à un acte d'achat évoqué dans le contexte précédent.

(102) 11423 Ebauches SA , dont le siège est à Besançon , s' est engagé à reprendre 273 des 422 employés d' ici à 1996 et à investir 30 millions de deutschemarks (120 millions de francs) dans sa nouvelle **acquisition** .
⟶ Ebauches SA n'investit des millions dans un acte d'achat, mais investit des millions dans l'entreprise achetée ⟶ pas évènementiel ⟶ REFTOP

(103) Carryfast est la seizième **acquisition** d' UPS en Europe depuis 1985 .
⟶ la construction avec copule + attribut du sujet indique que l'on a coréférence entre "Carryfast" et "la seizième acquisition ..." ⟶ donc on n'a pas ici stricto sensu le processus ⟶ MAIS 27/05/2015 : On a ici une interférence avec l'emploi prédicatif : ici effectivement il s'agit du "acquisition" qui s'il était *en emploi référentiel* réfèrerait aux biens achetés, mais il est ici en emploi prédicatif (cf. la section suivante 6.2). Si on adopte le même traitement que celui pour "cause" ou "conséquence" en emploi prédicatif, on doit considérer qu'*acquisition* ne réfère pas au bien acheté dans cet exemple!!! on a donc plusieurs manières pour "acquisition" de ne pas référer aux biens achetés : soit parce qu'il réfère à l'évènement, soit parce qu'il est en emploi prédicatif.

6.2 Noms relationnels et frames représentant une relation

On considère dans cette section qu'un nom est **prédicatif** s'il a une structure argumentale, i.e. au moins un argument syntaxique, correspondant à un argument sémantique du nom.

On a vu donc dans la section précédente des noms prédicatifs ambigus car pouvant référer à une éventualité ou bien à un participant ou au résultat d'une éventualité.
Par exemple "achat" réfère soit au processus d'acheter, ou bien aux biens effectivement achetés. Les deux sens sont annotés avec le même frame, mais avec ou sans annotation de type REFTOPARTICIPANT.

Nous abordons ici un autre caractéristique de certains noms prédicatifs, dits noms relationnels: leur sémantique met en relation deux entités (entités concrètes, ou bien elles-mêmes évènementielles) et est telle que lorsque le nom est employé comme tête d'un NP saturé, i.e. avec tous ses compléments, et en emploi "référentiel", le NP réfère à un des arguments de la relation. (On précisera dans la prochaine section (section 6.3) ce que l'on entend par emploi référentiel par opposition à emploi prédicatif.)

Typiquement le nom *père* évoque la relation de "paternité" entre deux individus, mais un NP saturé de tête "père" réfère à un des 2 individus: dans *J'ai rencontré le père de Paul*, les deux individus en question sont *Paul* et *le père de Paul*, on a bien que le NP réfère à un

des 2 individus mis en relation.

Par contraste, le nom *similitude* évoque lui-aussi une relation entre deux (ou plus) d'individus, mais en emploi référentiel, un NP de tête *similitude* réfère à la relation elle-même et pas à une des entités similaires.

(104) la similitude entre Paul et Pierre est troublante
$\longrightarrow$ le NP de tête similitude réfère à la relation et ni à Paul ni à Pierre.

Les noms d'évènements sont d'autres exemples typiques de noms prédicatifs qui ne sont pas relationnels. Par exemple le nom *cambriolage*, est prédicatif (cambriolage de X par Y), mais non relationnel cf. en emploi référentiel *cambriolage* réfère à un évènement, et ni au cambrioleur ni au cambriolé.

De fait, la modélisation en frames des noms relationnels pourrait suivre la syntaxe: i.e. considérer un frame avec un seul rôle pour "le-père-de-X", mais il est plus naturel de considérer un frame à deux rôles (pour "relation de paternité entre X et Y"). On a alors la particularité d'un frame évoqué par un nom, qui employé référentiellement réfère à un des rôles du frame.

Dans les domaines annotés dans ASFALDA, on a rencontré des noms relationnels en particulier dans le domaine de la causalité.

Par exemple le nom *cause* est relationnel en cela qu'il évoque une relation de cause à effet entre 2 entités (une cause et un effet), et réfère, en emploi référentiel, à l'entité jouant le rôle de la cause. On considère *cause* comme un déclencheur du frame à 2 rôles (Cause et Effect) Causation, et le NP de tête *cause* est le remplisseur du rôle Cause. Le complément en *de* du nom remplit le rôle Effect.[6] Exemple:

(105) 330: Les policiers ont ouvert une enquête afin de déterminer les **[causes** exactes de **[**cet incendie**]**$_{\text{Effect}}$ **]**$_{\text{Cause}}$.

De la même manière, le nom *conséquence* est modélisable avec le même frame Causation, mais réfère (en emploi référentiel) au rôle Effect:

(106) 10220: Refusant de " faire supporter aux travailleurs les **[conséquences** de **[**la politique de la direction**]**$_{\text{Cause}}$**]**$_{\text{Effect}}$" ...

Il s'agit donc d'une situation proche du cas vu section précédente, i.e. de nominalisations qui référant à un des participants, sauf qu'ici on n'a pas polysémie (i.e. il y a polysémie entre achat1="l'acte d'achat" et achat2="les biens achetés", alors que l'on n'a pas cette polysémie pour "cause" ou "conséquence"). Aussi par cohérence, on utilise aussi une annotation de type REFTOPARTICIPANT pour les noms relationnels en emploi référentiel : ils déclenchent le frame, et sont tête d'un des remplisseurs.

6.3 Usage prédicatif versus usage référentiel d'un déclencheur nominal

Dans cette section on aborde une distinction typique dans l'interprétation d'un nom, qui complique le traitement des noms relationnels vu à la section 6.2.

[6] Une alternative suivant la syntaxe aurait été de définir un frame "cause_of_situation_X", à un seul rôle pour X.

- **USAGE PREDICATIF** : soit le nom est utilisé uniquement pour prédiquer une propriété sur un autre élément de la phrase
  - par exemple dans "*L'assassin a utilisé un outil contondant. Cet outil était peut-être un COUTEAU.*" on considère que "un couteau" ne réfère pas à un objet précis, mais indique des propriétés sur le référent de "Cet outil".
- **USAGE REFERENTIEL** : ou bien, le plus fréquemment, le nom réfère directement à une entité (et l'usage de ce nom indique que cette entité en a les propriétés)
  Par exemple dans "*L'assassin a utilisé un COUTEAU*" $\longrightarrow$ "un couteau" sert à référer à un objet (en plus de dire que cet objet a les propriétés de couteau).

L'usage standard du nom est l'usage référentiel, mais l'usage prédicatif est relativement fréquent pour les noms relationnels.

ANNOTATION des usages REFERENTIELS: On considère 2 cas:

- si le déclencheur nominal (cas le plus fréquent) évoque un frame mais ne réfère pas à un des rôles du frame (comme par ex. *cambriolage* ou *similitude*), l'annotation n'a rien de spécial
- mais dans le cas d'un nom référant (en emploi référentiel) à un participant (par exemple un nom relationnel comme *cause* ou bien un des sens de *achat*), alors en emploi référentiel, le déclencheur en question est à la fois déclencheur du frame, et remplisseur d'un certain rôle du frame (plus précisément le déclencheur est la tête syntaxique d'un remplisseur du frame). On dit que le rôle est référé par le déclencheur. C'est une propriété du déclencheur que de déterminer à quel rôle il réfère dans le cas d'un usage référentiel (cf. les exemples (106) et (105) vu à la section précédent 6.2.

ANNOTATION des usages PREDICATIFS: le déclencheur en question est simplement déclencheur du frame. Le rôle qui serait référé dans le cas d'un emploi référentiel peut dans le cas prédicatif soit être présent dans la phrase (et annoté comme tel), soit omis (et il doit alors recevoir un des flags de NULL INSTANTIATION, voir infra). Exemple :

(107) **[**Cette fusion**]**$_{\text{Effect}}$ est une **conséquence** directe de **[**la fin de la guerre froide**]**$_{\text{Cause}}$.

On liste ci-dessous des exemples de cas typiques d'emplois prédicatifs, et quelques exemples d'emplois référentiels. Si vous trouvez un cas qui ne ressemble à aucun de ces exemples, signalez le via un commentaire wiki.

6.3.1 Liste non exhaustive des types d'emplois prédicatifs (où le nom déclencheur n'est pas en outre remplisseur)

Phrase copulative avec attribut nominal : NP1 + verbe être + NP-déclencheur: (le verbe être peut avoir toute sorte d'auxiliaire ou de modal)

(108) **[**Cette situation**]**$_{\text{Effect}}$ est le **résultat** de **[**plusieurs phénomènes qui se télescopent**]**$_{\text{Cause}}$.

NB: il faut que le déclencheur soit directement tête de l'attribut. Dans le cas contraire, on considère que l'emploi est référentiel. Exemple:

(109) Cette situation est l'une des **raisons** de la conduite de la Bundesbank
⟶ on considère le NP "(les) raisons de la conduite de la bundesbank" comme référentiel.

Phrase copulative "inversée" : dans les cas "NP1 + verbe être + NP2", c'est parfois en fait le NP1 qui prédique une propriété sur le NP2 référentiel (on parle de phrase copulative spécificationnelle). Le test pour repérer lequel des 2 NPs est référentiel est l'utilisation de la clivée. Mais pour simplifier ici, au vu des exemples, si le NP1 contient un nom du domaine causalité ⟶ on considère qu'il s'agit forcément d'un emploi prédicatif:

(110) Le **résultat** est [un taux d' épargne élevé , mais qui ne s' accroît plus]$_{\text{Effect}}$ (Cause=DNI)

(111) Le **résultat** est qu'[un décalage sensible se creuse sur le marché]$_{\text{Effect}}$ (Cause=DNI)

ATTENTION: on ne fait cette analyse que pour une phrase copulative (verbe être), dans d'autres cas de type NP1 verbe NP2, même si l'interprétation revient à identifier que NP1 et NP2 réfèrent à la même entité, on va considérer les deux NPs comme référentiels. Exemple:

(112) Le principal **motif** d' inquiétude vient du Japon , actuellement le premier consommateur mondial de métal_blanc ...
⟶ même si la phrase peut s'interpréter comme "le principal motif d'inquiétude EST le Japon...", qui serait une phrase copulative inversée, on colle à ce qui a effectivement été écrit, et on considère que l'emploi est référentiel ⟶ il est le remplisseur du rôle State_of_affairs de Reason (la cause), et on ne capture pas que c'est coréférent à "Japon".

Apposition sans déterminant : le nom déclencheur peut apparaître comme apposition sans déterminant, au sein d'un NP :

(113) Que ce soit l' impossibilité de loger les plus démunis ou , à l' autre bout de la chaîne , [l' importance des encours bancaires dans l' immobilier]$_{\text{Effect}}$, **résultat** d' [imprudences et]$_{\text{Cause}}$ de [spéculations]$_{\text{Cause}}$, rien ne va plus .

Phrase averbale (ou construction de type "connecteur") : nom déclencheur sans déterminant suivi de ":" ou bien ","

Dans les exemples suivants, c'est toute la phrase qui suit le ":" ou la "," qui est le remplisseur:

(114) **Résultat**: [leurs dirigeants forçaient les doses pour pousser les rendements]$_{\text{Effect}}$

(115) **Résultat**, [faute de pluies suffisantes pour les renouveler , leur niveau baisse]$_{\text{Effect}}$

(116) **A l'origine de** [ce désarroi]$_{\text{Effect}}$: [les premiers effets de l' application de la nouvelle PAC]$_{\text{Cause}}$

(117) **Conséquence** : [les services de Mme Véronique Neiertz ... ont décidé de renouveler l' expérience à intervalles réguliers]$_{\text{Effect}}$

Construction de type "avoir pour N":

(118) [Cette solution]$_{\text{Cause}}$ devrait avoir pour **conséquence** probable de [stopper le harcèlement judiciaire]$_{\text{Effect}}$

(119) [les rumeurs]$_{\text{Effect}}$ avaient pour **origine** [des propos au Nouvel Observateur]$_{\text{Cause}}$

Cas figé (mot composé de type préposition / adverbe / conjonction):
Le N déclencheur peut former un composé (comme *à l'origine de*, *en conséquence*) qui n'est pas un nom. On a alors forcément le cas standard où le déclencheur (ici composé) n'est pas en plus remplisseur:

(120) 5704 **À l'origine de** [cette glissade du dollar]$_{\text{Effect}}$, [...] on trouve [deux explications]$_{\text{Cause/NLI}}$.

6.3.2 Exemples d'emplois référentiels

En dehors des cas prédicatifs cités supra, on considère que l'emploi est référentiel, et donc si le nom est relationnel, on annote avec un des rôles rempli par le déclencheur lui-même (plus précisément le constituant dont il est tête). Exemples:

(121) 17761 [Cet abandon de compétence]$_{\text{Cause/NLI}}$ aurait des [**conséquences** très importantes]$_{\text{Effect}}$

(122) Les ventes françaises [en]$_{\text{Cause/NLI}}$ subiront [les **conséquences**]$_{\text{Effect}}$

(123) Et le [**résultat**]$_{\text{Effect}}$ a été conforme à leur attente : c' est devenu un frein à la mobilité .
⟶ attention, par cohérence le remplisseur du rôle Effect est bien "le résultat" ici, et on ne capture pas le lien d'identité avec la phrase derrière les ":".

Attention : dans le cas d'une phrase copulative avec attribut du sujet de type adjectival, ou bien être + locatif, le sujet est forcément référentiel (et donc annoté avec incorporation):

(124) 19466 [Les **conséquences** génétiques de [la technique du transfert d' embryon]$_{\text{Cause}}$]$_{\text{Effect}}$ sont potentiellement immenses.

6.4 Noms quantificateurs

Il existe des NP dont la tête syntaxique est un N qui correspond sémantiquement à une quantification, ne donnant pas le type sémantique du référent du NP global. Dans ces cas on annote quand même comme remplisseur tout le NP, incluant le N quantificateur (que l'accord soit au singulier ou au pluriel) :

(125) [(une partie / la majorité / la plupart ...) des 300 manteaux produits]$_{\text{Goods}}$ (a / ont) été **achetée** par ...

(126) [La bagatelle de 52 , 4 milliards de francs]$_{\text{Money}}$ ont été **versés** au titre de l' une des trois aides.

6.5 Constructions à verbe support

Il s'agit de constructions où un nom prédicatif est employé comme dépendant d'un verbe, et d'une part le verbe n'est pas employé dans un de ses sens habituels (comme "prêter" dans "prêter attention") et/ou c'est un verbe à sens très général (faire, avoir...). D'autre part le sujet du verbe est forcément un dépendant du nom prédicatif.
On a envisagé d'annoter les verbes support des noms prédicatifs, mais finalement dans un premier temps, on se concentre sur l'annotation des arguments du nom prédicatif (comme remplisseurs) sans noter le verbe support.

Une seule exception: lors de l'annotation des verbes qui sont déclencheurs de frames de nos domaines, si certaines occurrences sont hors-domaine mais correspondent clairement à un verbe support, on utilise le frame FR_Support (au lieu de Other_sense).

7 Annotation des remplisseurs: points divers

7.1 Remplisseurs collectifs versus rôle _1 / rôle _2

Pour les déclencheurs admettant l'alternance entre 2 rôles réalisés comme 2 constituants différents versus les 2 rôles réalisés comme un seul NP collectif, on reprend la solution du FrameNet anglais, où l'on dispose dans le frame de deux rôles distincts XXX_1 et XXX_2 et d'un rôle pour XXXs pour le collectif. Par exemple, dans Be_in_agreement_on_assessment, on a les rôles Cognizer_1, Cognizer_2 et Cognizers:

(127) réalisation comme NP collectif : [Les enfants]$_{\text{Cognizers}}$ sont d'accord (entre eux)

(128) réalisation comme 2 NP : [Paul]$_{\text{Cognizer_1}}$ est d'accord avec [Marie]$_{\text{Cognizer_2}}$

(129) [Les enfants]$_{\text{Cognizer_1}}$ sont d'accord avec [le professeur_2]$_{\text{Cognizer}}$.

(130) réalisation avec un seul NP, l'autre étant omis car présent dans le contexte: Le professeur propose de faire une pause. Les [Les enfants]$_{\text{Cognizer_1}}$ sont d'accord.

Pour choisir qui de X ou Y est XXX_1 : en général, le rôle XXX_1 est le rôle le plus saillant syntaxiquement (en particulier le sujet).

Attention, dans le cas d'un seul NP coordonné, on ne descend pas au sein de la coordination pour trouver les rôles XXX_1 et XXX_2, on utilise le rôle XXXs, cf. on peut avoir des coordinations n-aires (si compatible avec l'interprétation):

(131) 13267: Pendant la décennie 80 , le **commerce** entre [la colonie et Pékin]$_{\text{Participants}}$ a été multiplié par 8 ...

7.2 Astuce dans le cas de déclencheurs qui dépendent syntaxiquement d'un de leur arguments sémantiques

Dans le cas de déclencheurs qui dépendent syntaxiquement d'un de leurs remplisseurs, on est confronté au problème que ce remplisseur ne constitue pas un syntagme à lui tout seul. On utilise une astuce pour accélérer l'annotation, corrigée par post-traitement automatique.

7.2.1 Adjectifs et participes épithètes

Un adjectif épithète ou un participe épithète dépend syntaxiquement d'un élément (en général un nom) qui correspond à un de leurs arguments sémantiques, et donc en général à un de leurs remplisseurs:

(132) la **jolie** balle $\longrightarrow$ *balle* est argument sémantique de *jolie*

(133) l'appartement **acheté** par Deviers $\longrightarrow$ *l'appartement* est argument sémantique de *acheté*

(134) les personnes **voulant** annoter des corpus $\longrightarrow$ *les personnes* argument sémantique de *voulant*

On a la même situation typiquement pour une préposition ou une conjonction :

(135) il a fait cela **avant** la conférence ⟶ "*il a fait cela*" = le Trajector du frame FR_Temporal_relation déclenché par "avant"

(136) il a fait cela **parce que** Paul le lui a demandé ⟶ "*il a fait cela*" remplit le rôle Effect du frame Causation déclenché par *parce que*

Mais en syntaxe, la préposition ou conjonction modifie le verbe principal, et la principale sans ce modifieur ne forme pas un constituant (*il a fait cela* ne forme pas un constituant. Ou bien dit en termes de dépendances: il n'y a pas de noeud qui domine le remplisseur effectif (i.e. la principale sans le modifieur introduit par la prep ou la conjonction.

Par exemple il n'y a pas de noeud pour "il a fait cela" ⟶ le noeud dominant la tête "fait" inclut le modifieur "avant la conférence" / "parce que Paul le lui a demandé".

⟶ Pour accélérer l'annotation manuelle, on code comme remplisseur le noeud dominant le nom modifié / le verbe modifié. I.e. dans les exemples ci-dessus : le noeud NC pour "balle" / appartement / personnes, ou le noeud VPP pour "fait".

Post-traitement: On appliquera un post-traitement automatique corrigeant cela: si le déclencheur (plus tout le sous-arbre du déclencheur) est contenu dans un remplisseur, sans en être la tête syntaxique[7] , alors on interprétera en réalité le remplisseur comme étant celui annoté moins le déclencheur et son sous-arbre.

7.2.2 Incises

On utilise également ce raccourci dans le cas de déclencheurs verbaux qui sont tête d'une incise :

(137) 14023: Si un partenariat avec l' Afrique est nécessaire , **croient** les auteurs , c' est qu' elle est " la terre d' élection d' amitiés renforcées .

⟶ "croient les auteurs" est une incise, et "croient" est rattaché au verbe principal (ici le verbe "est"). Le verbe "croient" déclenche FR_Awareness-certainty-opinion et le rôle Content précis est toute la phrase sauf "croient les auteurs" ⟶ on annote par commodité sur le noeud V dominant le verbe principal "est", c'est le post-traitement qui s'appliquera et retirera l'incise.

ATTENTION: Ce raccourci ne doit pas être utilisé pour les cas où le verbe de l'incise est la tête du noeud qu'on souhaite annoter (cas d'erreurs dans l'annotation syntaxique sous-jacente). En effet, lorsqu'un rôle est annoté comme rempli par le noeud dominant directement le déclencheur de son frame, on interprétera qu'il s'agit d'un cas d'incorporation de rôle au déclencheur. L'heuristique de suppression du déclencheur (et de ses dépendants) du span du remplisseur ne sera donc pas appliquée.

ATTENTION: Ce raccourci ne doit pas non plus être utilisé si l'heuristique donne un remplisseur mal formé. Par exemple:

(138) 12995: celle - ci est loin d' être une réalité , comme en **témoignent** notamment les difficultés de la négociation sur le GATT .

[7] Restriction nécessaire, cf. la représentation des noms relationnels en emploi référentiel section 6.2).

On veut coder "celle - ci est loin d' être une réalité" comme rôle Proposition/FULLANT (et "en" est le remplisseur ANAPH). MAIS on ne peut pas utiliser le noeud V dominant le verbe principal "est" cf. cela couvrirait toute la phrase et si on enlève (automatiquement) le déclencheur et son sous-arbre on obtiendra au final un remplisseur mal formé (car incluant le "comme" : on aurait Proposition="celle - ci est loin d' être une réalité , comme"). Il faut alors annoter précisément le rôle Proposition (annotation "multi-pattes") pour obtenir exactement "celle - ci est loin d' être une réalité".

7.3 Remplisseur adjectival synonyme d'un PP

On a parfois la mention d'un remplisseur sous une forme elliptique, au travers d'un adjectif qui caractérise le remplisseur. ⟶ contrairement au choix de départ, on décide d'annoter ces adjectifs comme remplisseur, avec un flag "ELLIPTIC". Par exemple :

(139) les **achats** [chinois]$_{\text{Buyer/ELLIPTIC}}$ (⟶ on comprend "les achats des Chinois")

(140) 19201: ...les crédits affectés dans le budget 1990 de Paris aux **acquisitions** [foncières]$_{\text{Goods/ELLIPTIC}}$ ont subi une coupe ... (⟶ on comprend : les acquisitions de biens fonciers)

(141) 5140 Tout au long de la **discussion** [parlementaire]$_{\text{Interlocutors/ELLIPTIC}}$

Attention: le nom du flag ELLIPTIC est assez mal choisi. On ne l'utilise en fait pas pour les cas d'ellipse du nom comme:

(142) 54 % des fonctionnaires ont un second emploi et près de 30 % se livrent à une forme de commerce sur leur lieu de travail
⟶ pas de flag ELLIPTIC sur le participant "près de 30%"

On l'utilise bien en revanche pour des cas où le remplisseur a pour tête un nom qui a un type sémantique différent de celui attendu par le rôle, et que l'interprétation nécessite d'ajouter une tête au remplisseur:

(143) 286 Il n'en **justifie** pas moins [le décret]$_{\text{Eventuality/ELLIPTIC}}$ ···
⟶ "justifie" sélectionne une éventualité, ici on reconstruit "justifie (la prise du) décret"

8 Phénomènes de portée

8.1 remplisseurs phrastiques/infinitivaux: gestion des verbes introducteurs de complétives/d'infinitives

NB: cette section utilise des exemples contenant des déclencheurs du domaine Causalité. Certains (*parce que*) sont ambigus entre les frames Causation et Evidence (cf. le guide spécifique de désambiguisation Causation / Evidence). Dans cette section, on laisse de côté cette ambiguïté, en ne prenant que des exemples où *parce que* sera annoté avec le frame Causation (on utilise également des exemples avec le déclencheur *grâce à*, qui n'évoque que Causation et jamais Evidence).
On considère les cas de remplisseurs phrastiques ou infinitivaux contenant ou introduit par un verbe enchâssant. Le verbe enchâssant (souligné dans les exemples de cette section) est selon l'interprétation, inclus (exemple (144)) ou pas (exemple (145)) au remplisseur.

(144) [Paul pense que les enfants sont rentrés]$_{\text{Effect}}$ **parce qu'** [il a vu leur cartable dans l'entrée]$_{\text{Cause}}$.

(145) Paul pense que [les enfants sont rentrés]$_{\text{Effect}}$ **parce qu'** [ils avaient faim]$_{\text{Cause}}$.

En (144), c'est parce qu'il a vu leur cartable que Paul pense que les enfants sont rentrés, on inclut le verbe enchâssant *penser* au role Effect. En (145) au contraire, la portée est sur le verbe enchâssé uniquement: Paul pense que c'est parce qu'ils avaient faim que les enfants sont rentrés. Ce cas est le plus fréquent.
Dans ces deux exemples, l'ambiguïté est normalement levée par l'analyse syntaxique : *parce que* est normalement rattaché à *pense* en (144) et à *rentrés* en (145). Mais attention : on ne doit pas se fier aveuglément à la structure syntaxique proposée, qui peut comporter des erreurs: il vous faut annoter le remplisseur en fonction de votre interprétation de la phrase. En outre, on peut avoir structurellement un rattachement obligatoire du déclencheur sur le verbe introducteur, mais avec une portée du déclencheur sur le verbe enchâssé. C'est en particulier le cas en cas d'antéposition de la subordonnée introduite par le déclencheur:

8.1.1 Portée sémantique sur verbe enchâssé versus portée sur verbe introducteur

Portée sur verbe enchâssé :

(146) **Grâce à** [ces mesures]$_{\text{Cause}}$, [les intervenants]$_{\text{Effect}}$ déclarent [avoir économisé du temps]$_{\text{Effect}}$.

(147) **Grâce à** [ces mesures]$_{\text{Cause}}$, les intervenants déclarent qu'[ils ont économisé du temps]$_{\text{Effect}}$. $\longrightarrow$ l'interprétation est claire : les intervenants déclarent (qu'ils ont économisé grâce à ces mesures).

À contraster avec portée sur verbe introducteur :

(148) **Grâce à** [ces mesures]$_{\text{Cause}}$, [les intervenants sont autorisés à économiser du temps]$_{\text{Effect}}$. $\longrightarrow$ l'interprétation est claire : c'est grâce à ces mesure que les intervenants ont le droit d'économiser.

On annote les cas de portée sur verbe introducteur de façon régulière (pas de problème particulier). En revanche, pour les cas de portée sur V enchâssé, **il faut veiller à exclure du remplisseur le verbe introducteur**. On annote comme remplisseur l'infinitive ou la complétive enchâssée, et en outre dans le cas de l'infinitive, on inclut le sujet de l'infinitif (cf. infra section 9).

8.1.2 Difficultés de détermination de portée

Pour certains verbes introducteurs, trancher entre portée sur verbe introducteur et portée sur verbe enchâssé peut s'avérer assez difficile. C'est en particulier le cas pour les verbes épistémiques (exprimant un degré de croyance, *penser, croire...*)

(149) **Grâce à** ces mesures, les intervenants pensent avoir économisé du temps. $\longrightarrow$ l'ambiguïté est soit ce sont les mesures qui causent qu'ils pensent avoir économisé, soit ils pensent que ce sont les mesures qui leur ont fait économiser du temps.

On peut construire des exemples forçant clairement la portée. Par exemple ci-dessous, sur verbe épistémique introducteur:

(150) **Grâce aux** [mesures de propagande et de manipulation de la part des organisateurs]$_{\text{Cause}}$, [les intervenants pensent avoir économisé]$_{\text{Effect}}$.

Selon les différents types de verbe introducteur listés ci-dessous, on demande de désambiguiser en contexte la portée, ou bien on recommande une annotation systématique.

(151) **Grâce à** [ces mesures]$_{\text{Cause}}$, [les intervenants]$_{\text{Effect}}$ pensent [avoir économisé du temps]$_{\text{Effect}}$.

8.1.3 Gestion dans Asfalda : classes de verbes introducteurs

Verbes de communication/d'influence sur autrui
Pour les verbes dénotant la communication verbale : il ne semble pas difficile de faire effectivement la désambiguisation de la portée (cf. supra l'annotation différente pour (147) versus (148). On a en général plutôt un remplisseur n'incluant pas le verbe enchâssant de communication.
Liste non exhaustive de v de discours rapporté (cf. le guide annodis p.6) : clamer, proclamer, insister, soutenir, déclarer, affirmer, faire valoir, revendiquer, imposer, maintenir, témoigner, attester, confirmer, assurer, garantir, certifier, conclure, démontrer, montrer, souligner, informer, noter, révéler, annoncer, expliquer, dire, rapporter, raconter, narrer, relater, indiquer....
persuader, convaincre
Exemple:

(152) 2813: [Un dirigeant de la CG2A]$_{\text{Effect}}$, lors_de l' attribution du marché , a ainsi déclaré au juge [avoir acquitté de semblables factures au mois de février 1992]$_{\text{Effect}}$ " **parce que** [la commission d' appel d' offres se réunissait la semaine suivante]$_{\text{Cause}}$ " .

Verbes aspectuels
Pour les verbes aspectuels (=indiquant le début, la fin, la continuation, le futur proche d'une éventualité : commencer, finir, continuer, cesser, aller etc...), on décide de ne pas désambiguiser entre portée sur verbe enchâssé ou sur verbe introducteur: on inclut systématiquement le verbe aspectuel dans le remplisseur (il en est la tête). Donc un remplisseur de tête verbe aspectuel sera en fait sous-spécifié quant à la portée.
Exemples:

(153) 10936': **Grâce à** [cet accord]$_{\text{Cause}}$, [les hongrois vont élargir leur activité en intégrant un réseau européen]$_{\text{Effect}}$.

(154) 15690: [L' accès grandissant des femmes ingénieurs au milieu professionnel continuera sans doute de leur profiter]$_{\text{Effect}}$ **grâce à** [la pénurie d' ingénieurs diplômés]$_{\text{Cause}}$.

(155) 15827: **Grâce au** [président Delaporte]$_{\text{Actor}}$, [elle commence à s' ouvrir à l' extérieur]$_{\text{Effect}}$.

Verbes de tentative/ promesse / factifs
(essayer de , s'appliquer à ...)
(oser, avoir le toupet de, se permettre de)
⟶ il semble que l'on aura toujours portée sur verbe introducteur.

(156) 1813: Cette demande a été rejetée et [il est en prison]$_{\text{Effect}}$ **parce qu'** [il a osé , il y a quelque temps , proposer la recherche d' une solution politique au problème

kurde]$_{\text{Cause}}$.

Verbes modaux (pouvoir, devoir) / degré de croyance de la part de l'auteur (sembler, paraître)
X peut Vinf (épistémique: il est probable que Vinf/ déontique: X a la possibilité morale ou physique de Vinf)

X doit Vinf (épistémique: il est vraisemblable que Vinf / déontique: X a l'obligation morale ou physique de Vinf)

X semble Vinf / il semble que P

Bien qu'on puisse souvent avoir portée sur le verbe enchassé, on inclut systématiquement ces verbes introducteurs au remplisseur, sans chercher à désambiguiser la portée, cf. cela s'avère trop difficile.

(157) 4394: [Prague devrait devenir la plaque tournante aérienne d' Europe centrale]$_{\text{Effect}}$ **grâce à** [sa vocation touristique et à ses atouts industriels]$_{\text{Cause}}$

(158) 5974: **Grâce à** [l' appui technique de ces deux organismes]$_{\text{Cause}}$, [les choses devraient aller plus vite]$_{\text{Effect}}$ ⟶sur enchâssé : il est probable que (grâce à l' appui technique ..., les choses aillent plus vite)

(159) **Grâce à** [ces mesures]$_{\text{Cause}}$, [les intervenants]$_{\text{Effect}}$ doivent [avoir économisé du temps]$_{\text{Effect}}$.

Remarque: on peut faire le parallèle avec le cas d'un adverbe épistémique (indiquant un degré de croyance), où on suit la syntaxe, en incluant l'adverbe dès lors que le verbe qu'il modifie est inclus au remplisseur. On ne représente pas que sémantiquement l'adverbe porte sur toute la relation de causalité.

(160) **Grâce à** [ces mesures]$_{\text{Cause}}$, [les intervenants ont certainement économisé du temps]$_{\text{Effect}}$.

Verbes de degré de croyance d'un participant
X (pense / croit / imagine / suppose / estime / juge / a l'impression / devine / suppute / s'attend à ce que / anticipe / sent / est convaincu / est persuadé / savoir / être sûr / être certain...) que P

Il semble à X que P

X (pense / croit / imagine / suppose / estime / juge / a l'impression de / devine / suppute / s'attend à / anticipe / sent / est convaincu de / est persuadé de) Vinf

Exemples:

(161) 13725: Le groupe s' attend à [une hausse de 50 % de ses résultats]$_{\text{Effect}}$ **grâce à** [des plus - values de cession , notamment sur Sopharga]$_{\text{Cause}}$.

(162) 1701 modifié: Je pense qu' il s' agit d' un objectif important **parce que** ... nous n' avons d' autre choix que d' aller là où ils se trouvent . ⟶ le verbe épistémique introduit une opinion, ce cas est annoté avec le frame Evidence (voir guide de désamb Causation / Evidence)

Verbes d'orientation
Sentiment d'un cognizer par rapport à la réalisation future / ou passée d'une éventualité (souhaiter, espérer, vouloir, désirer, craindre, redouter, regretter, apprécier ...)
On fait l'interprétation en contexte, en privilégiant l'annotation n'incluant pas le verbe introducteur au remplisseur.

(163) 10132: **Grâce à** [ces mesures]$_{\text{Cause}}$, les intervenants espèrent enfin revoir [l'économie repartir]$_{\text{Effect}}$.

8.2 Portée de la négation et des modaux dans un remplisseur

On ne gère pas la portée de la négation au sein d'un remplisseur. Par exemple:

(164) 15107: En cas de troubles , guérillas ou autres soulèvements , la production n' est pas perturbée **grâce au** stockage flottant
$\longrightarrow$ = le stockage flottant cause la non perturbation de la production

versus

(165) 15107 modifié : La production n'est pas perturbée **à cause du** stockage flottant mais à cause de l'invasion par les sardines
$\longrightarrow$ = il n'est pas vrai que (le stockage flottant cause la perturbation de la production)

On ne traite pas cette subtilité de portée de la négation : dans les 2 cas, le remplisseur intègre la marque de négation (Effect= la production n'est pas perturbée).

D'une manière générale on ne gère pas non plus la portée d'un modal. Exemple interagissant avec un modal :

(166) 1667: ... ils ne peuvent être punis **parce qu'** ils viennent volontairement en aide au Kosovo .
lecture 1 : il n'est pas possible qu' (ils soient punis parce qu'ils viennent en aide au Kosovo)
versus
lecture 2 (erronée) : (il n'est pas possible qu'il soient punis) parce qu'ils viennent en aide au Kosovo

On ne gère pas la distinction de portée du modal, et on annote Effect="ils ne peuvent être punis".

9 Saturation des remplisseurs prédicatifs et divergence entre tête syntaxique et tête sémantique

Lorsqu'un rôle correspond sémantiquement à une éventualité (un évènement, un état), son remplisseur sera constitué d'une tête sémantique (le mot décrivant l'éventualité: un verbe, un nom prédicatif, un adjectif...), et des dépendants sémantiques de cette tête.
Le cas le plus simple est un remplisseur de type phrase (167) ou bien de type NP complètement saturé (168) (i.e. avec tous ses compléments).

(167) Paul a **appris** à 17h que [Pierre partirait bientôt]$_{\text{Phenomenon}}$

(168) Paul a **appris** à 17h [le départ imminent de Pierre]$_{\text{Phenomenon}}$

D'une part on a alors, sauf cas d'extraction, un remplisseur sans discontinuité, et d'autre part la tête *syntaxique* du remplisseur (le verbe partirait en (167) ou le nom départ en (168)) correspond bien à la tête *sémantique* du remplisseur.[8] On a alors la situation simple où on peut annoter le remplisseur simplement en pointant le noeud de sa tête à la fois sémantique et syntaxique.
Mais nous détaillons maintenant:

- des cas où le remplisseur est discontinu: il faut aller récupérer d'autres constituants pour "saturer" le remplisseur de tous ses arguments,
- des cas où cette saturation amène à des situations où la tête syntaxique d'un remplisseur n'est pas la tête sémantique voulue, cas pour lesquels on introduit donc un marquage spécial.

9.1 Saturation des remplisseurs infinitivaux

Lorsque la tête sémantique du remplisseur est un verbe non conjugué, le sujet de cette tête sém ne peut pas être réalisé comme un dépendant syntaxique de la tête sém du remplisseur. C'est typiquement le cas pour un remplisseur infinitival : le sujet de l'infinitif est soit non réalisé, soit réalisé non localement. Si la sémantique du rôle correspond bien à la prédication complète, alors on choisit d'inclure au remplisseur le sujet de l'infinitif.

(169) Anna **dit** que [Paul est parti à 23h]$_{\text{Message}}$

(170) Paul **dit** être parti à 23h $\longrightarrow$ Message = "Paul" + "être parti à 23h"

Attention cependant, si la définition du rôle n'inclut pas le sujet, alors on n'applique pas cette règle d'inclusion du sujet au remplisseur. C'est typiquement le cas pour les rôles Action des frames XXX.to_do. Par exemple dans Remembering_to_do, le frame prévoit 2 rôles Cognizer et Action, et pour tous les déclencheurs du frame, l'action a forcément pour sujet le Cognizer. Le périmètre du rôle Action est toute la prédication sauf le sujet:

(171) Pierre a **oublié** de [fermer le gaz]$_{\text{Action}}$ $\longrightarrow$ Cognizer = Pierre / Action= fermer le gaz

9.2 Saturation des remplisseurs adjectivaux

Dans le cas d'un remplisseur dont la tête *sémantique* est un adjectif, on doit lister les 3 positions syntaxiques principales pour un adjectif

- épithète : *une enfance heureuse*
- attribut du sujet : *mon enfance fut heureuse*
- attribut de l'objet : *il trouve son enfance heureuse*

Dans tous ces cas, l'argument sémantique premier de l'adjectif (qu'on appelle parfois le "sujet" de l'adjectif) n'est pas un dépendant syntaxique de l'adjectif (*enfance* dans les 3 exemples ci-dessus ne dépend pas syntaxiquement de *heureuse*).

[8] Dans tous les exemples de cette section, la tête sémantique du remplisseur est soulignée.

Comme pour les infinitifs, on choisit de saturer les remplisseurs, en y incluant les sujets des adjectifs.
Plus précisément: pour un attribut du sujet, dépendant donc d'un verbe d'état, on annote comme remplisseur simplement toute la phrase du verbe d'état.

(172) Anna a **dit** que [Paul **était** content de la fête]$_{\text{Message}}$.

Pour un attribut de l'objet, comme en (173), si l'analyse syntaxique ne comporte pas d'erreur, on aura dans ce cas un remplisseur constitué de 2 sous-arbres syntaxiques : un pour l'objet direct ("Pierre") et un pour l'attribut ("content"), qui est bien la tête sémantique de l'attribut. (Rem: dans ce cas le post-traitement (cf. section **??**) donnera l'adjectif comme tête syntaxique et tête sémantique du remplisseur).

(173) Paul **trouve** Pierre intelligent
⟶ Content = "Pierre" + "intelligent"

9.2.1 Remplisseur adjectif ou participe épithète et divergence entre tête syntaxique et sémantique

Le plus fréquemment l'adjectif est épithète. On se retrouve alors dans la situation où l'élément que l'adjectif modifie est le gouverneur syntaxique de l'adjectif, mais en est un argument sémantique. Si l'adjectif correspond à la tête sémantique d'un remplisseur, et que l'on sature ce remplisseur en récupérant le sujet de l'adjectif, alors le remplisseur sera tout le NP.
Ainsi pour un NP remplisseur, il faut distinguer le cas classique où c'est bien l'entité dénotée par le NP qui est sémantiquement le remplisseur, comme infra en (174), (tête sém = la tête synt = *enfance*) :

(174) Paul a longuement **évoqué** [son enfance heureuse]$_{\text{Topic}}$.

du cas où la tête sémantique est une prédication interne au NP : un adjectif (ou un participe) épithète, et c'est en saturant celui-ci que l'on récupère tout un NP. Ainsi en (175), le rôle Effect est sémantiquement "son enfance a été heureuse", et rôle Cause = "son enfance a été paisible", ce que l'on résume par "la tête sémantique du remplisseur est "heureuse"":

(175) Paul a évoqué longuement [son enfance heureuse]$_{\text{Effect}}$ **car** paisible.

Autre exemple réel cette fois-ci: Dans l'exemple ci-dessous (6875), le rôle Effect est sémantiquement la prédication équivalente à "l'écart de taux (était) très important il y a quelques années", mais syntaxiquement en incluant le "sujet" de "important" on inclut le NP de tête "écart", dont "important" est syntaxiquement un dépendant modifieur. On utilise une astuce pour bien repérer les cas où le remplisseur a une tête syntaxique (ici "écart") qui ne correspond pas à la tête sémantique (ici "important") : on annote sous Salto un remplisseur à 2 pattes une sur le noeud dominant le remplisseur complet (ici le NC dominant "écart") et en plus une patte sur le token qui est en fait la tête sémantique prédicative (ici le token "important").

(176) 6875: ..., [l' écart de taux, très important il y a quelques années]$_{\text{Effect}}$, **car** il traduisait les inquiétudes des milieux internationaux sur la santé du franc , s' est très fortement réduit ...
⟶ le rôle Effect a une tête synt (écart) distincte de la tête sém ⟶ on doit marquer la tête sém. (une patte en plus sur le token "important").

On a des situations similaires avec des participes, présents ou passés. Dans l'exemple ci-dessous, le remplisseur complet est "un PDG de thomson CSA, M. Gomez, soupçonné... blancs", il domine syntaxiquement à la fois le déclencheur soupçonné, et la tête sém du remplisseur correspondant au contenu du soupçon (tête sém = s'intéresser)

(177) 16016: Conjonction d' intérêts d' un PDG de Thomson SA , M. Gomez , **soupçonné** de ne s' être jamais intéressé à la vie de sa filiale de " produits blancs " (...)
⟶ rôle Content = le noeud NC dominant PDG, plus 2 pattes sur les tokens "s'" et "intéressé"

On peut avoir combinaison de récupérations de sujet. Dans l'exemple suivant, on a ce phénomène pour le rôle Effect (remplisseur dont la tête synt est "ensemble", et la tête sém est "grave"), et en outre, pour le rôle Cause, qui est un participe ("alimentée"), on va récupérer le sujet ("ensemble"). 11348: L' ensemble du système financier mondial subit aussi une crise immobilière bien plus grave que les précédentes CAR alimentée par la déréglementation des années 80 . ⟶ Causation, Effect="une crise immobilière .... précédentes" + patte sur "grave" ET AUSSI, rôle Cause ="une crise immobilière ... précédentes" + une patte sur "alimentée"

(178) 14758: Le chômage de longue durée , jusqu' alors éradiqué **grâce à** un suivi social systématique, ...
⟶ Effect = noeud NC dominant "chômage" , plus une patte sur "éradiqué"

Dans l'exemple suivant, on va récupérer le sujet de "acteur" et de "courant" :

(179) 5553: Acteur moyen de l' industrie des télécommunications et **donc** courant le risque d' être marginalisée , Matra Communication , filiale spécialisée du groupe Matra , a décidé de nouer une alliance
⟶ Effect= "Matra communication filiale...Matra" + "acteur moyen de l'industrie de télécoms" / Cause="Matra communication filiale...Matra" + "courant le risque d' être marginalisée"

On a une situation différente avec l'exemple suivant.

(180) 7158: De leur côté , les équipementiers européens ont fait des efforts pour adapter le système en réduisant son volume (30 litres au lieu de 60) , sa sophistication et **donc** son coût .

On peut reconstruire qu'une cause (les équipementiers ont réduit le volume et la sophistication du système) a causé un effet (les équipementiers ont réduit le coût de ce système). D'une part pour la cause, la tête sém ("réduisant") n'a pas son sujet localement (⟶ on va récupérer "les équipementiers" : rôle Cause = "les équipementiers" + "réduisant son volume ... sa sophistication"). Et d'autre part, pour l'effet, on a une coordination elliptique : l'objet direct du prédicat "réduisant" est coordonné, le prédicat n'est pas répété. On doit par cohérence récupérer Effect = "les équipementiers" + "réduisant" + "son coût". C'est-à-dire que l'on doit récupérer non seulement le sujet du prédicat remplisseur, mais le prédicat lui-même ...

10 Bibliographie

References

Anne Abeillé and Nicolas Barrier. 2004. Enriching a french treebank. In *Proc. of LREC'04*.

C. F. Baker, C. J. Fillmore, and J. B. Lowe. 1998. The Berkeley FrameNet project. In *COLING-ACL '98: Proc. of the Conference*, pages 86–90.

A. Burchardt, K. Erk, A. Frank, A. Kowalski, S. Padó, and M. Pinkal. 2006. SALTO – A Versatile Multi-Level Annotation Tool. In *Proc. of LREC'06*.

M. Candito and D. Seddah. 2012. Le corpus sequoia : annotation syntaxique et exploitation pour l'adaptation d'analyseur par pont lexical. In *Proc. of TALN 2012 (in French)*.

M. Candito, P. Amsili, L. Barque, F. Benamara, Ga. de Chalendar, M. Djemaa, P. Haas, R. Huyghe, Y. Y. Mathieu, P. Muller, B. Sagot, and L. Vieu. 2014. Developing a French FrameNet: Methodology and first results. In *Proc. of LREC'14*.

C. J. Fillmore. 2007. Valency issues in framenet. In Thomas Herbst and Katrin Gotz-Votteler, editors, *Valency: theoretical, descriptive and cognitive issues*, volume 187 of *Trends in Linguistics*, pages 129–162. Mouton de Gruyter.

J. Ruppenhofer, M. Ellsworth, M. R.L. Petruck, C. R. Johnson, and J. Scheffczyk. 2006. *FrameNet II: Extended Theory and Practice*. International Computer Science Institute, Berkeley, California. Distributed with the FrameNet data.

D. Seddah, R. Tsarfaty, S. Kübler, M. Candito, J. Choi, R. Farkas, J. Foster, I. Goenaga, K. Gojenola, Y. Goldberg, S. Green, N. Habash, M. Kuhlmann, W. Maier, J. Nivre, A. Przepiorkowski, R. Roth, W. Seeker, Y. Versley, V. Vincze, M. Woliński, A. Wróblewska, and E. Villemonte de la Clérgerie. 2013. Overview of the SPMRL 2013 Shared Task: A Cross-Framework Evaluation of Parsing Morphologically Rich Languages. In *Proc. of the 4th Workshop on Statistical Parsing of Morphologically Rich Languages: Shared Task*.

ANNEXE C

LEXIQUE QUANTIFIÉ EXTRAIT DES ANNOTATIONS DE QUELQUES CADRES

Nous présentons ici à titre d'exemple le lexique quantifié extrait des annotations de quatre cadres lexicaux du French FrameNet (un de chaque domaine). L'extrait fourni est au format cadre par cadre, qui permet de visualiser le lexique d'un cadre, mais nous proposons aussi un format lemme par lemme, qui permet de visualiser la polysémie annotée pour un lemme donné. Le lexique correspondant à l'intégralité des annotations de la ressource est visualisable dans ses deux formats à partir de la page d'accueil des données sur le site du French FrameNet : `http://asfalda.linguist.univ-paris-diderot.fr/frameIndex.xml`.

```
<?xml version="1.0" encoding="utf-8" ?>
<!DOCTYPE FrameNetFEN SYSTEM "FrameNetFEN-1.0.dtd">
<FrameNetFEN maxluid="0">
   <FRAME name="Commerce_buy" annotated_occ="587" domains="COM_TRANS" frameID="171">
      <FE name="Buyer"/>
      <FE name="Goods"/>
      <FE name="Money"/>
      <FE name="Rate"/>
      <FE name="Seller"/>
      <LU name="achat.n" annotated_occ="133" cat="n" frame="Commerce_buy" lemma="achat"/>
      <LU name="acheter.v" annotated_occ="91" cat="v" frame="Commerce_buy" lemma="acheter"/>
      <LU name="acquisition.n" annotated_occ="63" cat="n" frame="Commerce_buy" lemma="
          acquisition"/>
      <LU name="acquérir.v" annotated_occ="68" cat="v" frame="Commerce_buy" lemma="acquérir"/>
      <LU name="bonification.n" annotated_occ="3" cat="n" frame="Commerce_buy" lemma="
          bonification"/>
      <LU name="bonifier.v" annotated_occ="1" cat="v" frame="Commerce_buy" lemma="bonifier"/>
      <LU name="faire␣ses␣courses.v" annotated_occ="1" cat="v" frame="Commerce_buy" lemma="
          faire␣ses␣courses"/>
      <LU name="nationalisation.n" annotated_occ="12" cat="n" frame="Commerce_buy" lemma="
          nationalisation"/>
      <LU name="nationaliser.v" annotated_occ="16" cat="v" frame="Commerce_buy" lemma="
          nationaliser"/>
      <LU name="rachat.n" annotated_occ="53" cat="n" frame="Commerce_buy" lemma="rachat"/>
      <LU name="racheter.v" annotated_occ="47" cat="v" frame="Commerce_buy" lemma="racheter"/>
      <LU name="reprendre.v" annotated_occ="38" cat="v" frame="Commerce_buy" lemma="reprendre"/
          >
      <LU name="reprise.n" annotated_occ="26" cat="n" frame="Commerce_buy" lemma="reprise"/>
      <LU name="régler.v" annotated_occ="3" cat="v" frame="Commerce_buy" lemma="régler"/>
      <LU name="rémunération.n" annotated_occ="19" cat="n" frame="Commerce_buy" lemma="rémuné
          ration"/>
```

```
    <LU name="rémunérer.v" annotated_occ="13" cat="v" frame="Commerce_buy" lemma="rémunérer"/
        >
</FRAME>
<FRAME name="FR_Becoming_aware" annotated_occ="297" domains="COG_POS" frameID="10086">
    <FE name="Cognizer"/>
    <FE name="Evidence"/>
    <FE name="Phenomenon"/>
    <FE name="Topic"/>
    <LU name="comprendre.v" annotated_occ="13" cat="v" frame="FR_Becoming_aware" lemma="
        comprendre"/>
    <LU name="constat.n" annotated_occ="11" cat="n" frame="FR_Becoming_aware" lemma="constat"
        />
    <LU name="constatation.n" annotated_occ="3" cat="n" frame="FR_Becoming_aware" lemma="
        constatation"/>
    <LU name="constater.v" annotated_occ="50" cat="v" frame="FR_Becoming_aware" lemma="
        constater"/>
    <LU name="déceler.v" annotated_occ="4" cat="v" frame="FR_Becoming_aware" lemma="déceler"/
        >
    <LU name="découverte.n" annotated_occ="16" cat="n" frame="FR_Becoming_aware" lemma="dé
        couverte"/>
    <LU name="découvrir.v" annotated_occ="38" cat="v" frame="FR_Becoming_aware" lemma="dé
        couvrir"/>
    <LU name="détecter.v" annotated_occ="3" cat="v" frame="FR_Becoming_aware" lemma="détecter
        "/>
    <LU name="entrevoir.v" annotated_occ="7" cat="v" frame="FR_Becoming_aware" lemma="
        entrevoir"/>
    <LU name="noter.v" annotated_occ="27" cat="v" frame="FR_Becoming_aware" lemma="noter"/>
    <LU name="observer.v" annotated_occ="68" cat="v" frame="FR_Becoming_aware" lemma="
        observer"/>
    <LU name="prise␣de␣conscience.n" annotated_occ="6" cat="n" frame="FR_Becoming_aware"
        lemma="prise␣de␣conscience"/>
    <LU name="relever.v" annotated_occ="14" cat="v" frame="FR_Becoming_aware" lemma="relever"
        />
    <LU name="remarquer.v" annotated_occ="15" cat="v" frame="FR_Becoming_aware" lemma="
        remarquer"/>
    <LU name="repérer.v" annotated_occ="1" cat="v" frame="FR_Becoming_aware" lemma="repérer"/
        >
    <LU name="s'apercevoir.v" annotated_occ="9" cat="v" frame="FR_Becoming_aware" lemma="s'
        apercevoir"/>
    <LU name="se␣rendre␣compte.n" annotated_occ="9" cat="n" frame="FR_Becoming_aware" lemma="
        se␣rendre␣compte"/>
    <LU name="se␣rendre␣compte.v" annotated_occ="1" cat="v" frame="FR_Becoming_aware" lemma="
        se␣rendre␣compte"/>
    <LU name="voir.v" annotated_occ="2" cat="v" frame="FR_Becoming_aware" lemma="voir"/>
</FRAME>
<FRAME name="FR_Chatting-Discussion" annotated_occ="274" domains="COM_LANG" frameID="10010">
    <FE name="Interlocutor_1"/>
    <FE name="Interlocutor_2"/>
    <FE name="Interlocutors"/>
    <FE name="Message"/>
    <FE name="Topic"/>
    <LU name="causerie.n" annotated_occ="1" cat="n" frame="FR_Chatting-Discussion" lemma="
        causerie"/>
    <LU name="communication.n" annotated_occ="10" cat="n" frame="FR_Chatting-Discussion"
        lemma="communication"/>
    <LU name="communiquer.v" annotated_occ="2" cat="v" frame="FR_Chatting-Discussion" lemma="
        communiquer"/>
    <LU name="contact.n" annotated_occ="3" cat="n" frame="FR_Chatting-Discussion" lemma="
        contact"/>
    <LU name="conversation.n" annotated_occ="12" cat="n" frame="FR_Chatting-Discussion" lemma
        ="conversation"/>
```

```
    <LU name="dialogue␣de␣sourd.n" annotated_occ="2" cat="n" frame="FR_Chatting-Discussion"
        lemma="dialogue␣de␣sourd"/>
    <LU name="dialogue.n" annotated_occ="18" cat="n" frame="FR_Chatting-Discussion" lemma="
        dialogue"/>
    <LU name="dialoguer.v" annotated_occ="4" cat="v" frame="FR_Chatting-Discussion" lemma="
        dialoguer"/>
    <LU name="discussion.n" annotated_occ="61" cat="n" frame="FR_Chatting-Discussion" lemma="
        discussion"/>
    <LU name="discuter.v" annotated_occ="38" cat="v" frame="FR_Chatting-Discussion" lemma="
        discuter"/>
    <LU name="débat.n" annotated_occ="94" cat="n" frame="FR_Chatting-Discussion" lemma="débat
        "/>
    <LU name="débattre.v" annotated_occ="7" cat="v" frame="FR_Chatting-Discussion" lemma="dé
        battre"/>
    <LU name="délibération.n" annotated_occ="1" cat="n" frame="FR_Chatting-Discussion" lemma=
        "délibération"/>
    <LU name="délibérer.v" annotated_occ="2" cat="v" frame="FR_Chatting-Discussion" lemma="dé
        libérer"/>
    <LU name="faire␣causette.v" annotated_occ="1" cat="v" frame="FR_Chatting-Discussion"
        lemma="faire␣causette"/>
    <LU name="palabre.n" annotated_occ="1" cat="n" frame="FR_Chatting-Discussion" lemma="
        palabre"/>
    <LU name="parler.v" annotated_occ="2" cat="v" frame="FR_Chatting-Discussion" lemma="
        parler"/>
    <LU name="pourparler.n" annotated_occ="10" cat="n" frame="FR_Chatting-Discussion" lemma="
        pourparler"/>
    <LU name="tractation.n" annotated_occ="5" cat="n" frame="FR_Chatting-Discussion" lemma="
        tractation"/>
</FRAME>
<FRAME name="FR_Reason" annotated_occ="319" domains="CAUSE" frameID="10083">
    <FE name="Action"/>
    <FE name="Actor"/>
    <FE name="Agent"/>
    <FE name="State_of_Affairs"/>
    <LU name="acculer.v" annotated_occ="4" cat="v" frame="FR_Reason" lemma="acculer"/>
    <LU name="amener.v" annotated_occ="23" cat="v" frame="FR_Reason" lemma="amener"/>
    <LU name="astreindre.v" annotated_occ="1" cat="v" frame="FR_Reason" lemma="astreindre"/>
    <LU name="au␣motif␣que.conj" annotated_occ="2" cat="conj" frame="FR_Reason" lemma="au␣
        motif␣que"/>
    <LU name="au␣vu␣de.prep" annotated_occ="4" cat="prep" frame="FR_Reason" lemma="au␣vu␣de"/
        >
    <LU name="conduire.v" annotated_occ="33" cat="v" frame="FR_Reason" lemma="conduire"/>
    <LU name="contraindre.v" annotated_occ="46" cat="v" frame="FR_Reason" lemma="contraindre"
        />
    <LU name="contrainte.n" annotated_occ="2" cat="n" frame="FR_Reason" lemma="contrainte"/>
    <LU name="en␣vertu␣de.prep" annotated_occ="5" cat="prep" frame="FR_Reason" lemma="en␣
        vertu␣de"/>
    <LU name="fonder.v" annotated_occ="3" cat="v" frame="FR_Reason" lemma="fonder"/>
    <LU name="forcer.v" annotated_occ="16" cat="v" frame="FR_Reason" lemma="forcer"/>
    <LU name="inspirer.v" annotated_occ="5" cat="v" frame="FR_Reason" lemma="inspirer"/>
    <LU name="mobile.n" annotated_occ="1" cat="n" frame="FR_Reason" lemma="mobile"/>
    <LU name="motif.n" annotated_occ="17" cat="n" frame="FR_Reason" lemma="motif"/>
    <LU name="motivation.n" annotated_occ="1" cat="n" frame="FR_Reason" lemma="motivation"/>
    <LU name="motiver.v" annotated_occ="4" cat="v" frame="FR_Reason" lemma="motiver"/>
    <LU name="obliger.v" annotated_occ="52" cat="v" frame="FR_Reason" lemma="obliger"/>
    <LU name="puisque.conj" annotated_occ="1" cat="conj" frame="FR_Reason" lemma="puisque"/>
    <LU name="raison.n" annotated_occ="52" cat="n" frame="FR_Reason" lemma="raison"/>
    <LU name="soulever.v" annotated_occ="8" cat="v" frame="FR_Reason" lemma="soulever"/>
    <LU name="susciter.v" annotated_occ="37" cat="v" frame="FR_Reason" lemma="susciter"/>
    <LU name="voir.v" annotated_occ="1" cat="v" frame="FR_Reason" lemma="voir"/>
    <LU name="vu␣que.conj" annotated_occ="1" cat="conj" frame="FR_Reason" lemma="vu␣que"/>
</FRAME>
```

</FrameNetFEN>

ANNEXE D

HIÉRARCHIE DES MACRO-RÔLES

```
<?xml version="1.0" encoding="utf-8" ?>
<macroroles>
  <info>Hierarchy of macro-roles, grouping several macro-roles and/or several frame-specific roles.</info>
  <macrorole name="All_addressees">
    <submacrorole name="Co_addressee"/>
    <submacrorole name="Addressee"/>
  </macrorole>
  <macrorole name="All_cognizers">
    <submacrorole name="Cognizer"/>
    <submacrorole name="Co_cognizer"/>
  </macrorole>
  <macrorole name="All_speakers">
    <submacrorole name="Speaker"/>
    <submacrorole name="Co_speaker"/>
  </macrorole>
  <macrorole name="Content_or_entity">
    <submacrorole name="Action_entity"/>
    <submacrorole name="True_content"/>
    <submacrorole name="Legit_content"/>
    <submacrorole name="Real_entity"/>
    <submacrorole name="Action"/>
    <submacrorole name="Legit_entity"/>
    <subrole_main frame="FR_Attempt_supported_suasion" role="State_of_affairs" macro="Content_or_entity"/>
  </macrorole>
  <macrorole name="Action">
    <subrole_main frame="Appointing" role="Function" macro="Action"/>
    <subrole_main frame="Be_in_agreement_on_action" role="Obligation" macro="Action"/>
    <subrole_main frame="FR_Agree_or_refuse_to_act" role="Proposed_action" macro="Action"/>
    <subrole_main frame="FR_Agree_or_refuse_to_act_in_favor_of" role="Proposed_action" macro="Action"/>
    <subrole_main frame="FR_Attempt_suasion" role="Action" macro="Action"/>
    <subrole_main frame="FR_Attempt_suasion.to_do" role="Action" macro="Action"/>
    <subrole_main frame="FR_Be_in_agreement" role="Obligation" macro="Action"/>
    <subrole_main frame="FR_Cognizer_affecting" role="Action" macro="Action"/>
    <subrole_main frame="FR_Cognizer_affecting.to_do" role="Action" macro="Action"/>
    <subrole_main frame="FR_Commitment.to_do" role="Eventuality" macro="Action"/>
    <subrole_main frame="FR_Deciding" role="Decision" macro="Action"/>
    <subrole_main frame="FR_Expressing_stance" role="Action" macro="Action"/>
    <subrole_main frame="FR_Having_commercial_agreement" role="Obligation" macro="Action"/>
    <subrole_main frame="FR_Influence_of_event_on_cognizer-Subjective_influence" role="Action" macro="
        Action"/>
    <subrole_main frame="FR_Influence_of_event_on_cognizer-Subjective_influence" role="Activity" macro="
        Action"/>
    <subrole_main frame="FR_Make_agreement" role="Obligation" macro="Action"/>
    <subrole_main frame="FR_Means_for_purpose" role="Goal" macro="Action"/>
    <subrole_main frame="FR_Personal_stance" role="Action" macro="Action"/>
    <subrole_main frame="FR_Purpose" role="Goal" macro="Action"/>
    <subrole_main frame="FR_Remembering" role="Action" macro="Action"/>
    <subrole_main frame="FR_Renunciation" role="Eventuality" macro="Action"/>
    <subrole_main frame="FR_Request" role="Eventuality" macro="Action"/>
    <subrole_main frame="Make_agreement_on_action" role="Obligation" macro="Action"/>
    <subrole_main frame="Remembering_to_do" role="Action" macro="Action"/>
    <subrole_main frame="Respond_to_proposal" role="Proposal" macro="Action"/>
    <subrole_main frame="Willingness" role="Activity" macro="Action"/>
```

```
</macrorole>
<macrorole name="Action_entity">
  <subrole_main frame="Appointing" role="Body" macro="Action_entity"/>
  <subrole_main frame="Appointing" role="Role" macro="Action_entity"/>
  <subrole_main frame="FR_Commitment.to_do" role="Entity" macro="Action_entity"/>
  <subrole_main frame="FR_Renunciation" role="Benefit" macro="Action_entity"/>
  <subrole_main frame="FR_Request" role="Requested_entity" macro="Action_entity"/>
  <subrole_main frame="Remembering_to_do" role="Salient_entity" macro="Action_entity"/>
</macrorole>
<macrorole name="Addressee">
  <subrole_main frame="Adducing" role="Addressee" macro="Addressee"/>
  <subrole_main frame="Bragging" role="Addressee" macro="Addressee"/>
  <subrole_main frame="Communication_response" role="Addressee" macro="Addressee"/>
  <subrole_main frame="Complaining" role="Addressee" macro="Addressee"/>
  <subrole_main frame="Convey_importance" role="Addressee" macro="Addressee"/>
  <subrole_main frame="FR_Attempt_suasion" role="Cognizer" macro="Addressee"/>
  <subrole_main frame="FR_Attempt_suasion.legitimacy" role="Addressee" macro="Addressee"/>
  <subrole_main frame="FR_Attempt_suasion.to_do" role="Cognizer" macro="Addressee"/>
  <subrole_main frame="FR_Attempt_suasion.veracity" role="Addressee" macro="Addressee"/>
  <subrole_main frame="FR_Attempt_supported_suasion" role="Addressee" macro="Addressee"/>
  <subrole_main frame="FR_Chatting-Discussion" role="Interlocutor_2" macro="Addressee"/>
  <subrole_main frame="FR_Commitment" role="Addressee" macro="Addressee"/>
  <subrole_main frame="FR_Commitment.to_do" role="Addressee" macro="Addressee"/>
  <subrole_main frame="FR_Commitment.veracity" role="Addressee" macro="Addressee"/>
  <subrole_main frame="FR_Contacting" role="Addressee" macro="Addressee"/>
  <subrole_main frame="FR_Encoding" role="Addressee" macro="Addressee"/>
  <subrole_main frame="FR_Expressing_truth_of_proposition" role="Addressee" macro="Addressee"/>
  <subrole_main frame="FR_Grant_permission-Permitting" role="Grantee" macro="Addressee"/>
  <subrole_main frame="FR_Hail" role="Addressee" macro="Addressee"/>
  <subrole_main frame="FR_Insult" role="Addressee" macro="Addressee"/>
  <subrole_main frame="FR_Judgment_communication" role="Addressee" macro="Addressee"/>
  <subrole_main frame="FR_Justifying" role="Addressee" macro="Addressee"/>
  <subrole_main frame="FR_Proving" role="Addressee" macro="Addressee"/>
  <subrole_main frame="FR_Quoting" role="Addressee" macro="Addressee"/>
  <subrole_main frame="FR_Request" role="Addressee" macro="Addressee"/>
  <subrole_main frame="FR_Speak_on_topic" role="Addressee" macro="Addressee"/>
  <subrole_main frame="FR_Statement-manner-noise" role="Addressee" macro="Addressee"/>
  <subrole_main frame="FR_Telling" role="Addressee" macro="Addressee"/>
  <subrole_main frame="Judgment_direct_address" role="Addressee" macro="Addressee"/>
  <subrole_main frame="Prevarication" role="Addressee" macro="Addressee"/>
  <subrole_main frame="Quarreling" role="Arguer2" macro="Addressee"/>
  <subrole_main frame="Questioning" role="Addressee" macro="Addressee"/>
  <subrole_main frame="Reading_aloud" role="Addressee" macro="Addressee"/>
  <subrole_main frame="Respond_to_proposal" role="Interlocutor" macro="Addressee"/>
  <subrole_main frame="Reveal_secret" role="Addressee" macro="Addressee"/>
  <subrole_main frame="Successfully_communicate_message" role="Addressee" macro="Addressee"/>
  <subrole_main frame="Summarizing" role="Addressee" macro="Addressee"/>
  <subrole_main frame="Text_creation" role="Addressee" macro="Addressee"/>
  <subrole_secondary frame="FR_Chatting-Discussion" role="Interlocutors" macro="Addressee"/>
  <subrole_secondary frame="Quarreling" role="Arguers" macro="Addressee"/>
</macrorole>
<macrorole name="Beneficiary">
  <subrole_main frame="FR_Agree_or_refuse_to_act_in_favor_of" role="Beneficiary" macro="Beneficiary"/>
</macrorole>
<macrorole name="Cause">
  <subrole_main frame="Causation" role="Cause" macro="Cause"/>
  <subrole_main frame="Explaining_the_facts" role="Fact" macro="Cause"/>
  <subrole_main frame="FR_Attributing_cause" role="Cause" macro="Cause"/>
  <subrole_main frame="FR_Cause_enunciation" role="Cause" macro="Cause"/>
  <subrole_main frame="FR_Cause_to_start-Launch_process" role="Cause" macro="Cause"/>
  <subrole_main frame="FR_Reason" role="State_of_affairs" macro="Cause"/>
  <subrole_main frame="Preventing" role="Preventing_cause" macro="Cause"/>
  <subrole_main frame="Response" role="Trigger" macro="Cause"/>
</macrorole>
<macrorole name="Cause_actor">
  <subrole_main frame="Causation" role="Actor" macro="Cause_actor"/>
  <subrole_main frame="FR_Attributing_cause" role="Responsible_party" macro="Cause_actor"/>
  <subrole_main frame="FR_Cause_to_start-Launch_process" role="Actor" macro="Cause_actor"/>
  <subrole_main frame="FR_Reason" role="Actor" macro="Cause_actor"/>
  <subrole_main frame="Preventing" role="Agent" macro="Cause_actor"/>
  <subrole_main frame="Response" role="Sentient_trigger" macro="Cause_actor"/>
</macrorole>
```

```
<macrorole name="Co_addressee">
  <subrole_secondary frame="FR_Chatting-Discussion" role="Interlocutor_1" macro="Co_addressee"/>
  <subrole_secondary frame="Quarreling" role="Arguer1" macro="Co_addressee"/>
</macrorole>
<macrorole name="Co_cognizer">
  <subrole_main frame="Be_in_agreement_on_action" role="Party_2" macro="Co_cognizer"/>
  <subrole_main frame="Be_in_agreement_on_assessment" role="Cognizer_2" macro="Co_cognizer"/>
  <subrole_main frame="FR_Be_in_agreement" role="Party_2" macro="Co_cognizer"/>
  <subrole_main frame="FR_Having_commercial_agreement" role="Party_2" macro="Co_cognizer"/>
  <subrole_main frame="FR_Make_agreement" role="Party_2" macro="Co_cognizer"/>
  <subrole_main frame="FR_Make_agreement_on_assessment" role="Cognizer_2" macro="Co_cognizer"/>
  <subrole_main frame="Make_agreement_on_action" role="Party_2" macro="Co_cognizer"/>
  <subrole_secondary frame="FR_Having_commercial_agreement" role="Seller" macro="Co_cognizer"/>
</macrorole>
<macrorole name="Co_speaker">
  <subrole_secondary frame="FR_Chatting-Discussion" role="Interlocutor_2" macro="Co_speaker"/>
  <subrole_secondary frame="Quarreling" role="Arguer2" macro="Co_speaker"/>
</macrorole>
<macrorole name="Cognizer">
  <subrole_main frame="Be_in_agreement_on_action" role="Parties" macro="Cognizer"/>
  <subrole_main frame="Be_in_agreement_on_action" role="Party_1" macro="Cognizer"/>
  <subrole_main frame="Be_in_agreement_on_assessment" role="Cognizer_1" macro="Cognizer"/>
  <subrole_main frame="Be_in_agreement_on_assessment" role="Cognizers" macro="Cognizer"/>
  <subrole_main frame="Categorization" role="Cognizer" macro="Cognizer"/>
  <subrole_main frame="Coming_to_believe" role="Cognizer" macro="Cognizer"/>
  <subrole_main frame="Evoking" role="Cognizer" macro="Cognizer"/>
  <subrole_main frame="FR_Agree_or_refuse_to_act" role="Cognizer" macro="Cognizer"/>
  <subrole_main frame="FR_Agree_or_refuse_to_act_in_favor_of" role="Cognizer" macro="Cognizer"/>
  <subrole_main frame="FR_Awareness-Certainty-Opinion" role="Cognizer" macro="Cognizer"/>
  <subrole_main frame="FR_Be_in_agreement" role="Parties" macro="Cognizer"/>
  <subrole_main frame="FR_Be_in_agreement" role="Party_1" macro="Cognizer"/>
  <subrole_main frame="FR_Becoming_aware" role="Cognizer" macro="Cognizer"/>
  <subrole_main frame="FR_Being_in_favor_of" role="Cognizer" macro="Cognizer"/>
  <subrole_main frame="FR_Cognizer_affecting" role="Cognizer" macro="Cognizer"/>
  <subrole_main frame="FR_Cognizer_affecting.legitimacy" role="Cognizer" macro="Cognizer"/>
  <subrole_main frame="FR_Cognizer_affecting.to_do" role="Cognizer" macro="Cognizer"/>
  <subrole_main frame="FR_Cognizer_affecting.veracity" role="Cognizer" macro="Cognizer"/>
  <subrole_main frame="FR_Deciding" role="Cognizer" macro="Cognizer"/>
  <subrole_main frame="FR_Expectation" role="Cognizer" macro="Cognizer"/>
  <subrole_main frame="FR_Having_commercial_agreement" role="Buyer" macro="Cognizer"/>
  <subrole_main frame="FR_Having_commercial_agreement" role="Parties" macro="Cognizer"/>
  <subrole_main frame="FR_Having_commercial_agreement" role="Party_1" macro="Cognizer"/>
  <subrole_main frame="FR_Influence_of_event_on_cognizer-Subjective_influence" role="Cognizer" macro="
      Cognizer"/>
  <subrole_main frame="FR_Make_agreement" role="Parties" macro="Cognizer"/>
  <subrole_main frame="FR_Make_agreement" role="Party_1" macro="Cognizer"/>
  <subrole_main frame="FR_Make_agreement_on_assessment" role="Cognizer_1" macro="Cognizer"/>
  <subrole_main frame="FR_Make_agreement_on_assessment" role="Cognizers" macro="Cognizer"/>
  <subrole_main frame="FR_Memory-Remembering_experience-Remembering_information" role="Cognizer" macro="
      Cognizer"/>
  <subrole_main frame="FR_Personal_stance" role="Cognizer" macro="Cognizer"/>
  <subrole_main frame="FR_Purpose" role="Agent" macro="Cognizer"/>
  <subrole_main frame="FR_Questioning_oneself" role="Cognizer" macro="Cognizer"/>
  <subrole_main frame="FR_Reliance_on_expectation" role="Cognizer" macro="Cognizer"/>
  <subrole_main frame="FR_Remembering" role="Cognizer" macro="Cognizer"/>
  <subrole_main frame="FR_Renunciation" role="Cognizer" macro="Cognizer"/>
  <subrole_main frame="FR_Taking_sides" role="Cognizer" macro="Cognizer"/>
  <subrole_main frame="Judgment" role="Cognizer" macro="Cognizer"/>
  <subrole_main frame="Make_agreement_on_action" role="Parties" macro="Cognizer"/>
  <subrole_main frame="Make_agreement_on_action" role="Party_1" macro="Cognizer"/>
  <subrole_main frame="Regard" role="Cognizer" macro="Cognizer"/>
  <subrole_main frame="Religious_belief" role="Believer" macro="Cognizer"/>
  <subrole_main frame="Remembering_to_do" role="Cognizer" macro="Cognizer"/>
  <subrole_main frame="Willingness" role="Cognizer" macro="Cognizer"/>
  <subrole_secondary frame="FR_Attempt_suasion" role="Cognizer" macro="Cognizer"/>
  <subrole_secondary frame="FR_Attempt_suasion.legitimacy" role="Addressee" macro="Cognizer"/>
  <subrole_secondary frame="FR_Attempt_suasion.to_do" role="Cognizer" macro="Cognizer"/>
  <subrole_secondary frame="FR_Attempt_suasion.veracity" role="Addressee" macro="Cognizer"/>
  <subrole_secondary frame="FR_Attempt_supported_suasion" role="Addressee" macro="Cognizer"/>
  <subrole_secondary frame="FR_Attributed_information" role="Source" macro="Cognizer"/>
  <subrole_secondary frame="FR_Attributing_cause" role="Attributor" macro="Cognizer"/>
  <subrole_secondary frame="FR_Expressing_side" role="Speaker" macro="Cognizer"/>
```

```
    <subrole_secondary frame="FR_Expressing_stance" role="Speaker" macro="Cognizer"/>
    <subrole_secondary frame="FR_Judgment_communication" role="Communicator" macro="Cognizer"/>
    <subrole_secondary frame="FR_Judgment_communication_no_reason" role="Communicator" macro="Cognizer"/>
    <subrole_secondary frame="FR_Justifying" role="Speaker" macro="Cognizer"/>
    <subrole_secondary frame="FR_Proving" role="Addressee" macro="Cognizer"/>
    <subrole_secondary frame="FR_Request" role="Addressee" macro="Cognizer"/>
    <subrole_secondary frame="Judgment_direct_address" role="Communicator" macro="Cognizer"/>
    <subrole_secondary frame="Predicting" role="Speaker" macro="Cognizer"/>
    <subrole_secondary frame="Respond_to_proposal" role="Speaker" macro="Cognizer"/>
  </macrorole>
  <macrorole name="Effect">
    <subrole_main frame="Causation" role="Effect" macro="Effect"/>
    <subrole_main frame="Explaining_the_facts" role="State_of_affairs" macro="Effect"/>
    <subrole_main frame="FR_Attributing_cause" role="Known_state_of_affairs" macro="Effect"/>
    <subrole_main frame="FR_Cause_enunciation" role="Enunciative_effect" macro="Effect"/>
    <subrole_main frame="FR_Cause_to_start-Launch_process" role="Effect" macro="Effect"/>
    <subrole_main frame="FR_Reason" role="Action" macro="Effect"/>
    <subrole_main frame="Preventing" role="Event" macro="Effect"/>
    <subrole_main frame="Response" role="Response" macro="Effect"/>
  </macrorole>
  <macrorole name="Entity_predicated_on">
    <subrole_main frame="Categorization" role="Item" macro="Entity_predicated_on"/>
  </macrorole>
  <macrorole name="Evaluee">
    <subrole_main frame="FR_Judgment_communication" role="Evaluee" macro="Evaluee"/>
    <subrole_main frame="FR_Judgment_communication_no_reason" role="Evaluee" macro="Evaluee"/>
    <subrole_main frame="Judgment" role="Evaluee" macro="Evaluee"/>
    <subrole_main frame="Regard" role="Evaluee" macro="Evaluee"/>
    <subrole_secondary frame="Judgment_direct_address" role="Addressee" macro="Evaluee"/>
  </macrorole>
  <macrorole name="Evidence">
    <subrole_main frame="Coming_to_believe" role="Evidence" macro="Evidence"/>
    <subrole_main frame="FR_Attempt_supported_suasion" role="Support" macro="Evidence"/>
    <subrole_main frame="FR_Awareness-Certainty-Opinion" role="Evidence" macro="Evidence"/>
    <subrole_main frame="FR_Becoming_aware" role="Evidence" macro="Evidence"/>
    <subrole_main frame="FR_Cognizer_affecting" role="Evidence" macro="Evidence"/>
    <subrole_main frame="FR_Cognizer_affecting.legitimacy" role="Evidence" macro="Evidence"/>
    <subrole_main frame="FR_Cognizer_affecting.to_do" role="Evidence" macro="Evidence"/>
    <subrole_main frame="FR_Cognizer_affecting.veracity" role="Evidence" macro="Evidence"/>
    <subrole_main frame="FR_Expectation" role="Evidence" macro="Evidence"/>
    <subrole_main frame="FR_Proving" role="Support" macro="Evidence"/>
    <subrole_main frame="Predicting" role="Evidence" macro="Evidence"/>
  </macrorole>
  <macrorole name="Goods">
    <subrole_main frame="Carry_goods" role="Goods" macro="Goods"/>
    <subrole_main frame="Commerce_buy" role="Goods" macro="Goods"/>
    <subrole_main frame="Commerce_collect" role="Goods" macro="Goods"/>
    <subrole_main frame="Commerce_goods-transfer" role="Goods" macro="Goods"/>
    <subrole_main frame="Commerce_money-transfer" role="Goods" macro="Goods"/>
    <subrole_main frame="Commerce_pay" role="Goods" macro="Goods"/>
    <subrole_main frame="Commerce_scenario" role="Goods" macro="Goods"/>
    <subrole_main frame="Commerce_sell" role="Goods" macro="Goods"/>
    <subrole_main frame="Commercial_transaction" role="Goods" macro="Goods"/>
    <subrole_main frame="Earnings_and_losses" role="Goods" macro="Goods"/>
    <subrole_main frame="Exporting" role="Goods" macro="Goods"/>
    <subrole_main frame="FR_Getting_money" role="Goods" macro="Goods"/>
    <subrole_main frame="FR_Giving_money" role="Goods" macro="Goods"/>
    <subrole_main frame="FR_Spending" role="Goods" macro="Goods"/>
    <subrole_main frame="Importing" role="Goods" macro="Goods"/>
    <subrole_main frame="Renting" role="Goods" macro="Goods"/>
    <subrole_main frame="Renting_out" role="Goods" macro="Goods"/>
    <subrole_main frame="Repayment" role="Loan" macro="Goods"/>
  </macrorole>
  <macrorole name="Intrument">
    <subrole_main frame="FR_Becoming_aware" role="Instrument" macro="Intrument"/>
  </macrorole>
  <macrorole name="Legit_content">
    <subrole_main frame="FR_Attempt_suasion" role="Legitimate_content" macro="Legit_content"/>
    <subrole_main frame="FR_Attempt_suasion.legitimacy" role="State_of_affairs" macro="Legit_content"/>
    <subrole_main frame="FR_Being_in_favor_of" role="Content" macro="Legit_content"/>
    <subrole_main frame="FR_Cognizer_affecting" role="Legitimate_content" macro="Legit_content"/>
    <subrole_main frame="FR_Cognizer_affecting.legitimacy" role="Content" macro="Legit_content"/>
```

```
  <subrole_main frame="FR_Deserving" role="Eventuality" macro="Legit_content"/>
  <subrole_main frame="FR_Expressing_side" role="Content" macro="Legit_content"/>
  <subrole_main frame="FR_Expressing_stance" role="Legitimate_content" macro="Legit_content"/>
  <subrole_main frame="FR_Justifying" role="Eventuality" macro="Legit_content"/>
  <subrole_main frame="FR_Personal_stance" role="Legitimate_content" macro="Legit_content"/>
  <subrole_main frame="FR_Taking_sides" role="Content" macro="Legit_content"/>
</macrorole>
<macrorole name="Legit_entity">
  <subrole_main frame="FR_Being_in_favor_of" role="Side" macro="Legit_entity"/>
  <subrole_main frame="FR_Expressing_side" role="Side" macro="Legit_entity"/>
  <subrole_main frame="FR_Justifying" role="Responsible_entity" macro="Legit_entity"/>
  <subrole_main frame="FR_Taking_sides" role="Side" macro="Legit_entity"/>
</macrorole>
<macrorole name="Means">
  <subrole_main frame="Coming_to_believe" role="Means" macro="Means"/>
  <subrole_main frame="FR_Means_for_purpose" role="Means" macro="Means"/>
  <subrole_main frame="FR_Purpose" role="Means" macro="Means"/>
</macrorole>
<macrorole name="Medium">
  <subrole_main frame="Adducing" role="Medium" macro="Medium"/>
  <subrole_main frame="Be_in_agreement_on_action" role="Medium" macro="Medium"/>
  <subrole_main frame="Coming_to_believe" role="Medium" macro="Medium"/>
  <subrole_main frame="Communication_response" role="Medium" macro="Medium"/>
  <subrole_main frame="Complaining" role="Medium" macro="Medium"/>
  <subrole_main frame="Convey_importance" role="Medium" macro="Medium"/>
  <subrole_main frame="FR_Attempt_suasion" role="Medium" macro="Medium"/>
  <subrole_main frame="FR_Attempt_suasion.legitimacy" role="Medium" macro="Medium"/>
  <subrole_main frame="FR_Attempt_suasion.to_do" role="Medium" macro="Medium"/>
  <subrole_main frame="FR_Attempt_suasion.veracity" role="Medium" macro="Medium"/>
  <subrole_main frame="FR_Attempt_supported_suasion" role="Medium" macro="Medium"/>
  <subrole_main frame="FR_Attributed_information" role="Informational_content" macro="Medium"/>
  <subrole_main frame="FR_Cognizer_affecting" role="Medium" macro="Medium"/>
  <subrole_main frame="FR_Cognizer_affecting.legitimacy" role="Medium" macro="Medium"/>
  <subrole_main frame="FR_Cognizer_affecting.to_do" role="Medium" macro="Medium"/>
  <subrole_main frame="FR_Cognizer_affecting.veracity" role="Medium" macro="Medium"/>
  <subrole_main frame="FR_Commitment" role="Medium" macro="Medium"/>
  <subrole_main frame="FR_Commitment.to_do" role="Medium" macro="Medium"/>
  <subrole_main frame="FR_Commitment.veracity" role="Medium" macro="Medium"/>
  <subrole_main frame="FR_Deciding" role="Medium" macro="Medium"/>
  <subrole_main frame="FR_Encoding" role="Medium" macro="Medium"/>
  <subrole_main frame="FR_Expressing_decision" role="Medium" macro="Medium"/>
  <subrole_main frame="FR_Expressing_truth_of_proposition" role="Medium" macro="Medium"/>
  <subrole_main frame="FR_Grant_permission-Permitting" role="Principle" macro="Medium"/>
  <subrole_main frame="FR_Hail" role="Medium" macro="Medium"/>
  <subrole_main frame="FR_Having_commercial_agreement" role="Medium" macro="Medium"/>
  <subrole_main frame="FR_Judgment_communication" role="Medium" macro="Medium"/>
  <subrole_main frame="FR_Judgment_communication_no_reason" role="Medium" macro="Medium"/>
  <subrole_main frame="FR_Justifying" role="Medium" macro="Medium"/>
  <subrole_main frame="FR_Proving" role="Medium" macro="Medium"/>
  <subrole_main frame="FR_Purpose" role="Medium" macro="Medium"/>
  <subrole_main frame="FR_Quoting" role="Medium" macro="Medium"/>
  <subrole_main frame="FR_Request" role="Medium" macro="Medium"/>
  <subrole_main frame="FR_Speak_on_topic" role="Medium" macro="Medium"/>
  <subrole_main frame="FR_Statement-manner-noise" role="Medium" macro="Medium"/>
  <subrole_main frame="FR_Taking_sides" role="Medium" macro="Medium"/>
  <subrole_main frame="FR_Telling" role="Medium" macro="Medium"/>
  <subrole_main frame="Judgment_direct_address" role="Medium" macro="Medium"/>
  <subrole_main frame="Respond_to_proposal" role="Medium" macro="Medium"/>
  <subrole_main frame="Reveal_secret" role="Medium" macro="Medium"/>
  <subrole_main frame="Successfully_communicate_message" role="Medium" macro="Medium"/>
  <subrole_main frame="Summarizing" role="Medium" macro="Medium"/>
</macrorole>
<macrorole name="Message">
  <subrole_main frame="Bragging" role="Message" macro="Message"/>
  <subrole_main frame="Communication_response" role="Message" macro="Message"/>
  <subrole_main frame="Complaining" role="Complaint" macro="Message"/>
  <subrole_main frame="Convey_importance" role="Message" macro="Message"/>
  <subrole_main frame="FR_Attributed_information" role="Proposition" macro="Message"/>
  <subrole_main frame="FR_Commitment" role="Message" macro="Message"/>
  <subrole_main frame="FR_Commitment.veracity" role="Message" macro="Message"/>
  <subrole_main frame="FR_Encoding" role="Formulation" macro="Message"/>
  <subrole_main frame="FR_Expressing_truth_of_proposition" role="Message" macro="Message"/>
```

```
    <subrole_main frame="FR_Judgment_communication" role="Message" macro="Message"/>
    <subrole_main frame="FR_Judgment_communication_no_reason" role="Message" macro="Message"/>
    <subrole_main frame="FR_Making_speech" role="Speech" macro="Message"/>
    <subrole_main frame="FR_Proving" role="Content" macro="Message"/>
    <subrole_main frame="FR_Quoting" role="Message" macro="Message"/>
    <subrole_main frame="FR_Speak_on_topic" role="Message" macro="Message"/>
    <subrole_main frame="FR_Statement-manner-noise" role="Message" macro="Message"/>
    <subrole_main frame="FR_Telling" role="Message" macro="Message"/>
    <subrole_main frame="Predicting" role="Eventuality" macro="Message"/>
    <subrole_main frame="Questioning" role="Message" macro="Message"/>
    <subrole_main frame="Reading_aloud" role="Speech" macro="Message"/>
    <subrole_main frame="Reveal_secret" role="Information" macro="Message"/>
    <subrole_main frame="Spelling_and_pronouncing" role="Formal_realization" macro="Message"/>
    <subrole_main frame="Successfully_communicate_message" role="Message" macro="Message"/>
    <subrole_main frame="Summarizing" role="Summary" macro="Message"/>
    <subrole_main frame="Text_creation" role="Text" macro="Message"/>
  </macrorole>
  <macrorole name="Message_content">
    <subrole_main frame="FR_Contacting" role="Purpose" macro="Message_content"/>
    <subrole_main frame="FR_Encoding" role="Content" macro="Message_content"/>
    <subrole_main frame="Reading_aloud" role="Text" macro="Message_content"/>
    <subrole_main frame="Spelling_and_pronouncing" role="Sign" macro="Message_content"/>
    <subrole_main frame="Summarizing" role="State_of_affairs" macro="Message_content"/>
  </macrorole>
  <macrorole name="Money">
    <subrole_main frame="Commerce_buy" role="Money" macro="Money"/>
    <subrole_main frame="Commerce_collect" role="Money" macro="Money"/>
    <subrole_main frame="Commerce_goods-transfer" role="Money" macro="Money"/>
    <subrole_main frame="Commerce_money-transfer" role="Money" macro="Money"/>
    <subrole_main frame="Commerce_pay" role="Money" macro="Money"/>
    <subrole_main frame="Commerce_scenario" role="Money" macro="Money"/>
    <subrole_main frame="Commerce_sell" role="Money" macro="Money"/>
    <subrole_main frame="Commercial_transaction" role="Money" macro="Money"/>
    <subrole_main frame="Earnings_and_losses" role="Earnings" macro="Money"/>
    <subrole_main frame="FR_Cause_earning" role="Money" macro="Money"/>
    <subrole_main frame="FR_Getting_money" role="Money" macro="Money"/>
    <subrole_main frame="FR_Giving_money" role="Money" macro="Money"/>
    <subrole_main frame="FR_Having_commercial_agreement" role="Money" macro="Money"/>
    <subrole_main frame="FR_Reimbursement" role="Money" macro="Money"/>
    <subrole_main frame="FR_Spending" role="Money" macro="Money"/>
    <subrole_main frame="Renting" role="Money" macro="Money"/>
    <subrole_main frame="Renting_out" role="Money" macro="Money"/>
    <subrole_main frame="Repayment" role="Money" macro="Money"/>
  </macrorole>
  <macrorole name="Money_giver">
    <subrole_main frame="Commerce_buy" role="Buyer" macro="Money_giver"/>
    <subrole_main frame="Commerce_collect" role="Buyer" macro="Money_giver"/>
    <subrole_main frame="Commerce_goods-transfer" role="Buyer" macro="Money_giver"/>
    <subrole_main frame="Commerce_money-transfer" role="Buyer" macro="Money_giver"/>
    <subrole_main frame="Commerce_pay" role="Buyer" macro="Money_giver"/>
    <subrole_main frame="Commerce_scenario" role="Buyer" macro="Money_giver"/>
    <subrole_main frame="Commerce_sell" role="Buyer" macro="Money_giver"/>
    <subrole_main frame="Commercial_transaction" role="Buyer" macro="Money_giver"/>
    <subrole_main frame="Earnings_and_losses" role="Buyer" macro="Money_giver"/>
    <subrole_main frame="Exporting" role="Importer" macro="Money_giver"/>
    <subrole_main frame="FR_Getting_money" role="Giver" macro="Money_giver"/>
    <subrole_main frame="FR_Giving_money" role="Giver" macro="Money_giver"/>
    <subrole_main frame="FR_Reimbursement" role="Giver" macro="Money_giver"/>
    <subrole_main frame="FR_Spending" role="Spender" macro="Money_giver"/>
    <subrole_main frame="Importing" role="Importer" macro="Money_giver"/>
    <subrole_main frame="Renting" role="Lessee" macro="Money_giver"/>
    <subrole_main frame="Renting_out" role="Lessee" macro="Money_giver"/>
    <subrole_main frame="Repayment" role="Debtor" macro="Money_giver"/>
    <subrole_secondary frame="FR_Having_commercial_agreement" role="Buyer" macro="Money_giver"/>
  </macrorole>
  <macrorole name="Money_recipient">
    <subrole_main frame="Commerce_buy" role="Seller" macro="Money_recipient"/>
    <subrole_main frame="Commerce_collect" role="Seller" macro="Money_recipient"/>
    <subrole_main frame="Commerce_goods-transfer" role="Seller" macro="Money_recipient"/>
    <subrole_main frame="Commerce_money-transfer" role="Seller" macro="Money_recipient"/>
    <subrole_main frame="Commerce_pay" role="Seller" macro="Money_recipient"/>
    <subrole_main frame="Commerce_scenario" role="Seller" macro="Money_recipient"/>
```

```
  <subrole_main frame="Commerce_sell" role="Seller" macro="Money_recipient"/>
  <subrole_main frame="Commercial_transaction" role="Seller" macro="Money_recipient"/>
  <subrole_main frame="Earnings_and_losses" role="Earner" macro="Money_recipient"/>
  <subrole_main frame="Exporting" role="Exporter" macro="Money_recipient"/>
  <subrole_main frame="FR_Cause_earning" role="Earner" macro="Money_recipient"/>
  <subrole_main frame="FR_Getting_money" role="Getter" macro="Money_recipient"/>
  <subrole_main frame="FR_Giving_money" role="Recipient" macro="Money_recipient"/>
  <subrole_main frame="FR_Having_commercial_agreement" role="Seller" macro="Money_recipient"/>
  <subrole_main frame="FR_Reimbursement" role="Recipient" macro="Money_recipient"/>
  <subrole_main frame="FR_Spending" role="Seller" macro="Money_recipient"/>
  <subrole_main frame="Importing" role="Exporter" macro="Money_recipient"/>
  <subrole_main frame="Renting" role="Lessor" macro="Money_recipient"/>
  <subrole_main frame="Renting_out" role="Lessor" macro="Money_recipient"/>
  <subrole_main frame="Repayment" role="Loaner" macro="Money_recipient"/>
</macrorole>
<macrorole name="Name">
  <subrole_main frame="FR_Being_named" role="Name" macro="Name"/>
  <subrole_main frame="Name_conferral" role="Name" macro="Name"/>
  <subrole_main frame="Referring_by_name" role="Name" macro="Name"/>
</macrorole>
<macrorole name="Named_entity">
  <subrole_main frame="FR_Being_named" role="Entity" macro="Named_entity"/>
  <subrole_main frame="Name_conferral" role="Entity" macro="Named_entity"/>
  <subrole_main frame="Referring_by_name" role="Entity" macro="Named_entity"/>
</macrorole>
<macrorole name="Polarity">
  <subrole_main frame="FR_Attempt_suasion.legitimacy" role="Polarity" macro="Polarity"/>
  <subrole_main frame="FR_Attempt_supported_suasion" role="Polarity" macro="Polarity"/>
  <subrole_main frame="FR_Expressing_side" role="Polarity" macro="Polarity"/>
</macrorole>
<macrorole name="Rate">
  <subrole_main frame="Commerce_buy" role="Rate" macro="Rate"/>
  <subrole_main frame="Commerce_collect" role="Rate" macro="Rate"/>
  <subrole_main frame="Commerce_goods-transfer" role="Rate" macro="Rate"/>
  <subrole_main frame="Commerce_money-transfer" role="Rate" macro="Rate"/>
  <subrole_main frame="Commerce_pay" role="Rate" macro="Rate"/>
  <subrole_main frame="Commerce_scenario" role="Rate" macro="Rate"/>
  <subrole_main frame="Commerce_sell" role="Rate" macro="Rate"/>
  <subrole_main frame="Commercial_transaction" role="Rate" macro="Rate"/>
  <subrole_main frame="Earnings_and_losses" role="Rate" macro="Rate"/>
  <subrole_main frame="FR_Getting_money" role="Rate" macro="Rate"/>
  <subrole_main frame="FR_Giving_money" role="Rate" macro="Rate"/>
  <subrole_main frame="FR_Spending" role="Rate" macro="Rate"/>
  <subrole_main frame="Renting" role="Rate" macro="Rate"/>
  <subrole_main frame="Renting_out" role="Rate" macro="Rate"/>
</macrorole>
<macrorole name="Real_entity">
  <subrole_main frame="FR_Memory-Remembering_experience-Remembering_information" role="Salient_entity"
      macro="Real_entity"/>
  <subrole_main frame="Religious_belief" role="Element" macro="Real_entity"/>
</macrorole>
<macrorole name="Reason_of_judgment">
  <subrole_main frame="FR_Judgment_communication" role="Reason" macro="Reason_of_judgment"/>
  <subrole_main frame="Judgment" role="Reason" macro="Reason_of_judgment"/>
  <subrole_main frame="Judgment_direct_address" role="Reason" macro="Reason_of_judgment"/>
  <subrole_main frame="Regard" role="Reason" macro="Reason_of_judgment"/>
</macrorole>
<macrorole name="Reason_of_legitimacy">
  <subrole_main frame="FR_Deserving" role="Sufficient_reason" macro="Reason_of_legitimacy"/>
  <subrole_main frame="FR_Justifying" role="Sufficient_reason" macro="Reason_of_legitimacy"/>
</macrorole>
<macrorole name="Speaker">
  <subrole_main frame="Adducing" role="Speaker" macro="Speaker"/>
  <subrole_main frame="Appointing" role="Selector" macro="Speaker"/>
  <subrole_main frame="Bragging" role="Speaker" macro="Speaker"/>
  <subrole_main frame="Communication_response" role="Speaker" macro="Speaker"/>
  <subrole_main frame="Complaining" role="Complainer" macro="Speaker"/>
  <subrole_main frame="Convey_importance" role="Speaker" macro="Speaker"/>
  <subrole_main frame="FR_Attempt_suasion" role="Persuader" macro="Speaker"/>
  <subrole_main frame="FR_Attempt_suasion.legitimacy" role="Persuader" macro="Speaker"/>
  <subrole_main frame="FR_Attempt_suasion.to_do" role="Persuader" macro="Speaker"/>
  <subrole_main frame="FR_Attempt_suasion.veracity" role="Persuader" macro="Speaker"/>
```

```
    <subrole_main frame="FR_Attempt_supported_suasion" role="Persuader" macro="Speaker"/>
    <subrole_main frame="FR_Attributed_information" role="Source" macro="Speaker"/>
    <subrole_main frame="FR_Attributing_cause" role="Attributor" macro="Speaker"/>
    <subrole_main frame="FR_Chatting-Discussion" role="Interlocutor_1" macro="Speaker"/>
    <subrole_main frame="FR_Chatting-Discussion" role="Interlocutors" macro="Speaker"/>
    <subrole_main frame="FR_Cognizer_affecting" role="Persuader" macro="Speaker"/>
    <subrole_main frame="FR_Cognizer_affecting.legitimacy" role="Persuader" macro="Speaker"/>
    <subrole_main frame="FR_Cognizer_affecting.to_do" role="Persuader" macro="Speaker"/>
    <subrole_main frame="FR_Cognizer_affecting.veracity" role="Persuader" macro="Speaker"/>
    <subrole_main frame="FR_Commitment" role="Speaker" macro="Speaker"/>
    <subrole_main frame="FR_Commitment.to_do" role="Speaker" macro="Speaker"/>
    <subrole_main frame="FR_Commitment.veracity" role="Speaker" macro="Speaker"/>
    <subrole_main frame="FR_Contacting" role="Communicator" macro="Speaker"/>
    <subrole_main frame="FR_Encoding" role="Speaker" macro="Speaker"/>
    <subrole_main frame="FR_Expressing_decision" role="Speaker" macro="Speaker"/>
    <subrole_main frame="FR_Expressing_side" role="Speaker" macro="Speaker"/>
    <subrole_main frame="FR_Expressing_stance" role="Speaker" macro="Speaker"/>
    <subrole_main frame="FR_Expressing_truth_of_proposition" role="Speaker" macro="Speaker"/>
    <subrole_main frame="FR_Grant_permission-Permitting" role="Grantor" macro="Speaker"/>
    <subrole_main frame="FR_Hail" role="Speaker" macro="Speaker"/>
    <subrole_main frame="FR_Insult" role="Speaker" macro="Speaker"/>
    <subrole_main frame="FR_Judgment_communication" role="Communicator" macro="Speaker"/>
    <subrole_main frame="FR_Judgment_communication_no_reason" role="Communicator" macro="Speaker"/>
    <subrole_main frame="FR_Justifying" role="Speaker" macro="Speaker"/>
    <subrole_main frame="FR_Making_speech" role="Speaker" macro="Speaker"/>
    <subrole_main frame="FR_Proving" role="Persuader" macro="Speaker"/>
    <subrole_main frame="FR_Quoting" role="Speaker" macro="Speaker"/>
    <subrole_main frame="FR_Request" role="Speaker" macro="Speaker"/>
    <subrole_main frame="FR_Speak_on_topic" role="Speaker" macro="Speaker"/>
    <subrole_main frame="FR_Statement-manner-noise" role="Speaker" macro="Speaker"/>
    <subrole_main frame="FR_Telling" role="Speaker" macro="Speaker"/>
    <subrole_main frame="Judgment_direct_address" role="Communicator" macro="Speaker"/>
    <subrole_main frame="Name_conferral" role="Speaker" macro="Speaker"/>
    <subrole_main frame="Predicting" role="Speaker" macro="Speaker"/>
    <subrole_main frame="Prevarication" role="Speaker" macro="Speaker"/>
    <subrole_main frame="Quarreling" role="Arguer1" macro="Speaker"/>
    <subrole_main frame="Quarreling" role="Arguers" macro="Speaker"/>
    <subrole_main frame="Questioning" role="Speaker" macro="Speaker"/>
    <subrole_main frame="Ratification" role="Ratifier" macro="Speaker"/>
    <subrole_main frame="Reading_aloud" role="Speaker" macro="Speaker"/>
    <subrole_main frame="Referring_by_name" role="Speaker" macro="Speaker"/>
    <subrole_main frame="Respond_to_proposal" role="Speaker" macro="Speaker"/>
    <subrole_main frame="Reveal_secret" role="Speaker" macro="Speaker"/>
    <subrole_main frame="Spelling_and_pronouncing" role="Speaker" macro="Speaker"/>
    <subrole_main frame="Successfully_communicate_message" role="Communicator" macro="Speaker"/>
    <subrole_main frame="Summarizing" role="Communicator" macro="Speaker"/>
    <subrole_main frame="Text_creation" role="Author" macro="Speaker"/>
  </macrorole>
  <macrorole name="Time_of_eventuality">
    <subrole_main frame="FR_Expectation" role="Time_of_eventuality" macro="Time_of_eventuality"/>
    <subrole_main frame="FR_Means_for_purpose" role="Time_of_eventuality" macro="Time_of_eventuality"/>
    <subrole_main frame="FR_Purpose" role="Time_of_eventuality" macro="Time_of_eventuality"/>
    <subrole_main frame="FR_Reliance_on_expectation" role="Time_of_eventuality" macro="Time_of_eventuality"
        />
    <subrole_main frame="Predicting" role="Time_of_eventuality" macro="Time_of_eventuality"/>
  </macrorole>
  <macrorole name="Topic">
    <subrole_main frame="Be_in_agreement_on_action" role="Topic" macro="Topic"/>
    <subrole_main frame="Be_in_agreement_on_assessment" role="Topic" macro="Topic"/>
    <subrole_main frame="Bragging" role="Topic" macro="Topic"/>
    <subrole_main frame="Coming_to_believe" role="Topic" macro="Topic"/>
    <subrole_main frame="Communication_response" role="Topic" macro="Topic"/>
    <subrole_main frame="Complaining" role="Topic" macro="Topic"/>
    <subrole_main frame="FR_Attempt_suasion" role="Topic" macro="Topic"/>
    <subrole_main frame="FR_Attempt_suasion.legitimacy" role="Topic" macro="Topic"/>
    <subrole_main frame="FR_Attempt_suasion.to_do" role="Topic" macro="Topic"/>
    <subrole_main frame="FR_Attempt_suasion.veracity" role="Topic" macro="Topic"/>
    <subrole_main frame="FR_Attempt_supported_suasion" role="Topic" macro="Topic"/>
    <subrole_main frame="FR_Awareness-Certainty-Opinion" role="Topic" macro="Topic"/>
    <subrole_main frame="FR_Be_in_agreement" role="Topic" macro="Topic"/>
    <subrole_main frame="FR_Becoming_aware" role="Topic" macro="Topic"/>
    <subrole_main frame="FR_Chatting-Discussion" role="Topic" macro="Topic"/>
```

```
    <subrole_main frame="FR_Commitment.to_do" role="Topic" macro="Topic"/>
    <subrole_main frame="FR_Contacting" role="Topic" macro="Topic"/>
    <subrole_main frame="FR_Deciding" role="Topic" macro="Topic"/>
    <subrole_main frame="FR_Expectation" role="Topic" macro="Topic"/>
    <subrole_main frame="FR_Expressing_side" role="Topic" macro="Topic"/>
    <subrole_main frame="FR_Hail" role="Topic" macro="Topic"/>
    <subrole_main frame="FR_Having_commercial_agreement" role="Topic" macro="Topic"/>
    <subrole_main frame="FR_Insult" role="Topic" macro="Topic"/>
    <subrole_main frame="FR_Judgment_communication" role="Topic" macro="Topic"/>
    <subrole_main frame="FR_Judgment_communication_no_reason" role="Topic" macro="Topic"/>
    <subrole_main frame="FR_Make_agreement" role="Topic" macro="Topic"/>
    <subrole_main frame="FR_Make_agreement_on_assessment" role="Topic" macro="Topic"/>
    <subrole_main frame="FR_Memory-Remembering_experience-Remembering_information" role="Topic" macro="
        Topic"/>
    <subrole_main frame="FR_Questioning_oneself" role="Topic" macro="Topic"/>
    <subrole_main frame="FR_Request" role="Topic" macro="Topic"/>
    <subrole_main frame="FR_Speak_on_topic" role="Topic" macro="Topic"/>
    <subrole_main frame="FR_Statement-manner-noise" role="Topic" macro="Topic"/>
    <subrole_main frame="FR_Telling" role="Topic" macro="Topic"/>
    <subrole_main frame="Judgment" role="Topic" macro="Topic"/>
    <subrole_main frame="Judgment_direct_address" role="Topic" macro="Topic"/>
    <subrole_main frame="Make_agreement_on_action" role="Topic" macro="Topic"/>
    <subrole_main frame="Prevarication" role="Topic" macro="Topic"/>
    <subrole_main frame="Quarreling" role="Issue" macro="Topic"/>
    <subrole_main frame="Questioning" role="Topic" macro="Topic"/>
    <subrole_main frame="Reveal_secret" role="Topic" macro="Topic"/>
  </macrorole>
  <macrorole name="True_content">
    <subrole_main frame="Be_in_agreement_on_assessment" role="Opinion" macro="True_content"/>
    <subrole_main frame="Categorization" role="Category" macro="True_content"/>
    <subrole_main frame="Coming_to_believe" role="Content" macro="True_content"/>
    <subrole_main frame="Evoking" role="Phenomenon" macro="True_content"/>
    <subrole_main frame="FR_Attempt_suasion" role="True_content" macro="True_content"/>
    <subrole_main frame="FR_Attempt_suasion.veracity" role="State_of_affairs" macro="True_content"/>
    <subrole_main frame="FR_Awareness-Certainty-Opinion" role="Content" macro="True_content"/>
    <subrole_main frame="FR_Be_in_agreement" role="Opinion" macro="True_content"/>
    <subrole_main frame="FR_Becoming_aware" role="Phenomenon" macro="True_content"/>
    <subrole_main frame="FR_Cognizer_affecting" role="True_content" macro="True_content"/>
    <subrole_main frame="FR_Cognizer_affecting.veracity" role="Content" macro="True_content"/>
    <subrole_main frame="FR_Expectation" role="Phenomenon" macro="True_content"/>
    <subrole_main frame="FR_Expressing_stance" role="True_content" macro="True_content"/>
    <subrole_main frame="FR_Make_agreement" role="Opinion" macro="True_content"/>
    <subrole_main frame="FR_Make_agreement_on_assessment" role="Opinion" macro="True_content"/>
    <subrole_main frame="FR_Memory-Remembering_experience-Remembering_information" role="Mental_content"
        macro="True_content"/>
    <subrole_main frame="FR_Personal_stance" role="True_content" macro="True_content"/>
    <subrole_main frame="FR_Questioning_oneself" role="Question" macro="True_content"/>
    <subrole_main frame="FR_Reliance_on_expectation" role="Expectation" macro="True_content"/>
    <subrole_main frame="FR_Remembering" role="Content" macro="True_content"/>
    <subrole_main frame="Religious_belief" role="Content" macro="True_content"/>
    <subrole_secondary frame="FR_Attributed_information" role="Proposition" macro="True_content"/>
    <subrole_secondary frame="FR_Proving" role="Content" macro="True_content"/>
    <subrole_secondary frame="Predicting" role="Eventuality" macro="True_content"/>
  </macrorole>
  <macrorole name="Unit">
    <subrole_main frame="Commerce_buy" role="Unit" macro="Unit"/>
    <subrole_main frame="Commerce_collect" role="Unit" macro="Unit"/>
    <subrole_main frame="Commerce_goods-transfer" role="Unit" macro="Unit"/>
    <subrole_main frame="Commerce_money-transfer" role="Unit" macro="Unit"/>
    <subrole_main frame="Commerce_pay" role="Unit" macro="Unit"/>
    <subrole_main frame="Commerce_scenario" role="Unit" macro="Unit"/>
    <subrole_main frame="Commerce_sell" role="Unit" macro="Unit"/>
    <subrole_main frame="Commercial_transaction" role="Unit" macro="Unit"/>
    <subrole_main frame="Earnings_and_losses" role="Unit" macro="Unit"/>
    <subrole_main frame="FR_Getting_money" role="Unit" macro="Unit"/>
    <subrole_main frame="FR_Spending" role="Unit" macro="Unit"/>
    <subrole_main frame="Renting" role="Unit" macro="Unit"/>
    <subrole_main frame="Renting_out" role="Unit" macro="Unit"/>
  </macrorole>
</macroroles>
```

BIBLIOGRAPHIE

ABEILLÉ, A. (2004). *Guide des annotateurs : Annotation fonctionnelle*.

ABEILLÉ, A., CLÉMENT, L. et TOUSSENEL, F. (2003). Building a Treebank for French. *In Treebanks : Building and Using Parsed Corpora*, pages 165–188. Springer.

AHLBERG, M., BORIN, L., DANNÉLLS, D., FORSBERG, M., TOPOROWSKA GRONOSTAJ, M., FRIBERG HEPPIN, K., JOHANSSON, R., KOKKINAKIS, D., OLSSON, L.-J. et UPPSTRÖM, J. (2014). Swedish FrameNet++ The Beginning of the End and the End of the Beginning.

ASHWINI, V., CHOI, J., PALMER, M. et NARASIMHAN, B. (2011). Analysis of the Hindi Proposition Bank using Dependency Structure. *In Proceedings of the fifth Linguistic Annotation Workshop. ACL 2011*, pages 21–29, Portland.

BAKER, C. F. (2009). La sémantique des cadres et le projet FRAMENET : une approche différente de la notion de « valence ». *Langages*, 176.

BAKER, C. F., FILLMORE, C. J. et CRONIN, B. (2003). The Structure of the Framenet Database. *International Journal of Lexicography*, 16(3):281–296.

BAKER, C. F., FILLMORE, C. J. et LOWE, J. B. (1998). The Berkeley FrameNet Project. *In Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics - Volume 1*, ACL '98, pages 86–90, Stroudsburg, PA, USA. Association for Computational Linguistics.

BARZDINS, G. (2014). Framenet CNL : A Knowledge Representation and Information Extraction Language. *In* DAVIS, B., KALJURAND, K. et KUHN, T., éditeurs : *Controlled Natural Language : 4th International Workshop, CNL 2014, Galway, Ireland, August 20-22, 2014. Proceedings*, pages 90–101, Cham. Springer International Publishing.

BENAMARA, F., CHARDON, B., MATHIEU, Y. Y. et POPESCU, V. (2011). Towards Context-Based Subjectivity Analysis. *In The 5th International Joint Conference on Natural Language Processing IJCNLP 2011*, Chiang Mai, Thailand.

BERMAN, R. A. (2005). Introduction : Developing discourse stance in different text types and languages. *Journal of pragmatics*, 37(2):105–124.

BHATTACHARYYA, P. (2010). IndoWordNet. *In* Calzolari *et al.* (2010).

BIBER, D. et FINEGAN, E. (1989). Styles of stance in English : lexical and grammatical marking of evidentiality and affect. *Text*, 9:93–124.

BITTAR, A., AMSILI, P., DENIS, P. et DANLOS, L. (2011). French TimeBank : An ISO-TimeML Annotated Reference Corpus. *In ACL 2011 - 49th Annual Meeting of the Association for Computational Linguistics*, pages 130–134, Portland, Oregon, United States. Association for Computational Linguistics.

BLANCHE-BENVENISTE, C. (1991). Deux relations de solidarité utiles pour l'analyse de l'attribut. *In* GAULMYN, M., RÉMI-GIRAUD, S. et BASSET, L., éditeurs : *À la recherche de l'attribut*, pages 83–97. Presses universitaires de Lyon.

BOAS, H. C. (2002). Bilingual FrameNet Dictionaries for Machine Translation. *In* RODRÍGUEZ, M. G. et ARAUJO, C. P. S., éditeurs : *Proceedings of the Third International Conference on Language Resources and Evaluation*, volume IV, pages 1364–1371, Las Palmas.

BOAS, H. C., éditeur (2009a). *Multilingual FrameNets in computational lexicography : methods and applications*. Trends in linguistics. Mouton de Gruyter, Berlin, New York.

BOAS, H. C. (2009b). Semantic frames as interlingual representations for multilingual lexical databases. *In* Boas (2009a).

BONAMI, O. (1999). *Les constructions du verbe : le cas des groupes prépositionnels argumentaux*. Thèse de doctorat, Université Paris 7.

BONIAL, C., BONN, J., CONGER, K., HWANG, J. D. et PALMER, M. (2014). Propbank : Semantics of New Predicate Types. *In* CHAIR), N. C. C., CHOUKRI, K., DECLERCK, T., LOFTSSON, H., MAEGAARD, B., MARIANI, J., MORENO, A., ODIJK, J. et PIPERIDIS, S., éditeurs : *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland. European Language Resources Association (ELRA).

BONIAL, C., BONN, J., CONGER, K., HWANG, J. D., PALMER, M. et REESE, N. (2015). *English PropBank Annotation Guidelines*.

BONIAL, C., STOWE, K. et PALMER, M. (2013). Renewing and Revising SemLink. *In Proceedings of the 2nd Workshop on Linked Data in Linguistics (LDL-2013) : Representing and linking lexicons, terminologies and other language data*, pages 9 – 17, Pisa, Italy. Association for Computational Linguistics.

BORIN, L., DANNÉLLS, D., FORSBERG, M., TOPOROWSKA GRONOSTAJ, M. et KOKKINAKIS, D. (2009). Thinking Green : Toward Swedish FrameNet++. *In FrameNet Masterclass and Workshop*.

BORKIN, A. (1974). *Raising to Object Position*. Thèse de doctorat, University of Michigan.

BRANTS, S., DIPPER, S., HANSEN, S., LEZIUS, W. et SMITH, G. (2002). The TIGER treebank. *In Proceedings of the Workshop on Treebanks and Linguistic Theories, September 20-21 (TLT02)*, pages 24–41, Sozopol, Bulgaria.

BURCHARDT, A., ERK, K., FRANK, A., KOWALSKI, A. et PADO, S. (2006a). SALTO : A versatile multi-level annotation tool. *In Proceedings of LREC-2006*, Genoa, Italy.

BURCHARDT, A., ERK, K., FRANK, A., KOWALSKI, A., PADÓ, S. et PINKAL, M. (2006b). The SALSA Corpus : a German corpus resource for lexical semantics. *In Proceedings of LREC 2006*.

BURCHARDT, A., ERK, K., FRANK, A., KOWALSKI, A., PADÓ, S. et PINKAL, M. (2009a). Framenet for the semantic analysis of German : Annotation, representation and automation. *In* BOAS, H. C., éditeur : *Multilingual FrameNets in Computational Lexicography : Methods and Applications*, volume 200 de *Trends in Linguistics*, pages 209–244. Mouton de Gruyter.

BURCHARDT, A., ERK, K., FRANK, A., KOWALSKI, A., PADÓ, S. et PINKAL, M. (2009b). Using FrameNet for the semantic analysis of German : Annotation, representation, and automation. *In* Boas (2009a).

BURCHARDT, A. et PENNACCHIOTTI, M. (2008). FATE : a framenet-annotated corpus for textual entailment. *In Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco. European Language Resources Association (ELRA).

BURCHARDT, A., PENNACHIOTTI, M., THATER, S. et PINKAL, M. (2009c). Assessing the impact of frame semantics on textual entailment. *Natural Language Engineering*, 15(Special Issue 04):527–550.

BUTT, M. (2006). *Theories of Case*. Cambridge textbooks in linguistics. Cambridge University Press.

CALZOLARI, N., CHOUKRI, K., MAEGAARD, B., MARIANI, J., ODIJK, J., PIPERIDIS, S., ROSNER, M. et TAPIAS, D., éditeurs (2010). *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA).

CANDITO, M., AMSILI, P., BARQUE, L., BENAMARA, F., de CHALENDAR, G., DJEMAA, M., HAAS, P., HUYGHE, R., MATHIEU, Y. Y., MULLER, P., SAGOT, B. et VIEU, L. (2014a). Developing a French FrameNet : Methodology and First results. *In Proceedings of the 9th international conference on Language Resources and Evaluation (LREC 2014)*, Reykjavik, Iceland.

CANDITO, M., CRABBÉ, B. et DENIS, P. (2010). Statistical French Dependency Parsing : Treebank Conversion and First Results. *In* Calzolari *et al.* (2010).

CANDITO, M., PERRIER, G., GUILLAUME, B., RIBEYRE, C., FORT, K., SEDDAH, D. et VILLEMONTE DE LA CLERGERIE, É. (2014b). Deep Syntax Annotation of the Sequoia French Treebank. *In International Conference on Language Resources and Evaluation (LREC)*, Reykjavik, Iceland.

CANDITO, M. et SEDDAH, D. (2012). Le corpus Sequoia : annotation syntaxique et exploitation pour l'adaptation d'analyseur par pont lexical. *In Actes de la 19e conférence sur le Traitement Automatique des Langues Naturelles (TALN'2012)*, pages 321–334, Grenoble, France.

CARTER, R. (1976). Some linking regularities. *In* LEVIN, B. et TENNY, C., éditeurs : *On Linking : Papers by Richard Carter (1988)*. Lexicon Project, Center for Cognitive Science, Massachusetts Institute of Technology.

CHAFE, W. J. (1986). Evidentiality in English conversation and academic writing. *In* CHAFE, W. L. et NICHOLS, J., éditeurs : *Evidentiality : the Linguistic Coding of Epistemology*, pages 261–272. Ablex, Norwood, New Jersey.

CHARDON, B., BENAMARA, F., MATHIEU, Y., POPESCU, V. et ASHER, N. (2013). Measuring the Effect of Discourse Structure on Sentiment Analysis. *In 14th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing 2013)*, pages pp. 25–37, Samos, Greece.

CHOMSKY, N. (1981). *Lectures on government and binding : the Pisa lectures*. Studies in generative grammar. Foris Publications, Dordrecht.

CHOMSKY, N. (1982). *Some Concepts and Consequences of the Theory of Government and Binding*. MIT Press.

CHOMSKY, N. (1986). *Barriers*. Linguistic inquiry. MIT Press.

COLINET, M., DANLOS, L., DARGNAT, M. et WINTERSTEIN, G. (2014). Emplois de la préposition "pour" suivie d'une infinitive : description, critères formels et annotation en corpus. *SHS Web of Conferences*, 8:3041–3058.

CRUSE, D. A. (1973). Some thoughts on agentivity. *Journal of Linguistics*, 9(1):pp. 11–23.

CRUSE, D. A. (1982). On lexical ambiguity. *Nottingham Linguistic Circular*, 11(2):65–80.

CRUSE, D. A. (2000). Aspects of the Micro-structure of Word Meanings. *In* Ravin et Leacock (2000b).

D'AGOSTINO, E., ELIA, A. et VIETRI, S. (2004). Lexicon-Grammar, Electronic Dictionaries and Local Grammars of Italian. *In* LECLÈRE, C., LAPORTE, E., PIOT, M. et SILBERZTEIN, M., éditeurs : *Lexique, Syntaxe et Lexique-Grammaire. Papers in honour of Maurice Gross*, pages 125–136.

DANLOS, L. (1992). Support verb constructions : linguistic properties, representation, translation. *Journal of French Language Studies*, 2:1–32.

DANLOS, L., PRADET, Q., BARQUE, L., NAKAMURA, T. et CONSTANT, M. (2016). Un Verbenet du français. *Traitement Automatique des Langues*, 57(1):25.

DAS, D., SCHNEIDER, N., CHEN, D. et SMITH, N. A. (2010). Probabilistic frame-semantic parsing. *In Proc. of NAACL HLT 2010*, pages 948–956, Stroudsburg, PA, USA. Association for Computational Linguistics.

DERWOJEDOWA, M., ZAWISŁAWSKA, M. et LINDE-USIEKNIEWICZ, J. (2008). Framenet for Polish. *In Dritte Internationale Konferenz der Deutschen Gesellschaft für Kognitive Linguistik (DGKL-08/GCLA-08)*.

DJEMAA, M. (2014). Traitement FrameNet des constructions à attribut de l'objet. *In Proceedings of TALN 2014 (Volume 4 : RECITAL - Student Research Workshop)*, pages 13–24, Marseille, France. Association pour le Traitement Automatique des Langues.

DOLBEY, A. E. (2009). *BioFrameNet : a FrameNet Extension to the Domain of Molecular Biology*. Thèse de doctorat, University of California, Berkeley.

DOWTY, D. (1991). Thematic Proto-Roles and Argument Selection. *Language*, 67(3):547–619.

DUBOIS, J. et DUBOIS-CHARLIER, F. (1997). *Les verbes français*. Larousse-Bordas.

FABER, P., MÁRQUEZ LINARES, C. et VEGA EXPÓSITO, M. (2005). Framing Terminology : A Process-Oriented Approach. *Meta*, 50(4).

FABRE, C. et BOURIGAULT, D. (2008). Exploiter des corpus annotés syntaxiquement pour observer le continuum entre arguments et circonstants. *Journal of French Language Studies*, 18(1):87–102.

FELLBAUM, C., éditeur (1998). *WordNet : An Electronic Lexical Database*. Language, Speech and Communication. MIT Press.

FIELD, M. (1997). The role of factive predicates in the indexicalization of stance : A discourse perspective. *Journal of Pragmatics*, 27(6):799 – 814.

FILLMORE, C. J. (1968). The Case for Case. *In* BACH, E. et HARMS, R., éditeurs : *Universals in Linguistic Theory*. Holt, Rinehart, and Winston, New York.

FILLMORE, C. J. (1976). Frame semantics and the nature of language. *Annals of the New York Academy of Sciences*, 280(1):20–32.

FILLMORE, C. J. (1977a). The Case for Case reopened. *In* COLE, P. et SADOCK, J. M., éditeurs : *Syntax and Semantics*, volume 8 : Grammatical Relations, pages 59–81. Academic Press Inc.

FILLMORE, C. J. (1977b). Scenes-and-frames semantics. *In* ZAMPOLLI, A., éditeur : *Linguistic Structures Processing*, Fundamental Studies in Computer Science. North Holland Publishing.

FILLMORE, C. J. (1982). *Frame semantics*, pages 111–137. Hanshin Publishing Co., Seoul, South Korea.

FILLMORE, C. J. (1985). Frames and the semantics of understanding. *Quaderni di Semantica*, 6(2):222–254.

FILLMORE, C. J. (2007). Valency issues in FrameNet. *In* HERBST, T. et GÖTZ-VOTTELER, K., éditeurs : *Valency : Theoretical, descriptive and cognitive issues*, volume 187 de *Trends in Linguistics. Studies and Monographs*, pages 129–160. Walter de Gruyter.

FILLMORE, C. J. et ATKINS, B. T. (1992). Toward a frame-based lexicon : The semantics of RISK and its neighbors. *In Frames, fields, and contrasts : New essays in semantic and lexical organization*. Hillsdale : Lawrence Erlbaum Associates.

FILLMORE, C. J., JOHNSON, C. R. et PETRUCK, M. R. (2003). Background to Framenet. *International Journal of Lexicography*, 16(3):235–250.

FRANÇOIS, J., LE PESANT, D. et LEEMAN, D. (2007). Présentation de la classification des Verbes Français de Jean Dubois et Françoise Dubois-Charlier. *Langue française*, 1(153):3–19.

FRIBERG HEPPIN, K. et TOPOROWSKA GRONOSTAJ, M. (2014). Exploiting FrameNet for Swedish : Mismatch ? *Constructions and Frames*, 6(1):52–72.

GAIFMAN, H. (1965). Dependency systems and phrase-structure systems. *Information and Control*, 8(3):304 – 337.

GAWRON, J. M. (2011). Frame Semantics. *In* MAIENBORN, C., von HEUSINGER, K. et PORTNER, P., éditeurs : *Semantics : An International Handbook of Natural Language Meaning*, volume 1, chapitre 29, pages 664–87. Mouton de Gruyter, Berlin.

GEERAERTS, D. (1993). Vagueness's puzzles, polysemy's vagaries. *Cognitive Linguistics*, 4(3):223–272.

GHNEIM, N., KARHELY, E. et SAFI, W. (2009). First Step of Building an Arabic FrameNet (AFN). *In 13th International Business Information Management Conference (13th IBIMA), Arabic Information Processing*.

GOOSSENS, V. (2011). *Propositions pour une modélisation de la polysémie régulière des noms d'affect*. Theses, Université de Grenoble.

GROSS, M. (1975). *Méthodes en syntaxe : régime des constructions complétives*. Actualités scientifiques et industrielles. Hermann.

GRUBER, J. S. (1965). *Studies in lexical relations*. Thèse de doctorat, Massachussets Institute of Technology.

GUIMIER, É. (1998). Les constructions à prédicat de l'objet des verbes de perception. *In M. Forsgren, K. Jonasson & H. Kronning (éds), Prédication, assertion, information. Actes du colloque d'Uppsala, Uppsala : Acta Universitatis Upsaliensis*, pages 231–241.

GUIMIER, É. (1999). *Les constructions à prédicat de l'objet en français : aspects syntaxiques, interprétatifs et formels*. Thèse de doctorat, Université Paris VII-Denis Diderot.

HAO, X.-y., LIU, W., LI, R. et KAI-YING, L. (2007). Description Systems of the Chinese FrameNet Database and Software Tools. *JOURNAL OF CHINESE INFORMATION PROCESSING*, 21(5):96–100.

HAYOUN, A. et ELHADAD, M. (2015). The Hebrew FrameNet Project. *In ISCOL 2015*.

HAYS, D. G. (1964). Dependency Theory : A Formalism and Some Observations. *Language*, 40(4):511–525.

HOEKSTRA, T. (1992). Small clause theory. *Belgian Journal of Linguistics*, 7(1):125–151.

JACKENDOFF, R. (1987). The status of thematic relations in linguistic theory. *Linguistic Inquiry*, 18(3):pp. 369–411.

JACQUIN, C., MONCEAUX, L. et DESMONTILS, E. (2006). Systèmes Question-Réponse et EuroWordNet. *In 13ème conférence sur le Traitement Automatique des Langues Naturelles (TALN 2006)*, pages 188–197, Belgium.

JOHNSON, C. R., FILLMORE, C. J., PETRUCK, M. R., BAKER, C. F., ELLSWORTH, M. J., RUPPENHOFER, J. et WOOD, E. J. (2003). FrameNet : Theory and Practice.

JOHNSON, C. R., FILLMORE, C. J., WOOD, E. J., RUPPENHOFER, J., URBAN, M., PETRUCK, M. R. et BAKER, C. F. (2001). The FrameNet project : Tools for lexicon building.

KABBACH, A. et OHARA, K. H. (2015). Japanese FrameNet data quality overview : Toward full data integrity. *In MAPLEX2015*, Yamagata, Japan.

KARTTUNEN, L. (1971). Some observations on factivity. *Papers in Linguistics*, 1(4).

KIPARSKY, P. et KIPARSKY, C. (1970). Fact. *In* BIERWISCH, M. et HEIDOLPH, K. E., éditeurs : *Progress in Linguistics*, page 143–173. Mouton, The Hague.

KIPPER, K., DANG, H. T. et PALMER, M. (2000). Class-based construction of a verb lexicon. *In Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence*, pages 691–696. AAAI Press.

KIPPER, K., KORHONEN, A., RYANT, N. et PALMER, M. (2008). A large-scale classification of English verbs. *Language Resources and Evaluation*, 42(1):21–40.

KIPPER SCHULER, K. (2005). *VerbNet : A Broad-Coverage, Comprehensive Verb Lexicon*. Thèse de doctorat, University of Pennsylvania.

KLEIBER, G. (2008). Petit essai pour montrer que la polysémie n'est pas un sens interdit. *In Actes du CMLF 2008 - Congrès Mondial de Linguistique Française 2008*, pages 87–101. EDP Sciences.

LAKOFF, G. et JOHNSON, M. (1980). *Metaphors we Live by*. University of Chicago Press, Chicago.

LE GOFFIC, P. (1994). *Grammaire de la phrase française - Livre de l'élève - Edition 1994*. HU Langue française. Hachette Éducation.

LECLÈRE, C. (1990). Organisation du lexique-grammaire des verbes français. *Langue française*, 87(1): 112–122.

LENCI, A., JOHNSON, M. et LAPESA, G. (2010). Building an Italian FrameNet through semi-automatic corpus analysis. *In* Calzolari *et al.* (2010).

LEVIN, B. (1993). *English Verb Classes and Alternations : A Preliminary Investigation*. University of Chicago Press.

LEVIN, B. et RAPPAPORT HOVAV, M. (2005). *Argument Realization*. Research Surveys in Linguistics. Cambridge University Press.

LEWANDOWSKA-TOMASZCZYK, B. (2007). Polysemy, prototypes, and radial categories. *In* GEERAERTS, D. et CUYCKENS, H., éditeurs : *The Oxford Handbook of Cognitive Linguistics*, pages 139–169. Oxford University Press.

L'HOMME, M.-C. (2016). Terminologie de l'environnement et sémantique des cadres. *In Congrès Mondial de Linguistique Française - CMLF 2016*.

LOWE, J. B., BAKER, C. F. et FILLMORE, C. J. (1997). A Frame-Semantic Approach to Semantic Annotation. *In Tagging Text with Lexical Semantics : Why, What, and How ? Proceedings of the Workshop*, pages 18–24. Special Interest Group on the Lexicon, Special Interest Group on the Lexicon.

LUX-POGODALLA, V. et POLGUÈRE, A. (2011). Construction of a French lexical network : Methodological issues. *In Proceedings of the International Workshop on Lexical Resources (WoLeR 2011), ESSLLI*, Ljubljana.

LÖNNEKER-RODMAN, B. et BAKER, C. F. (2009). The FrameNet model and its applications. *Natural Language Engineering*, 15:415–453.

MARCUS, M., KIM, G., MARCINKIEWICZ, M. A., MACINTYRE, R., BIES, A., FERGUSON, M., KATZ, K. et SCHASBERGER, B. (1994). The Penn Treebank : Annotating predicate argument structure. *In Proceedings of the Workshop on Human Language Technology*, HLT '94, pages 114–119, Stroudsburg, PA, USA. Association for Computational Linguistics.

MARCUS, M. P., SANTORINI, B. et MARCINKIEWICZ, M. A. (1993). Building a large annotated corpus of English : The Penn Treebank. *Computational Linguistics*, 19(2):313–330.

MEL'ČUK, I., CLAS, A. et POLGUÈRE, A. (1995). *Introduction à la lexicologie explicative et combinatoire*. Duculot, Paris/Louvain-la-Neuve.

MERLO, P., STEVENSON, S., TSANG, V. et ALLARIA, G. (2002). A multilingual paradigm for automatic verb classification. *In Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL-02)*.

MICHALON, O. (2014). Modélisation probabiliste de l'interface syntaxe sémantique à l'aide de grammaires hors contexte probabilistes : Expériences avec framenet. *In Proceedings of RECITAL 2014*.

MICHALON, O., RIBEYRE, C., CANDITO, M. et NASR, A. (2016). Deeper syntax for better semantic parsing. *In Coling 2016 - 26th International Conference on Computational Linguistics*, Osaka, Japan.

MILLER, G. A. (1995). WordNet : A lexical database for english. *Communications of the ACM*, 38(11):39–41.

MILLER, P. H. (1997). Compléments vs. circonstants : une distinction syntaxique ou sémantique ? *In Congrès de la SAES 1997, "L'esprit des lieux : passages et rivages"*, pages 91–103, Nice, France. Souesme, J.-C. (ed.).

MILTSAKAKI, E., PRASAD, R., JOSHI, A. K. et WEBBER, B. L. (2004). The Penn Discourse Treebank. *In LREC*.

MØRDRUP, O. (1975). Présuppositions, implications et verbes français. *Revue romane*, 1(10):125–155.

MOREL, M. et FRANÇOIS, J. (2015). Le Dictionnaire Electronique des Synonymes du CRISCO : un outil de plus en plus interactif. *Revue Française de Linguistique Appliquée*, XX(1):9–28.

MOUSSER, J. (2010). A large coverage verb taxonomy for Arabic.

MOUTON, C. et de CHALENDAR, G. (2010). JAWS : Just another WordNet subset. *In* ATALA, éditeur : *Proc. of TALN 2010*.

MOUTON, C., de CHALENDAR, G. et RICHERT, B. (2010). Framenet Translation Using Bilingual Dictionaries with Evaluation on the English-French Pair. *In* Calzolari *et al.* (2010).

MULLER, C. (2000). Les constructions à adjectif attribut de l'objet, entre prédication seconde et complémentation verbale. *Langue française*, 127(1):21–35.

MUSHIN, I. (2001). *Evidentiality and Epistemological Stance : Narrative Retelling*. Numéro 87 de Pragmatics & Beyond New Series. John Benjamins Publishing company, Amsterdam.

NAM, J.-s. (2004). Some linguistic problems in building a Korean electronic lexicon of simple verbs. *In* LECLÈRE, C., LAPORTE, E., PIOT, M. et SILBERZTEIN, M., éditeurs : *Lexique, Syntaxe et Lexique-Grammaire. Papers in Honour of Maurice Gross*, pages 455–469.

NAM, S., PARK, J., KIM, Y., HAHM, Y., HWANG, D. et CHOI, K.-S. (2014). Korean FrameNet For Semantic Analysis. *In Submitted to ISWC*.

NILSEN, D. L. F. (1973). *The instrumental case in English ; syntactic and semantic considerations, by Don L. F. Nilsen*. Mouton, The Hague.

OHARA, K. H. (2009). Frame-based contrastive lexical semantics in Japanese FrameNet : The case of risk and kakeru. *In* Boas (2009a).

OHARA, K. H., FUJII, S., OHORI, T., SUZUKI, R., SAITO, H. et ISHIZAKI, S. (2004). The Japanese Framenet project : An introduction. *In Proceedings of the Satellite Workshop "Building Lexical Resources from Semantically Annotated Corpora"*, Lisbon, Portugal. The Fourth international conference on Language Resources and Evaluation (LREC 2004).

OLSSON, K. (1976). *La Construction verbe + objet direct + complément prédicatif en français : aspects syntaxiques et sémantiques*. Thèse de doctorat, Université de Stockholm.

OSTLER, N. (1979). *Case linking : A theory of case and verb diathesis, applied to Classical Sanskrit*. Thèse de doctorat, Massachusetts Institute of Technology.

OVCHINNIKOVA, E., VIEU, L., OLTRAMARI, A., BORGO, S. et ALEXANDROV, T. (2010). Data-driven and Ontological Analysis of FrameNet for Natural Language Reasoning. *In* Calzolari *et al.* (2010).

PADÓ, S. (2007). *Cross-Lingual Annotation Projection Models for Role-Semantic Information*. Thèse de doctorat, Saarland University.

PADÓ, S. et LAPATA, M. (2005). Cross-linguistic Projection of Role-Semantic Information. *In HLT/EMNLP 2005, Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, Vancouver, British Columbia, Canada. The Association for Computational Linguistics.

PADÓ, S. et LAPATA, M. (2006). Optimal constituent alignment with edge covers for semantic projection. *In* CALZOLARI, N., CARDIE, C. et ISABELLE, P., éditeurs : *ACL 2006, 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, Sydney, Australia. The Association for Computer Linguistics.

PADÓ, S. et PITEL, G. (2007). Annotation précise du français en sémantique de rôles par projection cross-linguistique. *In Proceedings of TALN-07*, Toulouse, France.

PALMER, F. R. (1986). *Mood and Modality*. Cambridge University Press, Cambridge, England.

PALMER, M. (2009). SemLink : Linking PropBank, VerbNet and FrameNet. *In Proceedings of the Generative Lexicon Conference GenLex-09*, Pisa, Italy.

PALMER, M., BABKO-MALAYA, O., BIES, A., DIAB, M., MAAMOURI, M., MANSOURI, A. et ZAGHOUANI, W. (2008). A Pilot Arabic Propbank. *In* CHAIR), N. C. C., CHOUKRI, K., MAEGAARD, B., MARIANI, J., ODIJK, J., PIPERIDIS, S. et TAPIAS, D., éditeurs : *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco. European Language Resources Association (ELRA).

PALMER, M., GILDEA, D. et KINGSBURY, P. (2005). The Proposition Bank : An annotated corpus of semantic roles. *Computational linguistics*, 31(1):71–106.

PALMER, M., RYU, S., CHOI, J., YOON, S. et JEON, Y. (2006). Korean PropBank.

PALMER, M. S., GILDEA, D. et XUE, N. (2010). *Semantic role labeling*. Synthesis lectures on human language technologies. Morgan & Claypool, San Rafael (Calif.).

PARENT, G., GAGNON, M. et MULLER, P. (2008). Annotation d'expressions temporelles et d'événements en français. *In* BÉCHET, F., éditeur : *Traitement Automatique des Langues Naturelles (TALN), Avignon, 09/06/08-13/06/08*. Association pour le Traitement Automatique des Langues (ATALA).

PITEL, G. (2009). Cross-lingual labeling of semantic predicates and roles : A low-resource method based on bilingual L(atent) S(emantic) A(nalysis). *In* Boas (2009a).

POLLARD, C. et SAG, I. A. (1987). *Information-based Syntax and Semantics : Vol. 1 : Fundamentals*. Center for the Study of Language and Information, Stanford, CA, USA.

PRADET, Q. (2015). *Annotation en rôles sémantiques du français en domaine spécifique*. Thèse de doctorat, Université Paris Diderot (Paris 7).

PRADET, Q., BAGUENIER DESORMEAUX, J., DE CHALENDAR, G. et DANLOS, L. (2013). WoNeF : amélioration, extension et évaluation d'une traduction française automatique de WordNet. *In TALN 2013 - 20ème conférence du Traitement Automatique du Langage Naturel*, pages 76–89, Les Sables d'Olonne, France.

PRADET, Q., DANLOS, L. et de CHALENDAR, G. (2014). Adapting VerbNet to French using existing resources. *In Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*.

PRADHAN, S. S., HOVY, E. H., MARCUS, M. P., PALMER, M., RAMSHAW, L. A. et WEISCHEDEL, R. M. (2007). OntoNotes : A Unified Relational Semantic Representation. *In Proceedings of the First IEEE International Conference on Semantic Computing (ICSC 2007), September 17-19, 2007, Irvine, California, USA*, pages 517–526.

QUINE, W. V. O. (1960). *Word and Object*. The MIT Press.

RAVIN, Y. et LEACOCK, C. (2000a). Polysemy : an overview. *In* Ravin et Leacock (2000b).

RAVIN, Y. et LEACOCK, C., éditeurs (2000b). *Polysemy : Theoretical and Computational Approaches*. Oxford University Press.

REHBEIN, I., RUPPENHOFER, J., SPORLEDER, C. et PINKAL, M. (2012). Adding nominal spice to SALSA - frame-semantic annotation of German nouns and verbs. *In 11th Conference on Natural Language Processing, KONVENS 2012, Empirical Methods in Natural Language Processing, Vienna, Austria, September 19-21, 2012*, pages 89–97.

RÉMI-GIRAUD, S. (1991). Adectif attribut et prédicat. approche notionnelle et morpho-syntaxique. *In* GAULMYN, M., RÉMI-GIRAUD, S. et BASSET, L., éditeurs : *À la recherche de l'attribut*, pages 151–157. Presses universitaires de Lyon.

RIBEYRE, C., CANDITO, M. et SEDDAH, D. (2014). Semi-Automatic Deep Syntactic Annotations of the French Treebank. *In The 13th International Workshop on Treebanks and Linguistic Theories (TLT13)*, Proceedings of TLT 13, Tübingen, Germany. Tübingen Universität.

RIEGEL, M. (1981). Verbes essentiellement ou occasionnellement attributifs. *L'information grammaticale*, 10(1):23–27.

RIEGEL, M. (1991). Pour ou contre la notion grammaticale d'attribut de l'objet : critères et arguments. *In* GAULMYN, M., RÉMI-GIRAUD, S. et BASSET, L., éditeurs : *À la recherche de l'attribut*, pages 99–118. Presses universitaires de Lyon.

RIEGEL, M., PELLAT, J.-C. et RIOUL, R. (2004). *Grammaire méthodique du français*. Presses Universitaires de France.

ROSCH, E. et MERVIS, C. B. (1975). Family resemblances : Studies in the internal structure of categories. *Cognitive psychology*, 7(4):573–605.

ROTHSTEIN, S. (1995). Small clauses and copular constructions. *Small clauses*, pages 27–48.

ROZE, C., DANLOS, L. et MULLER, P. (2012). LEXCONN : A French Lexicon of Discourse Connectives. *Discours*, (10).

RUPPENHOFER, J., ELLSWORTH, M., PETRUCK, M. R., JOHNSON, C. R. et SCHEFFCZYK, J. (2006). *FrameNet II : Extended Theory and Practice*. International Computer Science Institute, Berkeley, California. Distributed with the FrameNet data.

RUWET, N. (1982). *Grammaire des insultes et autres études*. Travaux Linguistiques. Éditions du Seuil.

SAGOT, B. (2010). The Lefff, a freely available and large-coverage morphological and syntactic lexicon for French. *In* Calzolari *et al.* (2010).

SAGOT, B. et FIŠER, D. (2008). Construction d'un wordnet libre du français à partir de ressources multilingues. *In TALN 2008 -Traitement Automatique des Langues Naturelles*, Avignon, France.

SALOMÃO, M. M. M. (2009). FrameNet Brasil : um trabalho em progresso. *Calidoscópio*, 7(2):171–182.

SAURÍ, R. (2008). *A Factuality Profiler for Eventualities in Text*. Thèse de doctorat, Brandeis University.

SCARTON, C., SANCHES DURAN, M. et ALUÍSIO, S. (2014). Using Cross-Linguistic Knowledge to Build VerbNet-Style Lexicons : Results for a (Brazilian) Portuguese VerbNet. *In PROPOR 2014 - Computational Processing of the Portuguese Language*.

SCHMIDT, T. (2009). The Kicktionary — a multilingual lexical resource of football language. *In* BOAS, H. C., éditeur : *Multilingual FrameNets in Computational Lexicography : Methods and Applications*, volume 200 de *Trends in Linguistics*, pages 101–132. Mouton de Gruyter.

SEDDAH, D., TSARFATY, R., K'UBLER, S., CANDITO, M., CHOI, J., FARKAS, R., FOSTER, J., GOENAGA, I., GOJENOLA, K., GOLDBERG, Y., GREEN, S., HABASH, N., KUHLMANN, M., MAIER, W., NIVRE, J., PRZEPIORKOWSKI, A., ROTH, R., SEEKER, W., VERSLEY, Y., VINCZE, V., WOLIŃSKI, M., WRÓBLEWSKA, A. et Villemonte de la CLÉRGERIE, E. (2013). Overview of the spmrl 2013 shared task : A cross-framework evaluation of parsing morphologically rich languages. *In Proc. of the 4th Workshop on Statistical Parsing of Morphologically Rich Languages : Shared Task*, Seattle, WA.

SHEN, D. et LAPATA, M. (2007). Using Semantic Roles to Improve Question Answering. *In Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, page 12–21.

STALMASZCZYK, P. (1996). Theta roles and the theory of theta-binding. *Papers and Studies in Contrastive Linguistics*, 31:97–110.

STAMOU, S., OFLAZER, K., PALA, K., CHRISTOUDOULAKIS, D., CRISTEA, D., TUFIS, D., KOEVA, S., TOTKOV, G., DUTOIT, D. et GRIGORIADOU, M. (2002). BALKANET : A Multilingual Semantic Network for Balkan Languages. *In Proceedings of the International Wordnet Conference, Mysore, India*, pages 21–25.

STOWELL, T. (1995). Remarks on clause structure. *Syntax and semantics*, 28:271–286.

SUBIRATS, C. (2009). Spanish FrameNet : A frame-semantic analysis of the Spanish lexicon. *In* Boas (2009a).

SUBIRATS, C. et PETRUCK, M. R. (2003). Surprise : Spanish FrameNet! *In* HAJIČOVÁ, E., KOTÉŠOVCOVÁ, A. et MIROVSKÝ, J., éditeurs : *Proceedings of the Workshop on Frame Semantics, XVII International Congress of Linguists (CIL)*. Matfyzpress.

TESNIÈRE, L. (1959). *Éléments de Syntaxe Structurale*. Klincksieck, Paris.

TOBBACK, E. (2005). *Les constructions à attribut de l'objet et le marquage de la relation prédicative seconde*. Thèse de doctorat, Université de Gand.

TONELLI, S. (2010). *Semi-automatic techniques for extending the FrameNet lexical database to new languages*. Thèse de doctorat, Università Ca' Foscari, Venezia.

VETULANI, Z., OBRĘBSKI, T. et VETULANI, G. (2007). Towards a Lexicon-Grammar of Polish : Extraction of Verbo-Nominal Collocations from Corpora. *In* WILSON, D. et SUTCLIFFE, G., éditeurs : *Poster session of 19th International Florida Artificial Intelligence Research Society Conference (FLAIRS'07)*, pages 267–268.

VIEU, L., MULLER, P., CANDITO, M. et DJEMAA, M. (2016). A general framework for the annotation of causality based on framenet. *In LREC 2016*.

VOSKAKI, O. (2011). *The lexicon-grammar of Modern Greek verbs : transitive non locative constructions with one direct object*. Theses, Université Paris-Est.

VOSSEN, P., éditeur (1998). *EuroWordNet : A Multilingual Database with Lexical Semantic Networks*. Kluwer Academic Publishers, Norwell, MA, USA.

WILLEMS, D. et DEFRANCQ, B. (2000). L'attribut de l'objet et les verbes de perception. *Langue française*, pages 6–20.

WILLIAMS, E. (1975). Small clauses in English. *Syntax and semantics*, 4:249–273.

XUE, N. et PALMER, M. (2003). Annotating Propositions in the Penn Chinese Treebank. *In Proceedings of the Second Sighan Workshop*, Sapporo, Japan.

YOU, L. et LIU, K. (2005). Building Chinese FrameNet database. *In Proceedings of 2005 IEEE International Conference on Natural Language Processing and Knowledge Engineering (IEEE NLP-KE '05)*, pages 301–306.