

Les rôles sémantiques

Théories et applications

Mémoire

présenté par

Clémence Filmont

MASTER 2

« Ingénierie Linguistique »

Mention *Linguistique générale et linguistique appliquée*

Parcours Recherche & développement

dirigé par

Delphine Battistelli

soutenu le

16 juin 2014

Table des matières

TABLE DES MATIERES.....	2
INTRODUCTION	4
1. LES ROLES SEMANTIQUES DANS L'HISTOIRE DE LA LINGUISTIQUE : POINT DE VUE DIACHRONIQUE	4
1.1. « L'INVENTION DES RÔLES THÉMATIQUES EN LIEN AVEC LA TENTATIVE DE RÉSOUDRE LE PROBLÈME DU LINKING » (KASPER, 2008)	4
1.1.1. <i>De la séparation nette entre syntaxe et sémantique</i>	5
1.1.2. <i>...A leur rapprochement</i>	6
1.1.3. <i>D'une conception où les rôles sémantiques sont le fondement de la théorie du Linking</i>	7
1.1.3.1. Les rôles sémantiques impliquent la théorie du Linking.....	7
1.1.3.2. Les défis à la théorie du Linking.....	7
1.1.4. <i>...A l'autonomie des rôles sémantiques vis à vis de la théorie du Linking</i>	8
1.2. THEORIES SUR LES ROLES SEMANTIQUES	8
1.2.1. <i>La Grammaire des cas de (Fillmore, 1968)</i>	8
1.2.2. <i>La Sémantique des Frames de (Fillmore, 1976) : les rôles sémantiques comme concepts constitutifs de frames</i>	9
1.2.2.1. Les frames, une notion abstraite issue de l'Intelligence Artificielle et de la linguistique - exemple avec le frame anglais « commercial transaction »	9
1.2.2.2. Les frames dans la langue : interaction lexicale - syntaxe - sémantique	10
1.2.3. <i>La Théorie des Principes et Paramètres (Chomsky, 1981, 1986) : les rôles sémantiques comme éléments constitutifs de la grille thématique d'un prédicat</i>	12
1.2.4. <i>(Levin et al., 1988) : les rôles sémantiques comme « relation de sens » (Hadouche, 2011) entre le prédicat et ses arguments</i>	14
1.2.4.1. Niveau de représentation lexico-syntaxique ou Structure Prédicat-Argument (SPA)	14
1.2.4.2. Niveau de représentation lexico-sémantique.....	15
2. LES ROLES SEMANTIQUES : POINT DE VUE DESCRIPTIF	15
2.1. TERTIUM COMPARATIONI AUX DIFFERENTES THEORIES	15
2.1.1. <i>Le Linking</i>	15
2.1.1.1. Des théories plus ou moins en accord avec le Linking	15
2.1.1.2. Les différentes théories du Linking.....	16
2.1.2. <i>Mode de représentation des rôles sémantiques</i>	16
2.1.2.1. Les rôles sémantiques sont-ils des primitifs ?	16
2.1.2.2. Les rôles sémantiques peuvent-ils être représentés sous forme de liste?	17
2.2. DIFFERENTS ASPECTS DES ROLES SEMANTIQUES	19
2.2.1. <i>Ambiguïté dans le rôle de certains participants</i>	19
2.2.2. <i>Verbes au sens voisin</i>	19
2.2.3. <i>Alternances syntaxiques impliquant les rôles sémantiques : alternances de transitivité</i>	19
2.2.3.1. Alternance causatif/inchoatif.....	20
2.2.3.2. Alternance transitive intermédiaire	20
2.2.4. <i>Variation syntaxique d'un participant : actant - circonstant</i>	20
3. RESSOURCES LEXICOGRAPHIQUES INTEGRANT LA NOTION DE ROLE SEMANTIQUE	21
3.1. FRAMENET : INTEGRATION DES ROLES SEMANTIQUES DANS DES CLASSES DE CONCEPTS APPELES « FRAMES »	22
3.1.1. <i>Mise en relation des rôles sémantiques et de leurs réalisations lexicales</i>	22
3.1.2. <i>Représentation implicite de la hiérarchie des rôles sémantiques</i>	23

3.2. VERBNET : INTEGRATION DES ROLES SEMANTIQUES DANS DES CLASSES LINGUISTIQUES SEMANTICO-SYNTAXIQUES	25
3.2.1. Définition des classes Verbnets	25
3.2.2. D'un système de classification niant la possibilité de lister les rôles sémantiques à une description intégrant une liste de rôles	26
3.2.3. Les obstacles à une transposition de Verbnets dans d'autres langues	28
3.2.3.1. Le caractère idiosyncrasique des alternances	29
3.2.3.2. L'absence de spécification des traits sémantiques syntaxiquement pertinents dans la ressource anglaise	30
3.3. VERS UN VERBNET FRANÇAIS ? ESSAI D'ADAPTATION DE VERBNET AU FRANÇAIS PAR LE CEA LIST ET PARIS 7	30
3.3.1. Objectifs et spécifications	31
3.3.2. Méthode	31
3.3.2.1. 1ère étape : projection des classes générales anglaises avec leurs frames dans des classes françaises et traduction des verbes anglais	31
3.3.2.2. 2ème étape : filtrage semi-automatique des traductions françaises, validation manuelle des frames et division manuelle des classes en sous-classes	32
3.3.3. Mise en perspective	33
3.3.3.1. Dans quelle mesure peut-on adapter les alternances syntaxiques sans affecter la nature des classes anglaises ni violer la définition générale des classes de (Levin, 1993)?	33
3.3.3.2. Quelle méthodologie ?	34
3.3.3.3. Le choix de Verbnets pour l'annotation en rôles sémantiques des textes français	36
4. LES ROLES SEMANTIQUES DANS LE TAL AUJOURD'HUI	36
4.1. L'ANNOTATION EN ROLES SEMANTIQUES	36
4.1.1. Principes	36
4.1.2. Exemple de système d'annotation en rôles sémantiques basé sur les règles extraites du Verbnets anglais : le programme développé au CEA List	37
4.1.2.1. Les ressources utilisées	37
4.1.2.2. Fonctionnement du programme	37
4.1.2.3. Evaluation du programme d'annotation	41
4.2. APPLICATIONS TAL UTILISANT DES CORPUS ANNOTES EN ROLES SEMANTIQUES	42
4.2.1. Applications pour le TAL : la désambiguïsation du sens des verbes	42
4.2.2. Applications pour les technologies du langage	43
4.2.2.1. Traduction automatique	43
4.2.2.2. Extraction d'informations d'après (Mouton, 2010)	43
CONCLUSION	44
TABLE DES FIGURES ET TABLEAUX	46
BIBLIOGRAPHIE	47

Introduction

Les rôles sémantiques sont une notion linguistique apparue et revendiquée en tant que telle à partir des années 1960, notamment avec les travaux de (Fillmore, 1968) sur la Grammaire des cas. La désignant sous des étiquettes diverses (« cas », « élément de frame », « rôles thématiques », « théta-rôles »...), les linguistes partagent la même conception globale des rôles sémantiques comme « généralisation[s] sur les arguments d'un prédicat en vue de capturer des régularités dans l'interprétation (sémantique) des représentations syntaxiques. » (Amsili *et al.*, 2012)

Nous étudierons dans un premier temps le contexte historique et linguistique d'apparition de cette notion. Puis nous verrons les principales théories associées, certaines axées sur l'interface sémantique-syntaxe (La Grammaire des cas de (Fillmore, 1968), et (Petrucci, 1996)), d'autres sur les aspects sémantiques (La Sémantique des Frames de (Fillmore, 1976)). Ensuite, nous nous intéresserons aux descriptions parfois opposées qui ont été proposées des rôles sémantiques, que ce soit dans leur globalité ou sous certains de leurs aspects. Les deux dernières parties sont consacrées aux principaux traitements lexicographique et informatique aussi bien effectifs que potentiellement réalisables autour de la notion.

Ces deux dernières sections seront aussi l'occasion d'aborder les deux sujets sur lesquels j'ai travaillé lors de mon stage de Master 2 effectué au CEA List. Le premier porte sur l'intégration du programme d'annotation en rôles sémantiques du CEA dans Lima, qui est un système dédié à l'analyse linguistique automatique des textes multilingues. Conçue à partir de la ressource anglaise Verbnet, ce programme s'applique à des textes anglais. L'objectif défini par le deuxième sujet de stage est de participer à « l'adaptation de [...] Verbnet au français » à mon niveau, c'est-à-dire en vérifiant la validité du mapping entre les classes Verbnet françaises validées et certaines classes de la ressource lexico-syntaxique Les Verbes Français (LVF+1). Le résultat de l'adaptation de Verbnet au français doit permettre au programme d'annotation de s'appliquer aux textes français.

1. Les rôles sémantiques dans l'histoire de la linguistique : point de vue diachronique

1.1. « L'invention des rôles thématiques en lien avec la tentative de résoudre le problème du Linking » (Kasper, 2008)

On peut affirmer avec (Kasper, 2008) que la notion de rôle sémantique naît d'une intuition face aux ressemblances entre les représentations respectives des phrases de deux domaines jusque-là bien distincts : la syntaxe et la sémantique. C'est ce lien général que Kasper nomme « le linking ». « Un » linking désigne quant à lui un ensemble donné d'associations binaires argument syntaxique – rôle sémantique, c'est-à-dire la mise en relation d'une structure en rôles sémantiques avec une structure syntaxique.

Selon le même auteur, ce sont les travaux de Chomsky qui, dans les années 1950, ont déclenché - par un paradoxe apparent - l'essor des théories qui traitent des rôles sémantiques.

Il faut, pour comprendre cette idée, revenir quelques années en arrière.

1.1.1. De la séparation nette entre syntaxe et sémantique...

En 1930, Blake propose dans *A Semantic Analysis of Case* un modèle de représentation logique des phrases qui rompt avec le modèle binaire sujet-prédicat proposé par Aristote au IV^{ème} siècle avant JC. Ainsi, pour une phrase à la voix active et dont la tête est un verbe transitif, la représentation logique est la suivante :

Verbe (x,y)

Figure 1 Représentation d'après la logique des prédicats d'une phrase construite autour d'un verbe transitif

Avec la parution en 1957 de *Syntactic Structures*, Chomsky, loin de s'intéresser aux questions de sémantique, cherche pour sa part à mettre au jour les principes universels de bonne formation des phrases. Pour lui, ces principes sont à situer au niveau de la syntaxe, module autonome et indépendant de la sémantique. Pour étayer cette idée de séparation entre syntaxe et sémantique, il prend l'exemple d'une phrase qui est à la fois grammaticale et incohérente sémantiquement :

(1) *Colorless green ideas sleep furiously.*

Il parvient à la généralisation suivante : pour toute phrase composée d'une proposition principale et dont le(s) verbe(s) conjugué(s) est/sont transitif(s), les règles sous-jacentes et la représentation syntaxique¹ sont :

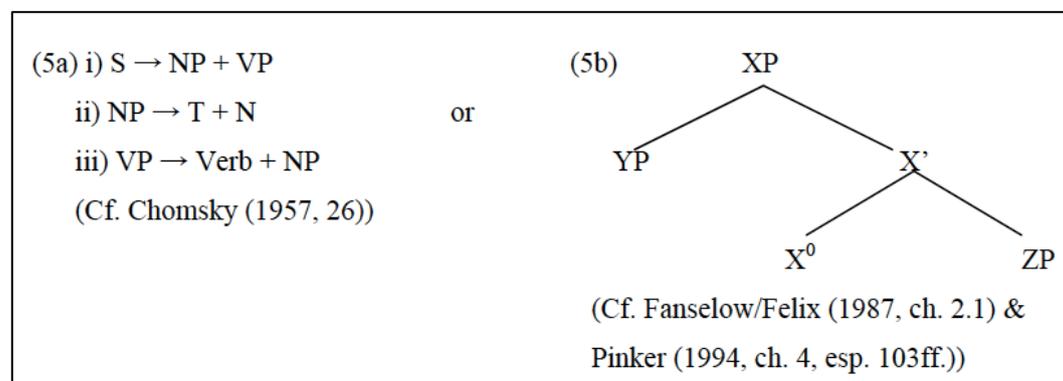


Figure 2 Règles de réécriture et représentation syntaxique des phrases à la voix active et composées d'une proposition principale en construction transitive (schéma emprunté à (Kasper, 2008))

Dans les deux figures, le symbole « P » signifie « syntagme » (de l'anglais « phrase »).

¹ Suivant la théorie X-barre

Sur la figure de gauche, le « S » (de l'anglais « sentence ») représente la phrase dans son ensemble. Les catégories morpho-syntaxiques du nom et du verbe sont représentées respectivement par les lettres « N » et « V ». « T » figure le déterminant.

Sur la figure de droite, « Y » représente la position structurale du spécifieur, et « Z » celle de complément.

1.1.2. ...A leur rapprochement

A partir des années 1960, le fait que Chomsky sépare les deux domaines relativement bien établis fait se poser la question à certains linguistes comme Gruber, (Fillmore, 1968) et Jackendoff de leur relation :

« Si l'on admet que la structure syntaxique d'une langue peut être vue comme constituée uniquement de catégories formelles du type de celles mentionnées dans la figure 2 et que la représentation sémantique d'une phrase donnée est représentée comme dans la figure 1, le problème peut donc être posé en terme de lien (*linking*). » (Kasper, 2008 : 21)

Kasper s'appuie sur la règle logique selon laquelle le rapprochement de plusieurs entités quelles qu'elles soient, impliquent qu'elles soient au départ séparées.

Le tableau suivant fait le lien entre les éléments sémantiques de la figure 1 et les éléments syntaxiques de la figure 2 :

figure	1	2 (schéma de gauche)	2 (schéma de droite)
Type de structure	logique	syntaxique (règles de réécriture)	syntaxique (représentation X-barre)
Le verbe	Verb	VP	XP
Elément 1	x	NP de (i)	L'adjonction YP
Elément 2	y	NP de (iii)	Le complément ZP

Tableau 1 Décomposition du *linking* de l'exemple des figures 1 et 2

L'idée d'un parallélisme entre les deux structures a ouvert la voie à la remise en cause des autonomies respectives de la syntaxe et de la sémantique :

« Les théories post-chomskiennes (1957), au premier rang desquelles celles de Chomsky (1965, 1981, 1995) lui-même - se sont empressées de s'affranchir de l'idée d'une syntaxe entièrement autonome, c'est-à-dire de l'idée selon laquelle les constructions syntaxiques ne sont en aucun cas influencées par la sémantique. » (Kasper, 2008 : 23)

Plus qu'un parallélisme, le *linking* dénote un lien de causalité - sous forme de règle « si x alors y » – comme le montre l'exemple linguistique suivant utilisé par Kasper :

« Il semble que le « meurtrier », [...] le « lecteur » et le « mangeur » apparaissent toujours en position sujet/au cas nominatif/à la position [Spec, VP] selon la théorie précise du

Linking dans laquelle on se situe. [...] Si un verbe-eur (par exemple un « mangeur ») est présent, il sera en position sujet/au cas nominatif/à la position [Spec, VP]. Les verbe-eur ont un point commun qui justifie que nous leur accordions un statut autonome dans la théorie du Linking : ils peuvent être appelés « agents », et « agent » est ainsi défini - de façon provisoire - comme la généralisation faite à partir d'un ensemble d'arguments [dont le référent, ndlr] « fait quelque chose ». » (Kasper, 2008 : 22-23)

1.1.3. D'une conception où les rôles sémantiques sont le fondement de la théorie du Linking...

1.1.3.1. Les rôles sémantiques impliquent la théorie du Linking

Ainsi, pour (Kasper, 2008), parler de « rôle sémantique » implique - du moins du point de vue de leur genèse - que l'on soit d'accord avec la théorie du Linking.

1.1.3.2. Les défis à la théorie du Linking

Toutefois, deux problèmes principaux se posent dès lors que l'on s'intéresse à l'interface entre syntaxe et sémantique.

Le premier est la possibilité pour un même ensemble de rôles de correspondre à deux structures syntaxiques distinctes. On a alors deux ensembles de correspondances rôle sémantique-argument syntaxique différents. Cela concerne principalement trois phénomènes linguistiques :

– L'alternance, définie comme la propriété qu'ont certains verbes de présenter au moins deux structures argumentales différentes, tout en gardant un noyau sémantique commun aux deux structures. Ce noyau commun est formé de l'ensemble des rôles sélectionnés par le verbe. Autrement dit, le rôle sémantique d'un ou plusieurs participant(s) relatif(s) à **une lexie verbale** est réalisé dans des syntagmes occupant des fonctions syntaxiques différentes en fonction des phrases, comme dans l'exemple suivant d'après (Amsili *et al.*, 2012) :

(2a) *Une imprimante imprime un fichier.*

(2b) *Un utilisateur imprime un fichier avec une imprimante.*

Le rôle d'instrument est réalisé en (2a) par le syntagme nominal occupant la fonction sujet « une imprimante » et en (2b) par le syntagme prépositionnel « avec une imprimante » si on considère ce dernier comme objet indirect - et non pas comme circonstant.

– La passivation, qui peut être vue comme un type particulier d'alternance : « Dans les emplois de la voix active et passive la relation qu'entretiennent les participants avec leurs prédicats verbaux n'est pas modifiée. Leurs fonctions syntaxiques changent mais pas leurs rôles sémantiques » (Hadouche, 2011). Ces deux premiers phénomènes pointent **l'absence de rapport bi-univoque** entre la fonction grammaticale et le jeu de rôles sémantiques.

– Les verbes converses et notamment les verbes psychologiques. (Amsili *et al.*, 2012) citent à ce titre **les lexies** « aimer » et « plaire » (dans « x aime y » et « y plaît à x ») pour lesquelles il y a bien deux « linkings différents pour apparemment les mêmes rôles ».

La seconde difficulté qui apparaît lorsque l'on se penche sur le Linking concerne l'alternance locative. Ce type d'alternance précis dépend de verbes impliquant dans ses arguments internes² les notions de Lieu et de Thème. C'est le cas du verbe « charger » dans :

² "Interne" caractérise le type de syntagme en jeu dans cette alternance. Est interne tout actant (vs. circonstant) ne remplissant pas la fonction de sujet.

(3a) *Marie charge des oranges dans le camion.*

(3b) *Marie charge le camion d'oranges.*

Comme dans l'alternance en (2), il y a deux fonctions syntaxiques distinctes pour une lexie verbale. Mais il y a en plus ici une différence de sens implicite entre la variante locative (3a) et la variante en « de » (3b). En effet en (3b), l'emploi de « charge » implique qu'à la fin de l'action, le camion soit rempli d'oranges alors qu'en (3a), le camion peut ne pas en être entièrement rempli. La question est de savoir si l'argument qui change de fonction, à savoir celui dont la tête est « camion », change aussi de rôle sachant que :

– si l'on postule des rôles différents et donc des linkings distincts (par exemple « locatif » pour « dans le camion » en (3a) et patient pour « le camion » en (3b)), on perd la notion de synonymie

– si l'on postule des rôles identiques, on assimile l'alternance locative à une simple alternance, on perd l'idée d'une différence de sens entre (3a) et (3b) et on retrouve le premier problème évoqué, à savoir la pluralité de linkings pour un même ensemble de rôles.

1.1.4. ...A l'autonomie des rôles sémantiques vis à vis de la théorie du Linking

Les rôles sémantiques, tout en étant une notion opératoire pour certaines descriptions et tâches en TAL, ne suffisent pas à *eux seuls*³ à résoudre les questions théoriques que nous venons de décrire.

Par la suite et selon les auteurs, les rôles sémantiques ont en effet été traités de façon plus ou moins indépendante des propriétés morpho-syntaxiques des arguments. En témoignent les théories qui traitent des rôles sémantiques tout en relativisant voire en niant la théorie du Linking (cf. section 2.1.1.1.). Ceci explique la polysémie du terme « rôles sémantiques » qui peut désigner aussi bien des catégories à l'interface de la sémantique et de la syntaxe que de pures étiquettes sémantiques dissociées de toute considération syntaxique.

1.2. Théories sur les rôles sémantiques

1.2.1. La Grammaire des cas de (Fillmore, 1968)

Les cas correspondent aux rôles sémantiques.

Dans la Grammaire des cas, l'ensemble des cas sélectionnés par un prédicat – verbal, adjectival ou nominal – est appelé **cadre casuel**. Aucun cas ne peut apparaître plus d'une fois dans une proposition à un prédicat. De plus, certains cas sont obligatoires.

Ce sont les cas qui déterminent la présence ou non de certaines fonctions grammaticales. Fillmore avance notamment l'idée d'une règle universelle de sélection du sujet avec une hiérarchie dans l'ordre d'application, représentable sous la forme :

Agent < Instrument < Objectif (≈ Thème)

³ C'est grâce à l'introduction d'autres notions en plus des rôles sémantiques (frame et perspective) que Fillmore a pu donner une explication aux verbes converses (cf. 1.2.2.2.b.)

Ainsi, si un prédicat a dans son cadre casuel un Agent, ce sera ce constituant qui assumera la fonction grammaticale sujet.

1.2.2. La Sémantique des Frames de (Fillmore, 1976) : les rôles sémantiques comme concepts constitutifs de frames

La sémantique des frames inscrit les rôles sémantiques dans une notion plus large : les frames.

1.2.2.1. Les frames, une notion abstraite issue de l'Intelligence Artificielle et de la linguistique - exemple avec le frame anglais « commercial transaction »

1.2.2.1.a. Le frame comme « système de concepts » (Petrucci, 1996)

Considérons la définition suivante des frames :

« Tout **système** de concepts reliés entre eux tel que la compréhension d'un concept implique la compréhension du système entier. » (Petrucci, 1996 : 1)

« L'ensemble des frames forme une structure ordonnée » où des frames s'insèrent dans d'autres frames de portée plus générale. « Le système de dépendance entre les frames forme une **structure** hiérarchique complexe » (Hamm, 2009)

Soit le frame « commercial transaction » d'après l'analyse de (Petrucci, 1996). Il s'insère dans un frame plus général exprimé de façon prototypique par le verbe ditransitif « give ». Les **éléments du frame** « commercial transaction » sont les concepts d'acheteur, de vendeur, de bien et d'argent qui, ensemble, décrivent une **situation** ou « **scène** ». Ils correspondent *grosso modo* aux rôles sémantiques et aux cas de (Fillmore, 1968).

1.2.2.1.b. Le frame comme « trame de fond et [...] motif des catégories représentées par les mots » (Petrucci, 1996)

Les frames sont sémantiquement reliés à des unités lexicales de catégories morpho-syntaxique diverses. Ces unités occupent soit la fonction de **déclencheur** (en anglais « trigger »), soit celle de remplisseur de rôle (en anglais « role filler »). (Petrucci, 1996) dit à propos des déclencheurs :

« Les mots, ou le matériau linguistique, **évoque** le frame (dans l'esprit du locuteur/de l'interlocuteur) ; l'interpréteur invoque le frame à partir de l'énoncé dans lequel le mot apparaît. » (Petrucci, 1996 : 1)

Pour « commercial transaction », les différents déclencheurs verbaux possibles sont « buy », « sell », « pay », « spend », « cost », et « charge ».

« Connaître le sens de tous ces verbes implique de savoir de quoi est composée une transaction commerciale et connaître le sens de chacun d'entre eux implique de connaître le sens des autres. La connaissance et l'expérience structurées dans le frame « commercial transaction » fournissent la **trame et le motif** des catégories représentées par les mots associés au frame⁴. » (Petrucci, 1996 : 1)

⁴ D'où la définition de (Petrucci, 1996) de la sémantique des Frames comme projet de recherche dont « l'un des objectifs est de découvrir le motif qui pousse telle communauté à créer une catégorie - médiatisée par un mot - et à intégrer ce motif dans la description-même du sens du mot »

Les verbes, et plus généralement les mots, représentent des « catégories d'**expérience** » (Petrucci, 1996) structurées par un ensemble de frames :

1.2.2.2.c. Le frame comme « description prototypique de scène » (Hamm, 2009)

Le frame « commercial transaction » décrit de façon **prototypique** une scène dans laquelle on « échange[e] de l'argent contre de la marchandise » (Hadouche, 2009).

Outre les verbes, le frame peut aussi être évoqué par des déclencheurs de type nominal : « tip », « ransom », « allowance », « refund », « honorarium », « bounty », « tuition », « retainer », « bonus », « rent », « fare », « child support », « bus money », « salary », « reward » et « alimony ».

« Faire référence à une somme d'argent par le biais d'un de ces noms implique de faire appel à des informations riches et détaillées appartenant à un contexte plus large où le transfert d'argent n'est qu'une toute petite partie. Par exemple, utiliser le mot "alimony" implique une situation telle que (1) deux personnes qui ont été mariées sont maintenant divorcées (2) un accord a été passé statuant sur la versation d'argent d'un conjoint à l'autre à intervalle régulier, le plus souvent chaque mois. [...] Tous les mots sont potentiellement porteurs de mini-scènes. » (Petrucci, 1996 : 2)

En linguistique, cela renvoie notamment à la notion d'**implicite** (linguistique et pragmatique).

(Hamm, 2009) formule deux critiques à l'encontre de la notion de prototype.

Il note tout d'abord que Fillmore n'utilise pas cette notion pour décrire les mots mais les frames ou les scènes. Or, les unités lexicales étant nécessairement associées à un/plusieurs frame(s), on peut se demander s'il y a un sens à parler de prototype s'agissant notamment de la catégorie des adjectifs : quel serait le prototype de « petit » par exemple ?

La seconde critique vise la notion d'**appartenance à une catégorie** qui peut être difficile à définir face à certains cas. Il cite l'exemple du pingouin qui n'est pas un oiseau prototypique mais que l'esprit humain n'a pas de mal à rattacher à cette catégorie.

1.2.2.2. Les frames dans la langue : interaction lexicale - syntaxe - sémantique

1.2.2.2.a. Un frame : plusieurs structures syntaxiques possibles : interface lexicale - syntaxe

(Petrucci, 1996) ajoute qu'une description complète des verbes déclencheurs de frames « doit aussi comprendre des informations relatives à leurs **propriétés grammaticales et syntaxiques** (structures syntaxiques dans lesquels il apparaissent). »

Poursuivons avec le frame « commercial transaction » présenté dans la section précédente. (Hamm, 2009) synthétise dans un tableau que nous reproduisons ci-après les propriétés syntaxiques des éléments de frames (colonne), en fonction des déclencheurs verbaux (ligne) - à la voix active - :

Déclencheur verbal\élément de frame	BUYER	GOODS	SELLER	MONEY	PLACE
« Buy »	sujet	objet	« from »	« for »	« at »
« Sell »	« to »				
« Cost »	objet indirect	sujet		objet	« at »
« Spend »	sujet	« on »		objet	« at »

Tableau 2 Caractéristiques syntaxiques et morphosyntaxiques des principaux déclencheurs verbaux du frame « commercial transaction » en anglais

Les propriétés sont, selon la réalisation syntaxique, soient des fonctions grammaticales soient des prépositions introduisant des objets indirects. Le tableau 2 montre que la réalisation syntaxique des rôles sémantiques dépend du verbe déclencheur.

(Hamm, 2009) précise que les verbes apparentés sémantiquement à « buy » tels que « sell » ont la « même structure de sens mais des comportements syntaxiques distincts » puisque le remplisseur du rôle Buyer passe de la fonction sujet avec « buy » à la fonction objet indirect introduite par « to » avec « sell ».

1.2.2.2.b. Un frame : plusieurs mises en perspective sémantiques : interface sémantique - syntaxe

Selon (Hadouche, 2011), les déclencheurs « non seulement sélectionnent des concepts individuels, mais indiquent aussi une **perspective** selon laquelle le frame est évoqué. » Ce que Hadouche désigne sous le nom de perspective, Petruck le nomme « aspect » :

« Les verbes d'un même frame « indexe[nt] ou évoque[nt] différents **aspects** du frame. » (Petruck, 1996)

En ajoutant aux rôles sémantiques les notions de frame et de perspective, Fillmore nous permet de dépasser la question des verbes converses.

(4a) *Julie achète à Rachelle un collier.*

(4b) *Rachelle vend un collier à Julie.*

Entre (4a) et (4b), il y a une différence de linking mais les étiquettes sémantiques des arguments exprimés restent les mêmes. La différence de linking s'explique par une différence de perspective. (Petruck, 1996) précise que « le verbe « buy » fait ressortir l'acheteur et le bien concerné, reléguant au second plan le vendeur et le prix » par opposition à « sell » qui « met en avant le vendeur et le bien au détriment de l'acheteur et du prix ». La synonymie entre (4a) et (4b) est quant à elle justifiée par la référence dans les deux phrases à la même scène, c'est-à-dire au même frame.

Concernant le rapport entre la sémantique et la syntaxe, (Petruck, 1996) donne l'exemple suivant :

(5) *Carla bought the computer from Sally for \$100.*

les éléments de frame relégués au second plan « from Sally » et « for \$100 » sont « syntaxiquement optionnels et remplissent en surface le rôle d'objet oblique. En

témoignent les différentes prépositions qui les introduisent [...]. De telles informations sur le rapport syntaxe-sémantique ne font pas partie de la description des verbes d'un frame. Elles sont toutefois déductibles de la description détaillée des éléments des frames. » (Petrucci, 1996).

(Hamm, 2009) formule clairement le lien entre différenciation des structures syntaxiques et différences sémantiques de mise en perspective. Selon lui, les différences de *perspective* (et non de rôles en tant que tels) sont à l'origine des différences de comportement syntaxique entre les verbes.

1.2.3. La Théorie des Principes et Paramètres (Chomsky, 1981, 1986) : les rôles sémantiques comme éléments constitutifs de la grille thématique d'un prédicat

Après avoir insisté sur la séparation entre syntaxe et sémantique, Chomsky montre en 1981 un infléchissement en avançant le principe de Projection. (Levin *et al.*, 1988) définissent ce principe par l'idée qu'« un grand nombre de propriétés syntaxiques d'une proposition sont déterminées par les propriétés sémantiques du prédicateur de la proposition. »

Dans le cadre de la théorie des Principes et Paramètres (notée « PPT ») de Chomsky (1986), il existe un premier « niveau de représentation syntaxique » (Ferret, 2011) des phrases appelé *deep-structure*. Ce niveau « encode les relations prédicat-argument, les **rôles thématiques** de la phrase et [...] fournit un compte rendu de la structure syntagmatique » (Ferret, 2011). Les « rôles thématiques », également appelés « théta-rôles » au sein de PPT, correspondent aux rôles sémantiques.

Les relations prédicats argument ne se réduisent pas aux rôles sémantiques, ceux-ci étant un sous-ensemble des relations.

La *deep-structure* comprend le module de la **théta-théorie**, centré autour de la notion de **grille thématique** (Stowell, 1981). La grille thématique d'un verbe est la « spécification de la structure argumentale des prédicats verbaux et le ou les *rôles que le prédicat assigne à ses arguments* » (Ferret, 2011) avec « argument » signifiant ici plus précisément « actant » ou « participant sémantique occupant une fonction essentielle dans la phrase », les circonstants ne portant dans cette théorie pas de rôle sémantique.

Un exemple de grille thématique est donné par (Levin *et al.*, 1988) pour le prédicat verbal « put » :

(6) *PUT* : <agent, thème, lieu>

Les rôles sont donnés dans le même ordre que l'ordre syntagmatique prototypique des arguments dans une phrase à la voix active. Leur ordre dans une phrase peut donc ne pas correspondre à la grille thématique.

La notion de grille thématique élargit le cadre de sous-catégorisation à l'argument qui en structure de surface (ou s-structure) remplit la fonction sujet. Elle permet de distinguer les **sujets thématiques**, qui ont un rôle sémantique et dont la présence est impliquée par le sens du verbe (ex. : « le concierge court » avec le verbe « courir » qui a comme grille thématique : courir <agent>), des sujets non thématiques comme les sujets explétifs

généralement associés à des verbes météorologiques (« il pleut » avec « pleuvoir » dont la grille thématique est vide : pleuvoir <>).

Le prédicat assigne un rôle thématique en respectant le **théta-critère** :

- Chaque théta-rôle doit être assigné à un et un seul argument
- Chaque argument doit recevoir un et un seul théta-rôle

Il existe trois types d'argument :

Type d'argument	contraintes - liées aux modules de la théorie X-barre et de la théorie du Gouvernement et du Liage (GB) - dans l'assignation d'un rôle thématique	Différentes positions structurales dans l'arbre syntaxique au niveau deep-structure
argument externe	assignation compositionnelle	
argument interne direct	sous gouvernement : - A gouverne B si et seulement si A c-commande B et B c-commande A (et A ne domine pas B) - A est un gouverneur (= une tête lexicale – 1ère version). A c-commande B si et seulement si le premier noeud branchant qui domine A domine également B (et A ne domine pas B)	
argument interne indirect	assignation indirecte : par l'intermédiaire de la préposition	

Tableau 3 Trois configurations syntaxiques possibles d'assignation des théta-rôles par le prédicat

Le linking n'est pas entre les rôles sémantiques et la fonction syntaxique mais entre les rôles sémantiques et la **position structurale** des arguments.

Dans leur représentation, (Williams, 1981) et (Marrantz, 1984) fusionnent les aspects syntaxique et sémantique. Dans la représentation de la structure prédicat-argument, les

rôles sémantiques sont associés à des symboles représentant la position structurale des argument. Sur le même exemple qu'en (6), on aura :

(7) *PUT* : agent <thème, lieu>

où le rôle en dehors des crochets correspond au syntagme en position [Spec, VP]⁵ et le rôle souligné se rapporte au syntagme en position [V⁰, Complément].

En outre, (Hirschbühler *et al.*, 1992) notent que « la grille argumentale d'un item lexical est généralement conservée à travers la dérivation morphologique ». Soit le nom déverbal « réparation » :

(8a) *Le charpentier répare le toit.*

(8b) *La réparation du toit par le charpentier.*

Le nom « réparation » en (8b) est un dérivé morphologique du verbe « réparer » présent en (8a) avec sens apparenté.

« Les participants réalisés en actants par une lexie verbale [...] sont réalisés en actants ayant les mêmes rôles sémantiques que ceux d'un dérivé [...]. Cette lexie et son dérivé ont des **sens apparentés** » (Hadouche, 2011 : 49)

« **L'ordre** respectif de la tête et des compléments [internes, ndlr] y est le même quelle que soit la catégorie de la tête : en français et en anglais, la tête du syntagme précède les compléments, que cette tête soit un verbe, un nom, un adjectif ou un adverbe. » (Hirschbühler *et al.*, 1992 : 143)

C'est ce constat qui a amené Chomsky à proposer la théorie X-barre, théorie selon laquelle tous les syntagmes sont formés selon la même structure de base (cf. schéma de droite, figure 2).

1.2.4. (Levin *et al.*, 1988) : les rôles sémantiques comme « relation de sens » (Hadouche, 2011) entre le prédicat et ses arguments

(Levin *et al.*, 1988) ont développé l'idée d'une séparation nette entre deux niveaux de représentation de la relation entre le prédicateur et ses arguments. Ils distinguent les variables logiques vides de sens, des rôles sémantiques qui sont porteurs de sens.

1.2.4.1. Niveau de représentation lexico-syntaxique ou Structure Prédicat-Argument (SPA)

La relation prédicat-argument, au centre de toutes les théories sur le lexique, rend compte de « la propriété essentielle des « prédicateurs » qui est [...] de posséder des arguments » (Levin *et al.*, 1988).

Le niveau de représentation lexico-syntaxique « ne comporte pas directement d'étiquettes de rôles thématiques ». Il est une « représentation qui précise comment en syntaxe, les [argument] qui correspondent à [chaque] variabl[e] se voient attribuer [un] rôl[e] thématiqu[e]. » (cf. tableau 3)

⁵ [Spec, VP] correspondant au système de notation de la figure 1 peut aussi s'écrire [Spec, V''] en accord avec le système de notation du tableau 3.

Au lieu d'utiliser des grilles thématiques du type de celle décrite en (6) et en (7), (Levin *et al.*, 1988) propose de représenter la structure prédicat-argument avec une structure argumentale inspirée de (Marrantz, 1984) et plus détaillée. Pour le prédicat « put », on aura :

(9) *PUT* : $x <_y, P \text{ loc } y >$

Avec « P loc » = préposition locative

« Toutes les distinctions faites entre les arguments d'un verbe le sont en termes de mode d'assignation des rôles thématiques [...] » (Levin *et al.*, 1988)

1.2.4.2. Niveau de représentation lexico-sémantique.

Le niveau qui traite des rôles sémantiques concerne plus spécifiquement la sémantique lexicale. Les rôles y sont définis comme « relation de sens » (Hadouche, 2011a) entre un actant et un prédicat.

2. Les rôles sémantiques : point de vue descriptif

2.1. *Tertium comparationi* aux différentes théories

2.1.1. Le Linking

2.1.1.1. Des théories plus ou moins en accord avec le Linking

Nous allons aborder deux des quatre théories présentées en 1.2. sous l'angle du Linking : celle de Levin et Rappaport dans *Lexique 7 : Lexique et syntaxe en grammaire générative* et celle d'Amsili et Candito dans « Analyse sémantique automatique : semantic role labeling ».

Pour (Levin *et al.*, 1988), il existe bien une « relation » entre la sémantique et la syntaxe, mais cette relation n'est pas « univoque » :

« Les processus syntaxiques ne sont pas influencés par les étiquettes de rôles thématiques [...] Il n'existe pas de relation univoque entre les rôles sémantiques et les modes d'assignation des rôles thématiques » (Levin *et al.*, 1988 : 17)

D'où leur distinction entre le mode d'assignation des rôles au niveau « lexical syntaxique » et les étiquettes des rôles - porteuses de sens - au niveau « lexical sémantique » (cf. 1.2.4.1. et 1.2.4.2.).

(Amsili *et al.*, 2012) adhèrent quant à eux à l'idée d'un Linking sinon systématique du moins régulier. S'inspirant de la définition que (Kasper, 2008) donne des rôles sémantiques, (Amsili *et al.*, 2012) définissent ces derniers en utilisant le terme de « régularités » :

« Généralisation sur les arguments d'un prédicat en vue de capturer des régularités dans l'interprétation (sémantique) des représentations syntaxiques. » (Amsili *et al.*, 2012 : 4)

2.1.1.2. Les différentes théories du Linking

Parmi les tenants de la théorie du Linking, les linguistes se répartissent en trois catégories en fonction du paramètre syntaxique précis qu'ils associent aux rôles sémantiques. Les paramètres peuvent être :

- la position structurale du syntagme avec en général une représentation des structures sous forme d'arbres. C'est le cas de la théorie des Principes et Paramètres,
- la fonction grammaticale avec les grammaires lexicales fonctionnelles (LFG en anglais).
- les cas. D'après (Kasper, 2008) et (Amsili *et al.*, 2012), le premier grammairien à aborder une notion proche des rôles sémantiques sous cet angle est Panini entre le Vème et le IVème siècle avant J.C.. Il propose en effet des règles de correspondance entre les cas et les rôles « Karaka », catégorie caractérisant un certain type d'argument et propre à certains verbes.

(Kasper, 2008) note que seuls les paramètres de la position structurale et des cas sont compatibles avec la revendication d'une théorie universelle, c'est-à-dire valable pour toutes les langues.

2.1.2. Mode de représentation des rôles sémantiques

Un autre tertium comparationis possible entre les théories traitant des rôles sémantiques selon (Kasper, 2008) est la façon dont les auteurs proposent de représenter cette notion. Nous traiterons le mode de représentation sous deux angles : les primitifs et la possibilité ou non d'établir une liste de rôles⁶.

2.1.2.1. Les rôles sémantiques sont-ils des primitifs ?

Pour (Jackendoff, 1972) et (Levin *et al.*, 1988) les rôles ne sont pas des primitifs sémantiques eux-mêmes mais ils sont *composés* de primitifs.

Jackendoff « déduit les rôles thématiques à partir des positions dans la structure conceptuelle des items lexicaux, lesquels consistent en des primitifs (comme CAUSER et CHANGER) » (Kasper, 2008).

Pour illustrer cette idée, (Levin *et al.*, 1988) prennent l'exemple du verbe anglais « to break » qui, selon la construction syntaxique dans laquelle il se trouve, se décompose sémantiquement de deux façons différentes :

(10a) *Sarah broke the cup.*

(10b) *The cup broke.*

(10a) et (10b) ont en commun le trait [y devenir CASSE]. « devenir » est un primitif potentiel et CASSE une « constante représentant l'élément de signification qui distingue l'état d'être cassé d'autres états ». Dans la variante causative (10a), le verbe « broke » contient un trait supplémentaire par rapport à la variante non causative (10b) : CAUSER. Ce trait « s'applique » au premier trait, ce qui peut être représenté pour (10a) de la façon suivante :

[x cause [y devenir CASSE]]

⁶ Il est à noter que certains linguistes (comme Bresnan) ont successivement (1982, 1989) adopté différents points de vue concernant ce paramètre.

par opposition à la représentation de (10b) :

[y devenir CASSE]

« Si l'on considère que les structures analysées peuvent capturer les traits de sens pertinents, alors le sens doit être décomposé de la « bonne façon ». Mais il n'existe pas de critère exact de décomposition du sens. » (Kasper, 2008 : 34)

A l'inverse, les approches que (Kasper, 2008) appelle « holistiques » (à savoir les théories chomskyennes, (Gruber, 1965), (Fillmore, 1968) et la Grammaire Lexicale Fonctionnelle à ses débuts avec Bresnan), considèrent que « les concepts dénotés par les rôles thématiques forment un ensemble universel de conceptualisations de relations prédicat-argument » (Kasper, 2008). En général, cette position va de paire avec l'idée que l'on peut établir une liste restreinte de rôles.

2.1.2.2. Les rôles sémantiques peuvent-ils être représentés sous forme de liste?

Les premiers linguistes à s'être intéressés aux rôles sémantiques dont (Gruber, 1965), (Fillmore, 1968), (Chomsky, 1981) et (Bresnan, 1982) ont d'abord suggéré une représentation des rôles sous forme de liste. Cette liste était dans un premier temps formée d'**unités discrètes** mises **sur le même plan**.

Puis, pour ajouter des contraintes à certains phénomènes comme la passivation ou les constructions réflexives, certains linguistes parmi lesquels (Jackendoff, 1972 et ultérieur), (Fillmore, 1977), (Bresnan, 1989) et (Primus, 2002) ont proposé de représenter les rôles sous la forme d'un **ensemble hiérarchisé**. Il n'y a toutefois pas de consensus entre ces linguistes sur les rôles impliqués et leur ordonnancement.

Notons que Fillmore est passé d'une représentation sous forme de liste restreinte de rôles à vocation universelle (à savoir l'Agent, l'Instrument et l'Objectif) à une liste étendue de rôles plus au moins spécifiques en fonction de la portée des frames auxquels ils se rapportent. (Ruppenhofer, 2010) tente de donner une explication à cela :

« Au fur et à mesure que la description des unités lexicales a progressé, il est apparu que les theta-rôles et les rôles assimilés à des cas ne couvraient qu'une partie des rôles existant. » (Ruppenhofer, 2010 : 85)

Puis il relève les conséquences et les implications de ce choix :

« Réduire l'ensemble de rôles utilisés pour les éléments de frame à un jeu infiniment plus petit (et certes plus facilement traitable par une machine) de théta-rôles n'est pas sans poser problème. L'un d'eux, bien connu dans la littérature, porte sur le choix des étiquettes. » (Ruppenhofer, 2010 : 85)

A l'opposé des auteurs tentant de construire des listes, (Levin *et al.*, 1988) contestent la possibilité de dresser une quelconque liste de rôles sémantiques :

« Le savoir lexical n'est pas assimilable à des listes mais [...] implique des processus de généralisation sur les classes de prédicateurs et sur les relations existant entre les prédicateurs et leurs arguments. » (Levin *et al.*, 1988 : 13)

Levin explicite son idée dans *English Verb Classes and Alternations: A Preliminary Investigation* :

« La complexité des schémas des comportements syntaxiques exhibée par les verbes dans les phénomènes d'alternance ne saurait s'expliquer par une représentation sémantique lexicale consistant en une liste de rôles sémantiques. » (Levin, 1993 : 16)

Pour étayer cette thèse, (Levin *et al.*, 1988) prennent l'exemple de l'alternance locative, définie en 1.1.3.2. comme mettant en jeu deux fonctions syntaxiques distinctes pour une lexie verbale, laquelle implique minimalement les notions de thème et de lieu.

Reprenons les exemples donnés en (3a) et (3b) pour mieux comprendre l'argumentation des auteurs :

(3a) *Marie charge des oranges dans le camion.*

(3b) *Marie charge le camion d'oranges.*

Les auteurs assignent trois conditions à toute description de l'alternance locative :

- la formulation et l'explication de la relation « quasi paraphrastique » entre les deux variantes.
- la capacité à formuler des règles de linking pour chaque argument de la phrase du type « si l'argument x occupe telle position structurale, alors lui assigner tel rôle sémantique ». Trois exemples de règle sont donnés par les auteurs :

**Associer le rôle d'agent à la variable argument externe de la SPA.
Associer le rôle de thème à la variable argument direct de la SPA.
Associer chaque rôle thématique restant à une variable argument indirect de la SPA qui est associée à une préposition appropriée.**

Figure 3 Exemple de règles d'association ou de « Linking » entre rôles sémantiques et position structurale

- l'explication de « l'interprétation « affectée » de l'argument « but » quand celui-ci est argument direct » (pour nous « le camion » en (3b)) (Levin *et al.*, 1988)

Or, comme le font remarquer les auteurs, quel que soit le nombre de listes de rôles sémantiques que l'on pose pour l'analyse de l'alternance, on ne remplit jamais simultanément l'ensemble de ces trois conditions :

Nombre de liste distinctes/conditions	Relation quasi paraphrastique	Règles de linking	Inteprération affectée
Une	oui	non	non
Deux	non	oui	oui

Tableau 4 Matrice de validation des conditions de description de l'alternance locative en fonction du nombre de listes de rôles

« Chacune de ces analyses prend en compte différentes facettes du phénomène d'alternance mais aucune ne rend compte, de manière exhaustive et satisfaisante, de l'ensemble des faits impliqués dans l'alternance. » (Levin *et al.*, 1988 : 24)

2.2. Différents aspects des rôles sémantiques

2.2.1. Ambiguïté dans le rôle de certains participants

L'ambiguïté dans le rôle de certains participants est liée à l'ambiguïté lexicale du noyau syntaxique des participants en question. (Hadouche, 2011) donne l'exemple suivant :

(11) *Ce guide nous a renseignés.*

Il y a ambiguïté entre les rôles Agent et Instrument. Ceci est dû à l'ambiguïté lexicale du noyau nominal « guide ». Il y a donc deux types d'ambiguïtés en (11).

Tant que l'on se situe dans une linguistique de type structurale où l'unité maximale considérée est la phrase, il nous est impossible de lever ces ambiguïtés. Il n'est pas absurde de penser qu'une linguistique du texte pourrait nous fournir des indices nous permettant de pencher vers telle ou telle interprétation du sens précis de « guide » et donc tel rôle sémantique. Par exemple, si l'on voulait traiter cet exemple comme un exemple non construit, on pourrait essayer de rechercher l'antécédant du déterminant de « guide » « ce », sous réserve d'une interprétation anaphorique de ce mot référentiellement non autonome.

2.2.2. Verbes au sens voisin

(Hadouche, 2011) note que « des lexies [verbales] différentes ayant des sens voisins peuvent également partager des actants ayant les mêmes rôles sémantiques. » Par exemple :

(12a) *Un pirate s'attaque aux Macs.*

(12b) *Les vers infectent les fichiers.*

(12c) *Les virus s'introduisent dans l'ordinateur⁷.*

Dans ces trois phrases, nous avons des termes prédicatifs qui ont des sens voisins représentables par la structure actancielle « Assaillant ~ Destination ». » ou si l'on veut la grille thématique x : <Assaillant, Destination>

2.2.3. Alternances syntaxiques impliquant les rôles sémantiques : alternances de transitivité

Certaines alternances (cf. définition en 1.1.3.2.) impliquent directement les arguments auxquels sont associés des rôles sémantiques

En plus de l'alternance locative abordée en 1.1.3.2. et 2.1.2.2., citons l'alternance causative/inchoative et l'alternance transitive intermédiaire. Les deux sont des alternances de transitivité.

⁷ La couche d'annotation en rôles sémantiques a été supprimée, ce qui nous intéresse ici étant le phénomène abstrait des verbes au sens voisins et non pas sa représentation sur corpus.

2.2.3.1. Alternance causatif/inchoatif

L'alternance causative/inchoatif peut être illustrée par le verbe « break » dans l'exemple donné en (10) :

(10a) *Sarah broke the cup.*

(10b) *The cup broke.*

Dans la variante inchoative (10b), le verbe est intransitif ; dans la variante causative (10a), le verbe est transitif et comprend le trait sémantique [être la cause de/provoquer].

2.2.3.2. Alternance transitive intermédiaire

L'alternance transitive intermédiaire reprend les mêmes règles syntaxiques que l'alternance causative/inchoatif. « To break » en fait d'ailleurs partie. Toutefois, elle s'en distingue par :

- le fait qu'elle exprime un événement localisable dans le temps
- le fait que dans les deux variantes, le rôle sémantique d'Agent soit
 - présent
 - remplit par le sujet

(13a) *Margaret cut the bread.*

(13b) *Margaret cut at the bread.*

En (13a), l'action implique qu'une tranche ait été coupée alors que (13b) n'implique pas nécessairement que l'action ait été menée jusqu'à son terme. (13b) peut signifier que Margaret n'a pas réussi à couper le pain. Cette nuance sémantique dépasse toutefois l'analyse en rôles sémantiques aussi parfois appelée « analyse sémantique de surface. »

2.2.4. Variation syntaxique d'un participant : actant - circonstant

Nous avons choisi le nom de « variation » pour distinguer ce phénomène de l'alternance. Contrairement à celle-ci, la variation met en jeu deux verbes différents. C'est justement cette distinction dans les verbes qui donne lieu pour un même argument à deux réalisations syntaxiques différentes.

« Une lexie peut réaliser un actant dans une phrase et un circonstant dans une autre. Par exemple, une unité lexicale qui indique le lieu peut être un actant ou un circonstant selon la lexie verbale avec laquelle elle est employée. Dans la phrase :

(14a) *Une personne consulte de l'information sur le Web.*

Le participant « sur le Web » joue le rôle d'un circonstant de rôle sémantique Lieu⁸ de la lexie verbale « consulter ». Cependant, dans la phrase :

⁸ Contrairement à Chomsky, (Hadouche, 2011) accorde aux circonstants le statut d'argument et donc la propriété d'avoir un rôle sémantique.

(14b) *Une personne navigue sur le Web.*

le participant « sur le Web » réalise un actant ayant comme rôle sémantique Lieu de la lexie verbale « naviguer ». » (Hadouche, 2011)

3. Ressources lexicographiques intégrant la notion de rôle sémantique

Parmi les différentes ressources linguistiques existantes figurent les lexiques (ou « ressources lexicographiques ») et les corpus annotés. Dans une perspective d'automatisation des traitements linguistiques, les lexiques servent plutôt à fournir une batterie de règles applicables selon les cas rencontrés alors que les corpus annotés servent plutôt de base pour un apprentissage automatique. Je m'intéresserai ici principalement aux lexiques dans la mesure où c'est sur ce type de ressources que j'ai travaillé durant mon stage de master au CEA.

La plupart des ressources lexicographiques intégrant de façon systématique des informations non seulement sur le lexique en général mais aussi sur les rôles sémantiques en particulier sont de **langue anglaise**. Parmi les plus importantes, citons FrameNet et le Verbnet⁹ informatisé dont aucun équivalent français n'existe encore à ce jour.

FrameNet et Verbnet sont assez différents à la fois du point de vue de leur conception générale et du point de vue de la description linguistique.

Sur le plan de la conception, FrameNet et Verbnet reposent sur des théories de domaine et de contenu différents. FrameNet « implémente » la sémantique des Frames de (Fillmore, 1976) ; le Verbnet informatisé s'inspire du système de classification sémantico-syntaxique des verbes de (Levin, 1993). Dans chacune de ces ressources, les rôles sémantiques sont décrits au sein d'autres catégories : les classes de prédicats dans Verbnet et les frames - formant chacun une classe - dans FrameNet. Les classes sont hiérarchisées selon des critères propres et à des degrés divers. Les critères de classification sont sémantiques et syntaxiques pour Verbnet et conceptuels pour FrameNet.

Du point de vue linguistique, les deux ressources associent à chaque classe des unités lexicales. Contrairement à Verbnet qui ne répertorie que les prédicats verbaux, FrameNet prend en compte une diversité de prédicats (les « déclencheurs de frame »), des verbes aux noms en passant par les adjectifs¹⁰. Dans FrameNet, les rôles sémantiques ont une définition plus large à la fois en intension (des rôles sont donnés pour les arguments syntaxiquement optionnels) et en extension que dans Verbnet. En revanche, Verbnet intègre des informations syntaxiques et les relie aux rôles sémantiques pour chaque argument là où la base FrameNet liste des éléments de frame sans leur associer d'informations morpho-syntaxiques. Seul le corpus annoté FrameNet relie les arguments syntaxiques aux rôles sémantiques. Souhaitant ne pas dépendre de corpus annotés, les concepteurs du programme d'annotation du CEA ont logiquement choisi de s'appuyer sur le lexique Verbnet.

⁹ Un autre lexique existe, qui a été construit à partir du corpus annoté en rôles sémantiques PropBank.

¹⁰ Certains frames contiennent même dans leur liste d'unités lexicales des prépositions. Par exemple, « Goal » contient la préposition « to ».

3.1. FrameNet : intégration des rôles sémantiques dans des classes de concepts appelés « frames »

La ressource lexicographique FrameNet a vu le jour au sein de L'International Computer Science Institute à Berkeley en Californie. Accessible en ligne¹¹, FrameNet répertorie l'ensemble des frames via un index trié par ordre alphabétique. Chaque frame se décompose en quatre rubriques :

- définition du frame sous forme de paraphrase contenant les éléments de frame (ou rôles sémantiques) et exemple linguistique contenant une réalisation lexicale pour chaque élément de frame
- liste des **éléments de frame** notée FE, définition et exemple linguistique pour chaque élément
- spécification des éventuelles relations (dont l'héritage, l'inclusion, la perspective) entre le frame et d'autres frames
- liste des unités lexicales associées¹² c'est-à-dire des déclencheurs de frame quand il y a lieu¹³

3.1.1. Mise en relation des rôles sémantiques et de leurs réalisations lexicales

Dans les deux premières rubriques, les éléments de frame et leur réalisation lexicale sont reliés entre eux par un système de surlignage associant une couleur à un élément. Par exemple, dans le frame « Achieving_first » les étiquettes des éléments [Cognizer]¹⁴ et [New_idea] sont respectivement mis en relation avec les réalisations lexicales « Edison's » et « of the light bulb » grâce au surlignage respectivement rouge et bleu. De plus, les mêmes couleurs sont associées aux mêmes éléments entre les rubriques Definition et FE :

¹¹ <https://framenet.icsi.berkeley.edu/fndrupal/index.php?q=frameIndex>

¹² Une unité lexicale se définit comme une paire constituée d'un lexème et d'un sens. Un lexème peut se retrouver dans plusieurs frames en fonction du sens qui participe à l'évocation du frame en question.

¹³ Certains frames dits « frames sémantiques non lexicaux » ont un niveau d'abstraction tel qu'ils ne sont associés à aucune unité lexicale.

¹⁴ Nous utilisons les crochets pour les éléments de frame afin de marquer la différence entre ces derniers qui sont de nature conceptuelle et les unités lexicales qui sont de nature linguistique.

Definition:

A **Cognizer** introduces a **New_idea** into society.
 It is difficult to imagine how different our lives would be without **Edison's INVENTION** of the light bulb.

FEs:

Core:

Cognizer [Cog] The **Cognizer** comes up with or conceptualizes the **New_idea**. It is normally expressed as an External Argument:
Semantic Type: Sentient **Kim Sung** **INVENTED** cinnamon cheez fries in 1988.

Figure 4 Co-indexation dans FrameNet des éléments de frame et de leur(s) réalisation(s) lexicale(s) dans les rubriques respectives Definition et FE et entre ces rubriques

Le noir est utilisé pour marquer les différentes réalisations lexicales du déclencheur de frame. Ainsi pour le frame « Achieving_first », le déclencheur est, dans l'exemple de la définition, le nom « invention » et le verbe fléchi « invented » dans l'exemple du premier élément de frame [Cognizer]. Ces unités lexicales prédicatives sont ensuite reprises et décrites dans la rubrique « Lexical Units ».

3.1.2. Représentation implicite de la hiérarchie des rôles sémantiques

La couleur permet également de relier entre eux les éléments de deux ou plusieurs frames quand les éléments entrent dans une relation d'héritage. L'élément [Cognizer], surligné en rouge comme dans la figure 4, est relié à l'élément [Creator] également surligné en rouge dans le frame « Intentionally_create » :

Intentionally_create

Definition:

The **Creator** creates a new entity, the **Created_entity**, possibly out of **Components**.
The Libyans were **GENERATING** arms for export.

Figure 5 Co-indexation des éléments de frame entretenant une relation d'héritage dans FrameNet grâce au système de surlignage

Nous pouvons déduire le sens de la relation entre [Cognizer] et [Creator] grâce à la relation d'héritage décrite dans la rubrique « frame-frame Relations » de l'un ou l'autre des frames auquel appartiennent ces éléments, à savoir « Achieving_first » et « Intentionally_create » :



Figure 6 Déduction du sens de la relation d’héritage entre les éléments de frame grâce à la rubrique « frame-frame Relations » dans FrameNet

La rubrique « frame-frame Relations » de « Achieving_first » (figure 6) nous informe du fait que ce frame hérite de « Intentionally_create », ce qui nous indique le sens de la relation entre leur élément de frame respectif représentés en rouge : [Cognizer] de « Achieving_first » hérite de [Creator] présent dans « Intentionally_create ». En effet, la relation d’héritage implique que les sous-frames héritent de certaines propriétés sémantiques des éléments de frame de leur « parent ».

Comme la relation d’héritage entre les frames, la relation d’héritage entre les éléments peut s’étendre sur plusieurs niveaux. Ainsi [Creator], « sur-élément » de [Cognizer], est lui-même le « sous-élément » de l’élément « Agent » du frame parent de « Intentionally_create » à savoir « Intentionally_act » :

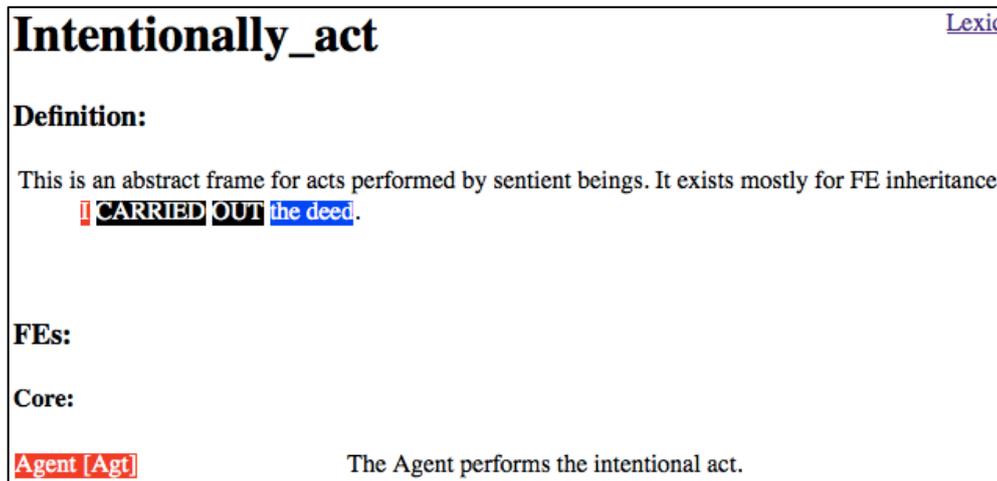


Figure 7 Classe FrameNet dont hérite « Intentionally_create » et contenant l’élément de frame générique « Agent »

Les rôles sémantiques dont on a vu qu’ils s’identifiaient aux éléments de frame sont dans FrameNet en **nombre plus important** que dans la théta-théorie par exemple puisqu’ils décrivent des notions en fonction du sens et de la portée du frame auquel ils sont associés. Un élément de frame est en effet défini sur le site de FrameNet comme « rôle sémantique **spécifique à un frame** et unité de base du frame¹⁵. »

¹⁵ « frame-specific defined semantic role that is the basic unit of a frame ».

3.2. *Verbnet : intégration des rôles sémantiques dans des classes linguistiques sémantico-syntaxiques*

3.2.1. Définition des classes Verbnet

Verbnet (Kipper-Schuler, 2006) est une ressource lexicographique anglaise créée à l'Université du Colorado aux Etats-Unis et disponible en ligne¹⁶. Il est inspiré du système de classification des verbes de (Levin, 1993). Nous le considérerons ici dans sa dernière version (la 3.2.), qui a été conçue pour enrichir le système de Levin.

L'unité indexée dans Verbnet n'est pas le concept comme dans FrameNet mais le verbe. Environ 4000 verbes distincts sont traités. Ils sont répartis dans 274 classes, chacune décomposable en sous-classes selon des considérations d'ordre essentiellement syntaxique. Le fondement de Verbnet reste fidèle à celui de Levin dans la mesure où le critère de définition des classes est le même. « Certaines propriétés syntaxiques, dit (Levin, 1993), sont liées à certaines classes de verbes », et ces classes sont constituées non pas de verbes sémantiquement proches mais de verbes ayant en commun un **ensemble de traits sémantiques syntaxiquement pertinents**¹⁷, c'est-à-dire « serv[ant] minimalement à opposer les verbes qui ocurrent dans les **alternances** relevées¹⁸ ».

Outre le niveau de granularité de description des prédicats, le Verbnet informatisé et la classification de Levin ont ainsi en commun le fait de pouvoir opposer syntaxiquement et sémantiquement les classes de verbes entre elles. Toutefois, l'opposition sur le plan sémantique n'est en pratique pas réalisable de façon objective étant donné que ni la classification de (Levin, 1993) ni le Verbnet informatisé ne mentionne les traits sémantiques définitoires des classes dans leur description. Les oppositions syntaxiques entre les verbes d'une même classe sont représentées dans le Verbnet informatisé par la division en une ou plusieurs sous-classe(s).

Du point de vue des rôles sémantiques, une différence essentielle existe entre la classification de (Levin, 1993) et Verbnet.

¹⁶ <http://verbs.colorado.edu/verb-index/vn/class-h.php>

¹⁷ On peut voir cette définition comme la tentative d'appliquer à la sémantique la méthode structuraliste de la phonologie découverte par Trubetsky et les linguistes du cercle de Prague selon laquelle il existe des traits phonétiques phonologiquement pertinents, c'est-à-dire servant à opposer deux mots du point de vue de leur sens.

¹⁸ Si les verbes « dump » et « load » pourtant proches sémantiquement car désignant mutuellement un mouvement localisé, ne sont pas dans une même classe, c'est qu'ils ne partagent pas de trait sémantique particulier motivant un comportement syntaxique donné.

Il est à noter que dans la classification de Levin, la plupart des alternances considérées concernent exclusivement des propositions simples ne contenant pas de complétive. La dernière version de Verbnet en particulier vise entre autres à pallier cette insuffisance. On peut ainsi résumer l'hypothèse linguistique de Levin sous forme de règle générale : « [propriété du verbe]+ ⇒ [trait sémantique verbal]+ avec ⇒ = implique et + = de 1 à ∞ fois répété » et appliquer cette règle aux corrélations respectives données par (Levin, 1993) entre les deux alternances syntaxiques vues en 2.2.3. et certains traits sémantiques :

- alternance causatif/inchoatif ⇒ [changement d'état]. Ex. : « to break »

- alternance transitive intermédiaire ⇒ [changement d'état][cause]. Ex. : « to break », « to cut »

3.2.2. D'un système de classification niant la possibilité de lister les rôles sémantiques à une description intégrant une liste de rôles

Verbnet est un exemple de ressource qui montre comment on peut passer d'une analyse linguistique fondée sur le refus de lister les rôles sémantiques, à une description des prédicats verbaux intégrant une liste de rôles et qui plus est, dont l'un des **objectifs** pratiques revendiqués est de pouvoir **servir de base à l'annotation en rôles sémantiques**¹⁹.

Alors que la thèse de Levin conduit celle-ci à nier aux rôles sémantiques le statut de liste (cf. 2.1.2.2.), le Verbnet informatisé induit à partir des arguments communs à tous les verbes de chaque classe une liste de rôles sémantiques.

En fonction de la documentation, il y a entre 21 et 30 rôles sémantiques dont : actor, agent, asset, attribute, beneficiary, cause, location qui, avec destination et source appartient à la classe générique « location, destination, source », « utilisée pour les localisations spatiales », experiencer, extent, instrument, material, product, patient, predicate, recipient, stimulus, theme, time et topic.

On distingue trois rubriques au sein de chaque classe. Dans l'ordre de leur présentation dans Verbnet²⁰, il y a :

- « Members » contenant les prédicats verbaux
- « Roles » énumérant les rôles communs aux arguments des verbes membres de la classe
- « Frames » constituée des « frames ». Un frame est une association admise par l'ensemble des verbes de la classe entre un patron (ou « une structure ») syntaxiques et un ensemble de rôles sémantiques syntagmatiquement ordonné²¹. Précisons que les patrons syntaxiques non admis par les verbes de certaines classes (ou « agrammaticaux ») et entrant dans la définition des alternances de ces verbes selon (Levin, 1993) ne sont de manière générale pas mentionnés dans la version informatisée²².

Pour illustrer l'organisation de Verbnet, considérons la classe `put_spatial-9.2` contenant les verbes suivants :

¹⁹ « Cela devra finalement permettre de tester à grande échelle l'utilité des classes syntaxiques pour la performance des parseurs syntaxiques et de l'annotation en rôles sémantiques sur de nouveaux domaines. » <http://verbs.colorado.edu/~mpalmer/projects/verbnet.html>

²⁰ L'ordre de présentation de ces trois rubriques diffère de l'ordre de leur mise à jour selon la méthode préconisée par (Levin, 1993).

²¹ Le caractère sémantico-syntaxique du frame dans Verbnet s'oppose à la définition essentiellement sémantique (ou conceptuel) du frame de FrameNet.

²² Ainsi, pour une alternance formée d'un patron grammatical et d'un autre non grammatical, la classe contiendra systématiquement un seul patron dans la rubrique « Frames », le patron grammatical.

MEMBERS
LAY (FN 1; WN 1, 2; G 1)
PROP (WN 1; G 1)
UNDERLAY (WN 1, 2, 3; G 1)

Figure 8 Rubrique « Membres » de la classe Verbnets put_spatial-9.2

Les rôles sémantiques – mentionnés spécifiquement dans la ressource informatisée - correspondant sont les suivants :

ROLES
• AGENT [+ANIMATE]
• THEME [+CONCRETE]
• DESTINATION [+LOCATION & -REGION]

Figure 9 Rubrique « Roles » de la classe Verbnets put_spatial-9.2

Les indications en orange sur la figure 9 sont des traits sémantiques binaires qui restreignent le champ d'application des rôles. Il en existe 36 différents dans Verbnets. Pour la classe put_spatial-9.2, l'agent est un être animé, le thème une entité concrète et la destination un lieu²³ précis.

Les rôles sont ensuite **donnés dans l'ordre** de la structure syntaxique à laquelle ils sont associés dans la rubrique « Frames » (cf. ligne « Syntax »). Pour put_spatial-9.2²⁴, l'alternance est constituée des deux structures suivantes :

²³ On remarque que l'étiquette « location » est utilisée dans certaines classes comme trait binaire (comme ici avec put_spatial-9.2) et dans d'autres comme rôle sémantique (exemple : search-35.2.)

²⁴ La notion de frame a été ajoutée à la ressource en ligne afin de pouvoir distinguer plus facilement les différentes alternances au sein d'une même classe. Dans la classification de Levin, les alternances internes aux classes sont distingués typographiquement.

FRAMES	
NP V NP PP.DESTINATION	
EXAMPLE	"Cheryl stood the books on the shelf."
SYNTAX	<u>AGENT</u> V <u>THEME</u> {{+LOC}} <u>DESTINATION</u>
SEMANTICS	MOTION(DURING(E), THEME) NOT(PREP(START(E), POSITION(END(E), THEME, POS) CAUSE(AGENT, E)
NP V NP ADVP	
EXAMPLE	"Cheryl stood the books there."
SYNTAX	<u>AGENT</u> V <u>THEME</u> <u>DESTINATION</u> <+ADV_LOC>
SEMANTICS	MOTION(DURING(E), THEME) NOT(PREP(START(E), POSITION(END(E), THEME, POS) CAUSE(AGENT, E)

Figure 10 Différents ordres d'apparition des rôles sémantiques en fonction de l'ordre syntaxique impliqué dans l'alternance propre à un frame dans la classe Verbnnet put_spatial-9.2

Dans la figure 10, il y a deux ordres possibles pour les mêmes rôles : un pour la structure NP V NP PP.Destination et un pour la structure NP V NP ADVP. Dans d'autres frames d'autres classes, la distinction des structures en rôles peut se faire sur leur ordre et/ou sur leur nombre (certaines étiquettes apparaîtront dans un frame et pas dans l'autre) en fonction de la structure syntaxique associée.

Les rôles des sous-classes héritent des rôles de la classe mère²⁵. Cela est signalé par la mention « no roles » dans la rubrique « Role » de la sous-classe, comme dans put-spatial-9-2.1, sous-classe de put-spatial-9.2 :

ROLES
NO ROLES

Figure 11 Mode de représentation de l'héritage des rôles de la classe mère dans Verbnnet

Je vais maintenant m'intéresser à plusieurs questions théoriques que je me suis posées à l'occasion de mon stage de M2 portant sur l'« adaptation de la ressource lexico-syntaxique Verbnnet au français » et encadré par le CEA List et l'Université Paris 7.

3.2.3. Les obstacles à une transposition de Verbnnet dans d'autres langues

Tout d'abord, que serait une transposition exacte de Verbnnet dans d'autres langues ? Etant donnée la définition de (Levin, 1988) des classes anglaises comme corrélation entre une ou plusieurs alternance(s) syntaxique(s) et un ou plusieurs trait(s) sémantique(s), une transposition exacte de Verbnnet renvoie sur le plan théorique à une **hypothèse linguistique forte** selon laquelle seraient **universels** :

²⁵ Une sous-classe Verbnnet peut être définie comme un ensemble de verbes partageant non seulement l'/les alternance(s) syntaxique(s) et les rôles sémantiques de la classe mère mais aussi certaines alternances qui lui est propre.

- les **traits sémantiques** syntaxiquement pertinents dégagés pour l'anglais
- les **structures syntaxiques** constitutives des alternances
- l'**association** entre ces traits sémantiques et les alternances syntaxiques définissant proprement les classes telles qu'elles sont établies par (Levin, 1993) pour l'anglais. Autrement dit, les mêmes traits sémantiques détermineraient les mêmes alternances syntaxiques dans toutes les langues.

(Levin, 1993), tout en soutenant l'ensemble de ces trois hypothèses, émet une réserve sur un point :

« Les alternances – y compris les alternances du type de celle trouvées pour l'anglais – se manifestent à travers les langues par des verbes à chaque fois apparentés sémantiquement. [...] Ce qui ne signifie pas que toutes les langues ont le même inventaire de verbes ni d'alternances. » (Levin, 1993 : 10-11)

3.2.3.1. Le caractère idiosyncrasique des alternances

(Levin, 1993) tente d'éluder la question du **caractère** potentiellement **idiosyncrasique des alternances** en même temps qu'elle fait surface. En effet, elle enchaîne son argumentation avec l'hypothèse suivante :

« Même quand les alternances sont propres à certaines langues, elles sont souvent sensibles à des traits sémantiques verbaux servant à expliquer le comportement [syntaxique, nldr] des verbes dans d'autres langues. » (Levin, 1993 : 10-11)

On peut paraphraser cette proposition en disant que pour les mêmes verbes *Va* d'une classe anglaise *Ca* dont l'alternance *x* est spécifique à l'anglais, il existerait des traits sémantiques tels que ces traits motiveraient une alternance *y* dans une ou plusieurs autres langues et l'on pourrait remplacer *x* par *y* dans une classe de langue différente mais *de même nature* que *Ca*.

A supposer que cela soit valable, cela étend clairement le champ de définition d'une classe sous-entendu dans un premier temps par (Levin, 1993). De corrélation entre un trait sémantique et une alternance syntaxique donnée, on passe à un ensemble de traits sémantiquement pertinents pour une alternance syntaxique quelle qu'elle soit.

Cette redéfinition des classes couplée à l'affirmation d'une **adaptation** possible **des alternances syntaxiques** permet à Levin de maintenir la possibilité d'une transposition exacte. Elle ne veut pas d'une définition stricte des classes pour laquelle dire que l'on peut calquer les mêmes traits sémantiques et *éventuellement* les mêmes verbes et les mêmes alternances syntaxiques d'une classe anglaise à une classe d'une autre langue revient à admettre une transposition partielle seulement.

(Levin, 1993) affirme donc au prix d'une redéfinition des classes et d'une hypothèse linguistique forte que l'on peut, dans le cas des alternances propres à l'anglais, *adapter* les classes dans une autre langue en partant d'abord du/des trait(s) sémantique(s) syntaxiquement pertinent(s) dégagé(s) pour chacune d'entre elles, puis en cherchant quelle(s) alternance(s) propre(s) à quel(s) verbe(s) ces traits « servent à expliquer ».

3.2.3.2. L'absence de spécification des traits sémantiques syntaxiquement pertinents dans la ressource anglaise

Une remarque s'impose toutefois concernant l'adaptation des classes à partir des traits sémantiques syntaxiquement pertinents proposée par (Levin, 1993) dans le cas des alternances spécifiques à l'anglais.

En effet, l'élargissement de la définition des classes ne fait que déplacer le problème d'un endroit à un autre. Et pour cause, (Levin, 1993) fait l'impasse sur l'ajout des traits sémantiques pertinents dont elle préconise pourtant de faire appel pour la première étape de l'adaptation des classes. On peut avancer deux explications à cela. La première est que (Levin, 1993) conçoit l'adaptation des alternances sur le plan théorique plus que sur le plan pratique. La seconde est liée à la façon dont sont constituées les classes. Selon la méthode définie par (Levin, 1993), l'établissement des traits sémantiques intervient - contrairement à ce qui est préconisé pour la méthode de l'adaptation des alternances à d'autres langues - à la fin du processus de création des classes anglaises, les deux premières étapes étant le listing des alternances syntaxiques (chapitre 1 de *English Verb Classes and Alternations: A Preliminary Investigation*) et la classification des verbes en fonction des alternances observées (chapitre 2 du même ouvrage).

Prenons pour exemple l'alternance causative/inchoatif - formée sur le même modèle que (10a) et (10b) - du frame de la classe break-45.1²⁶ :

(15a) *The window broke.*

(15b) *Tony broke the window.*

Cette alternance n'est pas directement transposable au français puisque l'équivalent syntaxique français *stricto sensu* du verbe « broke » en (15a), à savoir « a cassé », ne correspond pas à ce même verbe du point de vue sémantique. Le verbe français correspondant sémantiquement à (15a) serait « s'est cassé » mais il est syntaxiquement différent du verbe anglais. Or, en l'absence de trait sémantique dans la classe break-45.1, on ne peut se baser sur cette unité pour affirmer que l'alternance avec la variante pronominale - ou quelque autre alternance - en découlerait.

Faute de disposer des traits syntaxiquement pertinents, on peut s'appuyer sur les sens Wordnet et FrameNet reliés à chaque verbe anglais dans les classes (cf. les liens hypertextes oranges dans la figure 8), mais l'on n'est pas assuré qu'ils contiennent les traits pertinents en question. L'adaptation est donc sinon impossible, du moins non objectivable en l'état actuel de la ressource anglaise.

3.3. Vers un Verbnet français ? Essai d'adaptation de Verbnet au français par le CEA List et Paris 7

En dépit de ces obstacles, des chercheurs du CEA List et de l'Université Paris 7 ont entrepris depuis 2011 d'adapter Verbnet au français. Encore en phase de développement,

²⁶ Dans le Verbnet anglais : <http://verbs.colorado.edu/verb-index/vn/break-45.1.php#break-45.1> ; dans le Verbnet français : <http://verbenet.pradet.me/class/45/>

VerbNet est toutefois accessible en ligne et les modifications effectuées sont visibles instantanément sur le même site²⁷.

Nous allons d'abord nous intéresser aux objectifs de ce projet, puis détailler la méthode adoptée pour ensuite traiter les difficultés liées aux phénomènes évoqués en 3.2.3. et rencontrées lors du stage.

3.3.1. Objectifs et spécifications

De façon implicite, c'est la définition large des classes de prédicats qui a été choisie par (Pradet *et al.*, 2014) pour leur projet d'adaptation de Verbnet au français.

Le résultat de l'adaptation de la ressource, tout en visant une description linguistique fiable, doit être *in fine* utilisé pour une tâche précise en TAL : **l'annotation en rôles sémantiques** de textes français multidomaines.

Les auteurs du Verbnet français souhaitent, comme c'est le cas pour le Verbnet anglais :

- mettre en relation cette nouvelle ressource lexicographique avec les autres existantes sur le Web (Les Verbes Français ou « LVF » et Lexique-Grammaire pour le français et FrameNet pour l'anglais).
- rendre VerbNet interopérable grâce à l'utilisation des mêmes formats de données et d'échange que ceux de Verbnet (XML)

3.3.2. Méthode

3.3.2.1. 1ère étape : projection des classes générales anglaises avec leurs frames dans des classes françaises et traduction des verbes anglais

La première étape a été réalisée grâce à deux traitements automatiques : la projection des classes générales anglaises avec leurs frames dans des classes françaises et la traduction des verbes anglais.

(Pradet *et al.*, 2014) ont reproduit l'ensemble des classes générales de Verbnet dans la ressource française dans le but de garder par la suite à peu près la même structure de base qu'en anglais. Pour chaque classe anglaise, ils ont ensuite cherché automatiquement pour chaque verbe les traductions françaises communes aux deux dictionnaires bilingues SCI-FRAN-EURADIC et Wiktionary, ont rassemblé ces dernières dans la classe française correspondant à la classe anglaise et ont associé à la classe française la classe source anglaise via un lien hypertexte.

Aux termes de cette extraction, environ la moitié des verbes anglais sont associés à une traduction française (2128 verbes français contre 4154 en anglais). Les résultats sont « précis et cohérent à la fois sémantiquement et syntaxiquement » (Pradet *et al.*, 2014).

(Pradet *et al.*, 2014) ont ensuite dû supprimer manuellement certaines classes comme pit-10.7 n'incluant que des dénominaux n'ayant pas d'équivalent en français (« to bark », « to bone »...)

²⁷ <http://verbenet.pradet.me/class/9/>

3.3.2.2. 2ème étape : filtrage semi-automatique des traductions françaises, validation manuelle des frames et division manuelle des classes en sous-classes

Pour cette étape actuellement en cours de développement, les changements peuvent être proposés par n'importe quelle personne autorisée via l'interface Web. Chaque modification effectuée est gardée en mémoire de sorte que les concepteurs gardent une trace des versions précédentes et peuvent facilement revenir en arrière.

Grâce à l'utilisation des ressources lexicographiques LVF, Lexique Grammaire et Dicovalence²⁸, des filtres sont automatiquement ajoutés aux traductions communes aux dictionnaires bilingues. Cinq couleurs sont utilisées dans l'interface Web pour les sous-ensembles : le noir représente les **verbes les plus fréquents** d'après le Dicovalence (par opposition aux verbes gris), le rouge représente les verbes présents dans LVF, le vert ceux figurant dans Lexique-grammaire et le violet les verbes présents dans l'ensemble de ces trois ressources (intersection des trois ensembles dans le schéma suivant).

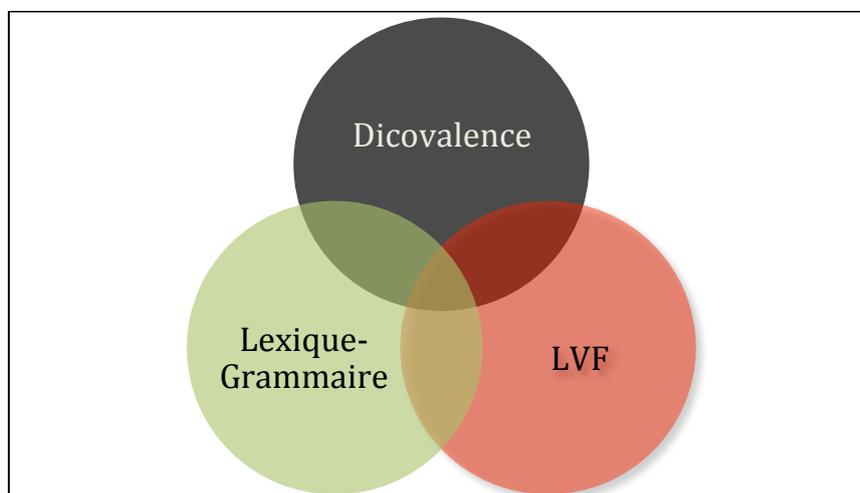


Figure 12 Diagramme des sous-ensembles de traductions communes aux dictionnaires bilingues pour l'adaptation de Verbnét au français

Les verbes violets sont donc considérés comme les plus pertinents.

Ensuite, la validation des frames consiste à vérifier si les frames automatiquement projetés de l'anglais ont un équivalent acceptable en français et sinon, chercher une autre alternance pour les verbes français traduits, c'est-à-dire « adapter les alternances syntaxiques de Verbnét au français. » (Pradet *et al.*, 2014)

Cette tâche est très liée à celle consistant à créer d'éventuelles sous-classes à partir des classes générales projetées vers le français en suivant les mêmes critères que pour le Verbnét anglais.

Pour l'un de mes deux sujet de stage, co-encadré par le CEA List et L'Université Paris 7, ma mission est d'associer aux classes Verbnét françaises dont les frames et les rôles

²⁸ Aucun des ces lexiques ne comprend d'informations relatives aux rôles sémantiques.

sémantiques ont été validés, les classes LVF dont les verbes correspondent auxdits rôles sémantiques et frames. Développé par Jean Dubois et Françoise Dubois-Charlier, LVF est une ressource lexicographique destinée à classer les verbes français selon des critères sémantiques et syntaxiques. Il est hiérarchisé selon cinq niveaux d'analyse. Le premier niveau distingue 14 classes sémantiques ; le second 4 classes sémantico-syntaxiques (combinaison entre trait +/-animé pour telle fonction grammaticale et aspect figuré ou non du sens des verbes) ; le troisième un certain nombre de classes également sémantico-syntaxiques ; et les deux derniers, des classes essentiellement sémantiques. Les classes LVF considérées dans le mapping LVF VerbNet sont celles du niveau 3.

3.3.3. Mise en perspective

3.3.3.1. Dans quelle mesure peut-on adapter les alternances syntaxiques sans affecter la nature des classes anglaises ni violer la définition générale des classes de (Levin, 1993)?

Il arrive que l'édition des frames des classes générales conduise à des classes qui n'ont plus grand chose à voir avec celles définies par (Levin, 1993).

Par exemple, la classe anglaise dine-39.5 se définit par l'alternance syntaxique composée d'une variante intransitive NP V et d'une variante transitive indirecte NP V PP.Patiente introduite par la préposition « on » :

FRAMES	
NP V	
EXAMPLE	"Cynthia breakfasted."
SYNTAX	<u>AGENT</u> V
SEMANTIC	TAKE_IN(DURING(E), AGENT, ?PATIENT)
NP V PP.PATIENT	
EXAMPLE	"Cynthia breakfasted on peaches."
SYNTAX	<u>AGENT</u> V {ON} <u>PATIENT</u>
SEMANTIC	TAKE_IN(DURING(E), AGENT, PATIENT)

Figure 13 Alternance syntaxique de la classe anglaise dine-39.5

Cette alternance se rencontre pour les verbes suivants :

MEMBERS		
BANQUET (WN 2)	GRAZE (WN 1, 5; G 1)	PICNIC (WN 1)
BREAKFAST (FN 1; WN 1)	JUNKET (WN 3; G 3)	SNACK (FN 1; WN 1)
BRUNCH (WN 1)	LUNCH (FN 1; WN 1)	SUP (FN 1; WN 1)
DINE (FN 1; WN 1; G 1)	LUNCHEON	
FEAST (FN 1; WN 1)	NOSH (FN 1; WN 1)	

Figure 14 Verbes membres de la classe anglaise dine-39.5

Or, on ne voit pas de verbes français d'Ingestion et à fortiori de traductions françaises de ces verbes qui, comme pour les verbes anglais :

– Admettent

- une construction intransitive (1^{er} frame du Verbnet informatisé)

- une construction transitive indirecte introduite par la préposition « sur » (2^{ème} frame du Verbnet informatisé)

– Et excluent

- Une construction transitive directe (cf. Levin, 1993 : 216)

La plupart des verbes français apparentés sémantiquement n'admettent pas d'alternance avec une variante transitive indirecte²⁹ (« participer à un banquet » pour « banquet », « prendre un petit déjeuner » pour « breakfast », « bruncher » pour « brunch », « dîner » pour « dine », « picniquer » pour « picnic »...). A défaut :

– de connaître les traits sémantiques anglais motivant les alternances en anglais

– de savoir si ces traits motivent d'autres alternances en français

– de savoir quelles alternances ces traits motiveraient

il paraît hasardeux de faire une classe française avec seulement la traduction du premier frame de la classe anglaise, à moins de renoncer à l'affirmation du lien de parenté entre les classes anglaise et française en renommant cette dernière par un nom autre que « dîner-39.5 » et en supprimant le lien hypertexte de la classe anglaise vers « dine-39.5 ».

3.3.3.2. Quelle méthodologie ?

Un point a attiré notre attention lors de l'édition de Verbnet : il est techniquement possible de valider sur la plateforme en ligne des frames et des rôles sémantiques à partir des verbes automatiquement traduits et filtrés par LVF, Grammaire-Lexique et Dicovalence et ensuite d'ajouter une classe LVF qui a pour effet de remplacer certains anciens verbes par de nouveaux ne correspondant plus à certains patrons syntaxique et/ou aux rôles sémantiques auparavant validés. Par exemple, pour la sous-classe française put-9-1.2, la classe LVF ayant orienté par la suite le choix et la validation des frames grâce à un traitement automatisé est L3b :

²⁹ « se régaler de » admet bien un objet indirect mais il est introduit par la préposition « de » et non pas par « sur » qui serait la traduction la plus proche de « on ».

Classe 9.1-2 L3b ↗ **38LD** ↗

∅

- Paragon : ∅
- Membres : place put set stick
- Traductions : appliquer apposer appuyer classer coller disposer enchâsser enfoncer fichier fourrer installer insérer introduire logger mettre percher piquer placer planter poser poster programmer ranger repiquer engager fixer investir plonger reculer tendre antéposer encadrer exposer imposer monter nouer passer plâtrer pointer prendre présenter reclasser sertir situer [+]
- Roles : Theme [+abstract]

NP V PP.destination NP		x
Exemple	Ils ont placé sur moi une grande responsabilité.	
Syntaxe	Agent V {sur} Destination Theme	
Sémantique	motion(during(E), Theme) not(Prep(start(E), Theme, Destination)) Prep(end(E), Theme, Destination) cause(Agent, E)	

Figure 15 Validation du frame de la classe française Verbnets put-9-1.2 à partir du filtrage des traductions

Lors de mon stage sur Verbnets, je me suis penchée sur le mapping de cette sous-classe validée en termes de frames et de rôles sémantiques. J'ai alors proposé de remplacer L3b par la classe L4b dont la définition « mettre qc qp ou trouver qc qp * Figurés de L3b » convient mieux que celle de L3b au trait +abstrait du rôle Thème :

Classe 9.1-2 L4b ↗ **38LD** ↗

∅

- Paragon : ∅
- Membres : place put set stick
- Traductions : appuyer engager fixer investir logger mettre placer appliquer apposer classer coller disposer enchâsser enfoncer fichier fourrer installer insérer introduire percher piquer planter plonger poser poster programmer ranger reculer repiquer tendre encadrer exposer régler [+]
- Roles : Theme [+abstract]

Figure 16 Verbes filtrés automatiquement après changement du mapping LVF – Verbnets de la classe Verbnets put-9-2.1

Ce qui a eu, comme on pouvait s'y attendre, pour effet de modifier la liste des verbes violets et rouges et de rendre caduc la structure syntaxique NP V PP NP du frame pourtant validé. En effet, les verbes ajoutés s'insèrent avantage dans une structure du type NP V NP PP (« mettre/fixer qch sur qch ») voire NP se V PP (« s'appuyer sur qch »).

On remarque d'ailleurs qu'il n'y a pas de réelle distinction syntaxique entre les verbes de put-9.1 (dont le patron est NP V NP PP) et ceux que l'on vient d'associer à put-9-1.2. La distinction entre la classe et sa sous-classe est essentiellement sémantique (avec pour le thème un trait +concret en 9.1 et +abstrait en 9-1.2) alors que la plupart des sous-classes répondent à un besoin de division en frames, c'est-à-dire en structures

syntaxiques. Il s'agit là d'un autre point qui n'a pas été définitivement fixé par les concepteurs de Verbnet.

3.3.3.3. Le choix de Verbnet pour l'annotation en rôles sémantiques des textes français

Le choix de Verbnet pour servir *in fine* de ressource à l'annotation en rôles sémantiques de textes français oblige à se focaliser sur l'adaptation des alternances anglaises en français. Or indépendamment des problèmes linguistiques et méthodologiques que cela pose, il n'est pas nécessaire de disposer des alternances pour réaliser la tâche d'annotation dans quelque langue que ce soit. En effet, ce qui compte pour l'annotation est non pas les alternances en soi mais plutôt les frames au sens Verbnet, c'est-à-dire les associations entre les patrons syntaxiques et les structures en rôles sémantiques pour chaque verbe du lexique, quelque soit ses similitudes de comportement syntaxique avec d'autres verbes.

En anglais, où Verbnet est, à notre connaissance la seule ressource contenant des associations entre les rôles sémantiques et les structures syntaxiques, il paraît fondé de s'appuyer sur cette ressource pour créer un système d'annotation basé sur des règles. En français, où aucune ressource ne recense de telles associations syntaxico-sémantiques, la question peut se poser de la pertinence du choix de l'adaptation de Verbnet au français.

4. Les rôles sémantiques dans le TAL aujourd'hui

On peut définir le TAL comme l' « ensemble des méthodes et programmes qui traitent des données langagières à condition de tenir compte des spécificités des langues » (Cori, 2013). L'annotation automatique sur corpus est un de ces traitements.

4.1. L'annotation en rôles sémantiques

Il est possible d'annoter un texte selon différentes catégories linguistiques dont par exemple les entités nommées, les adverbiaux calendaires et les rôles sémantiques (on parle alors d' « annotation en rôles sémantiques »).

4.1.1. Principes

Selon (Amsili *et al.*, 2012), l'annotation ou étiquetage en rôles sémantiques est une tâche de la **sémantique lexicale computationnelle** au même titre que la désambiguïsation lexicale ou la reconnaissance d'inférence textuelle.

(Maurel, 2010) note que cette tâche fait appel à des traits tels que la catégorie morpho-syntaxique des constituants, celle de leur tête ou le « chemin qui relie [les arguments] au verbe ».

L'automatisation de l'annotation en rôles sémantiques se décompose généralement en deux grandes étapes : l'identification des arguments et l'assignation à ces arguments d'étiquettes correspondant à des rôles auparavant identifiés. Les principales difficultés sont, pour la première étape, la distinction entre arguments et modifieurs et pour la seconde étape, la détermination des étiquettes de rôle ainsi que la distinction entre patient et thème.

L'expérience de (Hadouche, 2011) s'appuie sur deux modèles de description d'unités lexicales que sont la Lexicologie Explicative et Combinatoire (LEC) implémentée dans le DicoInfo³⁰ et la sémantique des frames de Fillmore. Elle se décompose en trois étapes :

- L'identification des arguments actant et circonstants confondus, avec une f-mesure à 74% pour une approche par règle et 84,8% pour l'apprentissage automatique.
- La distinction actants/circonstant dont l'apprentissage supervisé produit une f-mesure à 96%
- L'annotation en rôles sémantiques des actants, produite par apprentissage automatique

4.1.2. Exemple de système d'annotation en rôles sémantiques basé sur les règles extraites du Verbnet anglais : le programme développé au CEA List

L'approche fondée sur un système de règles facilite l'analyse des erreurs car cela permet de remonter à la source de l'application des règles. Elle est de plus adaptée à une des spécifications qui est d'« annoter de nouveaux domaines ne disposant pas de corpus annoté ».

Le programme se veut simple et reproductible.

4.1.2.1. Les ressources utilisées

Dans l'ordre chronologique des traitements effectués, l'annotation fait appel aux ressources suivantes :

- MST 0.5.0., programme qui transforme le format des données relatives aux dépendances syntaxiques (on passe du format Conll vers un ensemble de patrons syntaxiques).
- **VerbNet 3.2.** C'est cette ressource qui fournit à proprement parler les règles à appliquer pour l'annotation en rôles sémantiques. Les règles sont encodées implicitement dans l'algorithme du programme (cf. 4.2.2.2.). Elles se basent sur l'association [patron syntaxique-structure en rôles sémantiques] dans chaque variante syntaxique d'un frame (cf. figure 10). Verbnet traite des prédicats qui ne sont pas spécifiques à un domaine particulier.
- table de transformation des frames Verbnet pour la voix passive. Ceci permet de réduire le taux d'erreur de -15,7%.
- Semlink 1.2.2c, contenant une batterie de règles permettant de réécrire certaines associations entre des rôles FrameNet et des frame FrameNet en rôles sémantiques FrameNet 1.5 (ex. : rôle Verbnet Agent dans un Frame FrameNet Statement-> rôle FrameNet Speaker). Cette ressource est utilisée lors de l'évaluation.

4.2.2.2. Fonctionnement du programme

Le programme d'annotation en rôles sémantiques prend en entrée un fichier de type Conll. Conll est un format de représentation d'analyses linguistiques. Il prévoit une ligne par token et pour chaque token, un ensemble d'attributs de base dont le lemme (colonne 3), deux types de catégories morpho-syntaxiques - issues du Pen TreeBank - en fonction du grain (colonne 4 et 5) et des informations relatives aux dépendances syntaxiques (colonne 7 et 8 avec liens de dépendances syntaxiques dans le sens

³⁰ <http://olst.ling.umontreal.ca/dicoinfo>

gouverneur -> dépendant). Alors que la structure globale est fixe, les valeurs d'attributs peuvent ne pas être spécifiées en fonction des besoins.

Le système s'applique non seulement aux phrases ne contenant qu'un verbe conjugué mais aussi aux phrases complexes, comme nous allons le voir avec la proposition « However, in 2002 Russia declared it will eliminate its tactical nuclear weapons by the end of 2004 » dont la représentation Conll requise par le programme est la suivante :

1	however	however	RB	RB	-	6	VMOD	-	-		
2	,	,	,	,	-	6	P	-	-		
3	in	in	IN	IN	-	6	VMOD	-	-		
4	<num>	<num>	CD	CD	-	3	PMOD	-	-		
5	russia	russia	NNP	NNP	-	6	SUB	-	-		
6	declared	declared			VBD	VBD	-	0	ROOT	-	-
7	it	it	PRP	PRP	-	8	SUB	-	-		
8	will	will	MD	MD	-	6	VMOD	-	-		
9	eliminate	eliminate			VB	VB	-	8	VC	-	-
10	its	its	PRP\$	PRP\$	-	13	NMOD	-	-		
11	tactical	tactical			JJ	JJ	-	13	NMOD	-	-
12	nuclear	nuclear	JJ	JJ	-	13	NMOD	-	-		
13	weapons	weapons	NNS	NNS	-	9	OBJ	-	-		
14	by	by	IN	IN	-	9	VMOD	-	-		
15	the	the	DT	DT	-	16	NMOD	-	-		
16	end	end	NN	NN	-	14	PMOD	-	-		
17	of	of	IN	IN	-	16	NMOD	-	-		
18	<num>	<num>	CD	CD	-	17	PMOD	-	-		
19	-	6	P	-	-		

Figure 17 Exemple de fichier contenant une phrase au format Conll et servant d'entrée pour le programme d'annotation en rôles sémantique du CEA

Telles quelles, les informations de dépendances syntaxiques présentes dans la huitième colonne ne permettent pas de distinguer les cas d'ambiguïté syntaxique comme dans « Jean mange une glace au café ». L'ambiguïté porte sur le niveau de rattachement de l'argument « au café » qui est soit un actant - auquel doit donc être associé un rôle lors de l'annotation - soit un circonstant. Les règles heuristiques qui ont été utilisées par le programme d'annotation pour identifier les arguments à annoter (donc les actants) utilisent les étiquettes de dépendance du fichier Conll sans chercher spécifiquement à résoudre ce problème linguistique précis.

Le résultat du programme d'annotation en sortie est le fichier Conll d'entrée augmentée d'une colonne spécifiant un rôle sur chaque ligne correspondant à la tête syntaxique de l'argument annoté (figure 19).

Nous proposons, à partir de la lecture de (Pradet *et al.*, 2013) et de questions posées à son auteur, un algorithme décrivant les principaux traitements du programme d'annotation et les résultats intermédiaires pour la phrase d'exemple donnée ci-avant :

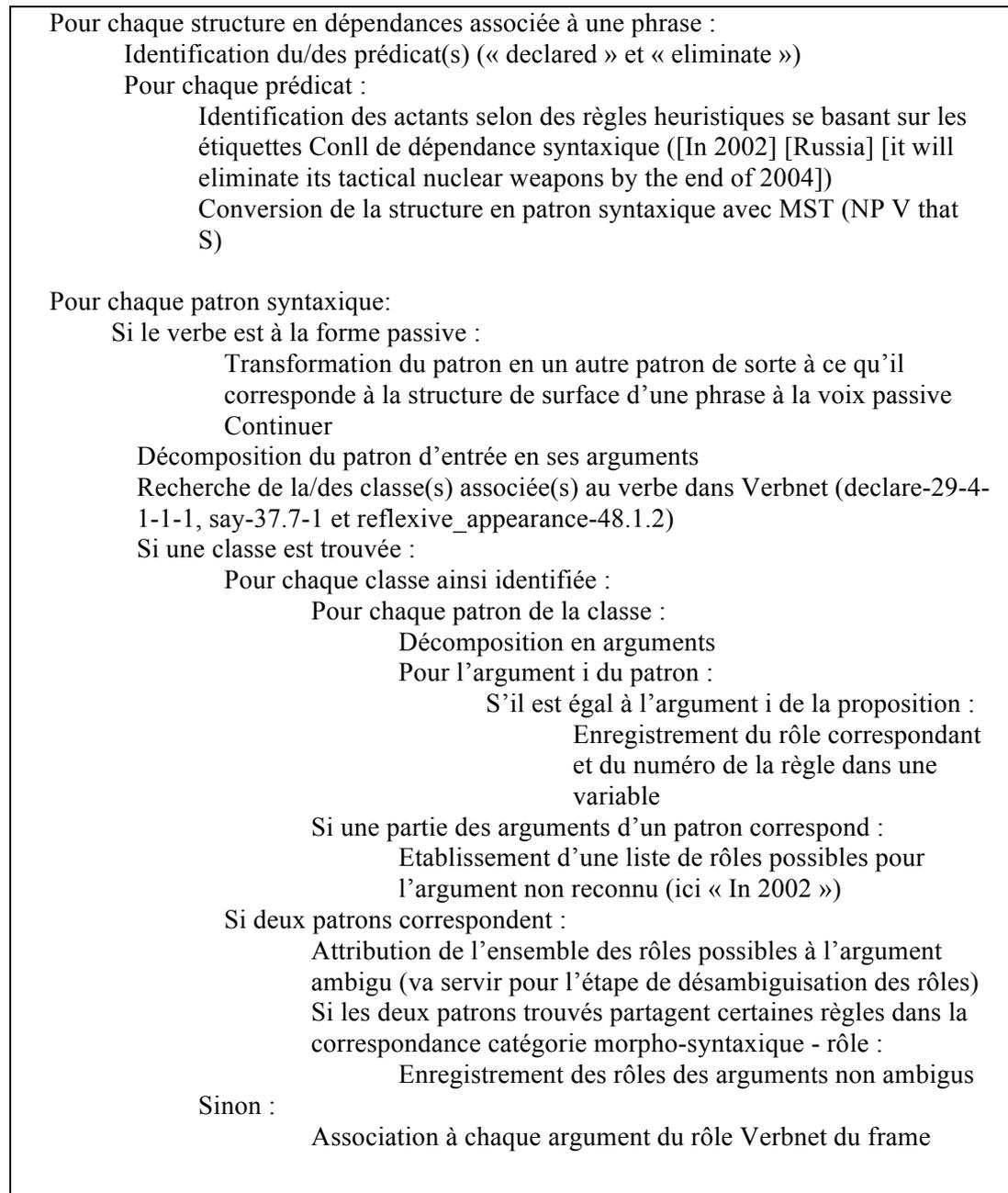


Figure 18 *Détail de l'algorithme associé au programme d'annotation en rôles sémantique du CEA d'après (Pradet et al., 2013)*

Un syntagme prépositionnel peut être reconnu comme actant sans par la suite être associé à un rôle sémantique. C'est le cas lorsqu'un patron Verbnets est trouvé, qui contient tous les arguments de la phrase d'entrée dans leur ordre d'apparition sauf le syntagme prépositionnel en question. Par exemple, l'argument [in 2002] du premier

prédicat de la phrase d'exemple n'a été associé à aucun rôle, comme on peut le voir dans le fichier Conll de sortie :

1	however	however	RB	RB	-	6	VMOD	-	-	-	-	-
2	,	,	,	,	-	6	P	-	-	-	-	-
3	in	in	IN	IN	-	6	VMOD	-	-	-	-	-
4	<num>	<num>	CD	CD	-	3	PMOD	-	-	-	-	-
5	russia	russia	NNP	NNP	-	6	SUB	-	-	-	Agent	-
6	declared	declared			VBD	VBD	-	0	ROOT	-	-	-
say-37.7-1												
7	it	it	PRP	PRP	-	8	SUB	-	-	-	-	Agent
8	will	will	MD	MD	-	6	VMOD	-	-	-	Topic	-
9	eliminate	eliminate			VB	VB	-	8	VC	-	-	-
remove-10.1												
10	its	its	PRP\$	PRP\$	-	13	NMOD	-	-	-	-	-
11	tactical	tactical	JJ	JJ	-			13				
NMOD												
12	nuclear	nuclear	JJ	JJ	-	13	NMOD	-	-	-	-	-
13	weapons	weapons	NNS	NNS	-	9	OBJ	-	-	-	-	Theme
14	by	by	IN	IN	-	9	VMOD	-	-	-	-	-
15	the	the	DT	DT	-	16	NMOD	-	-	-	-	-
16	end	end	NN	NN	-	14	PMOD	-	-	-	-	-
17	of	of	IN	IN	-	16	NMOD	-	-	-	-	-
18	<num>	<num>	CD	CD	-	17	PMOD	-	-	-	-	-
19	-	6	P	-	-	-	-	-

Figure 19 Sortie Conll du programme d'annotation en rôles sémantiques du CEA avec les étiquettes de Verbnnet

Seuls les arguments « Russia » et « It will eliminate its tactical nuclear weapons by the end of 2004 » sont associés à un rôle Verbnnet (respectivement Agent et Topic).

Le but de mon deuxième sujet de stage au CEA est dans un premier temps d'intégrer le programme d'annotation en rôles sémantiques dans un programme plus large d'analyse linguistique appelé Lima. Pour ce faire, je dois créer deux programmes en C++ à l'intérieur de l'architecture de Lima : un « dumper » produisant, à partir des variables Lima, le fichier de sortie Conll attendu en entrée par le programme d'annotation (cf. figure 17) et un « loader » lisant les données du fichier de sortie du programme d'annotation (cf. figure 19) pour les intégrer dans de nouvelles variables Lima. Ainsi, on peut résumer l'intégration en deux étapes principales chacune correspondant au programme crée :

- mise en correspondance de l'entrée du programme d'annotation et de la sortie du dumper de Lima
- chargement des résultats de la sortie du programme d'annotation dans Lima

On appelle cette façon d'intégrer un programme dans un autre l'externalisation³¹ d'une unité de traitement (ici le programme d'annotation).

J'ai, pour le dumper, travaillé à partir d'un dumper existant sur l'extraction d'entités nommées. Le code de mon programme est consultable au fur et à mesure de mes

³¹ « externalisation » du point de vue du programme principal dans le sens où le programme d'annotation est externe (et autonome) au programme Lima.

envois de modifications sur le serveur Github³² permettant de gérer des projets collaboratifs.

4.1.2.3. Evaluation du programme d'annotation

Pour l'évaluation, (Pradet *et al.*, 2013) ont pris comme corpus de test un corpus full-text de FrameNet comportant des textes de domaines variés dont des articles du Wall Street Journal. Deux batteries de test ont été effectuées : une portant uniquement sur l'attribution des étiquettes de rôles sémantiques, une autre portant sur la fonction d'identification des arguments en plus de l'assignation des rôles. Pour chacune des deux batteries, étant donné que les étiquettes servant de référence (FrameNet) et celles produites par le système (Verbnet) ne sont pas les mêmes, les auteurs ont dû ajouter une couche à l'annotation en rôles sémantiques Verbnet : la mise en correspondance entre les classes Verbnet et les classes FrameNet. Ce système additionnel est compris dans l'évaluation.

Au sein de chaque batterie de test, on distingue trois composants. Le composant de base est l'annotation en rôles sémantiques Verbnet et la mise en correspondance Verbnet FrameNet, auquel on peut ajouter deux composants supplémentaires : l'annotation des arguments ambigus à partir d'un modèle appris et la prise en compte de la voix passive. Pour les tests effectués sur le corpus dont le découpage en arguments est maintenu, deux indicateurs sont considérés :

- la précision, c'est-à-dire le nombre de rôles corrects sur le nombre de rôles générés
- la f-mesure ou la moyenne harmonique entre la précision et le rappel (à savoir le nombre de rôles générés par rapport au nombre attendu de rôles).

32

https://github.com/clemance/lima/blob/master/lima_linguisticprocessing/src/linguisticProcessing/core/AnalysisDumpers/ConllDumper.cpp

	f-mesure	précision
Mise en correspondance Verbnet FrameNet	70,48%	53,09%
Mise en corresponaence Verbnet FrameNet + désambiguïsation des rôles	72,02%	58,28%
Mise en correspondance Verbnet FrameNet + désambiguïsation des rôles + voix passive	76,40%	62,72%

Tableau 5 *Evaluation du programme d'annotation en rôles sémantiques d'arguments préalablement identifiés développé par (Pradet et al., 2013)*

Pour les test effectués sur le corpus ne comprenant aucune indication ni sur les rôles ni sur les arguments à annoter, les taux diminuent :

	f-mesure	précision
Mise en correspondance Verbnet FrameNet	46,75%	29,12%
Mise en correspondance Verbnet FrameNet + désambiguïsation des rôles	46,78%	33,49%

Tableau 6 *Evaluation du programme d'identification des arguments et d'annotation en rôles sémantiques développé par (Pradet et al., 2013)*

4.2. Applications TAL utilisant des corpus annotés en rôles sémantiques

Les applications TAL pouvant s'appuyer sur des corpus annotés en rôles sémantiques sont des « applications [...] nécessitant des informations sémantiques » (Hadouche, 2011). Elles peuvent se classer en deux catégories selon le domaine d'application visé.

4.2.1. Applications pour le TAL : la désambiguïsation du sens des verbes

Dans l'annotation en rôles sémantiques, les verbes peuvent être marqués comme « prédicat » ou « déclencheur », mais il n'est pas nécessaire de leur attribuer une étiquette les caractérisant plus précisément. La tâche de désambiguïsation du sens des verbes, tout en s'appuyant sur le résultat de l'annotation en rôles sémantiques et en étant susceptible de lui servir de ressource en retour, s'inscrit en dehors de cette tâche.

La désambiguïsation suppose que l'annotation des rôles sémantiques ait été réalisée manuellement, c'est-à-dire que l'annotateur ait assigné des rôles aux arguments en fonction de son interprétation implicite du sens du verbe. Si on annote des arguments avec des étiquettes de rôles induisant un sens précis des verbes de façon automatique, cela signifie en effet que le sens du verbe a déjà été déterminé en amont.

Plus les rôles sémantiques sont précis, plus l'on a de chance de pouvoir distinguer les différents sens d'un verbe.

4.2.2. Applications pour les technologies du langage

4.2.2.1. Traduction automatique

Le résultat de l'assignation de rôles sémantiques aux participants peut être exploité pour la traduction automatique dans le sens où il permet de « capturer la similarité sémantique entre les langues même si leur structure syntaxique est différente ». (Hadouche, 2011).

Soit les phrases :

(16a) *He swam across the river.*

(16b) *Il traversa la rivière à la nage.*

Bien que présentant des arguments internes différents, (16a) et (16b) ont la même structure en rôles sémantiques.

4.2.2.2. Extraction d'informations d'après (Mouton, 2010)

(Mouton, 2010) distingue trois intérêts principaux de l'annotation en rôles sémantiques pour la recherche d'information au sens large. Cela suppose toutefois de faire appel aux frames tels que ceux proposés dans FrameNet, en plus des seules étiquettes de rôle.

Les deux premiers concernent les **moteurs de recherche** fonctionnant par **mots-clés**. D'abord, une analyse en rôles sémantiques des mots clés de la **requête** pourrait servir à élargir le champs de recherche de documents. Mouton donne l'exemple de la requête complexe « école de voile du port » où « école », ayant été identifié comme déclencheur du frame « local_by_use » peut être commuté par d'autres déclencheurs nominaux de ce frame comme « centre » ou « club » dans une requête élargie utilisant des expressions régulières du type « [centre/club/école] de voile du port ».

Ensuite, une analyse des documents lors de l'indexation peut réduire le silence : si un mot de la requête a été trouvé dans le corpus mais que le frame auquel il a été rattaché ne correspond pas à celui associé au mot de la requête, on ne considèrera pas le document comme pertinent. L'analyse sémantique des documents indexés peut aussi permettre à l'utilisateur d'élargir les résultats grâce à la non spécification de certains termes de la requête. Par exemple, à partir de la requête « école de voile du port », l'utilisateur pourra choisir de reformuler l'interprétation « n'importe quelle école proposant la voile comme activité et située dans un port » en « n'importe quelle école proposant n'importe quelle activité et située dans un port ».

Enfin, l'annotation en rôles sémantique est d'un intérêt notable pour les systèmes de question réponse pour lesquels la requête est exprimée sous forme non pas de mot(s)

clé(s) mais de question contenant un verbe voire un prédicat non verbal et pouvant être ainsi analysée en termes de rôles sémantiques. Le corpus interrogé doit également avoir été analysé en rôles sémantiques. La résultat du système est la phrase (voire un ensemble de phrases) considérée comme répondant à la question.

En fait, pour une requête sous forme de question, on va chercher parmi les phrases du corpus contenant le frame identifié à partir de la question, celles qui contiennent le rôle sémantique non seulement de l'argument sur lequel porte l'interrogation (en général un pronom interrogatif) mais aussi de ceux des autres arguments de la question, et ce, quel que soit leur ordre dans les phrases du corpus. Le déclencheur du frame peut être identique ou non à celui de la requête en termes de lexème et de catégorie morphosyntaxique (verbes, noms...).

Ainsi - d'après l'exemple et la représentation de (Mouton, 2010) - pour une question comme :

(17a) *Qui a inventé l'imprimerie ?*

dont l'annotation est :

(17b) *QuiCognizer [a inventé]Achieving_first l'imprimerieNew_idea*

on la reformulera à l'aide d'expressions régulières sous la forme :

(17c) **Cognizer [*]Achieving_first l'imprimerieNew_idea*

et la phrase trouvée pourra être :

(17d) *Si les ChinoisCognizer [sont] bien [à l'origine de]Achieving_first l'imprimerieNew_idea pour un grand nombre d'historiens, l'imprimerie moderneNew_idea [reste le fait de]Achieving_first l'Allemand Johan Gutenb erg qui, vers 1440, fond des caractères mobiles pouvant être réutilisés indéfinimentCognizer*

Pour ce genre d'application, l'annotation des textes en rôles peut être encodée dans des langages de représentation de connaissances et remplir des bases de connaissances interrogeables après transformation en langage formel de la requête en langue naturelle.

On voit que l'intérêt de l'extraction d'informations à partir des rôles sémantiques réside dans le niveau d'abstraction relativement élevé de ces derniers (du moins par rapport au critère classique de présence ou non d'un mot dans un texte). Ce niveau d'abstraction est représentable par la question composée uniquement de pronoms relatifs indéfinis : « QUI fait QUOI, OU, QUAND et COMMENT ? »

On peut envisager l'exploitation des corpus annotés en rôles sémantiques pour d'autres applications comme le résumé automatique de texte.

Conclusion

Nous avons ainsi pu nous rendre compte que les principales théories traitant des rôles sémantiques ancrèrent leur description soit dans une perspective sémantique comme avec la Sémantique des Frames de (Fillmore, 1976) soit à l'interface sémantique-syntaxique avec la grammaire de Cas du même auteur voire la Théorie des Principes et Paramètres. Nous avons ensuite montré comment s'était effectué le passage de certaines de ces théories à des systèmes de description du lexique pour l'anglais. Puis nous avons vu en

quoi ces descriptions pouvaient être exploitées pour l'annotation en rôles sémantiques basée sur des règles avec l'exemple du programme développé au CEA, et en quoi les annotations en rôles sémantiques pouvaient à leur tour servir à des systèmes TAL nécessitant en particulier des informations sémantiques.

L'analyse en rôles sémantiques est une analyse sémantique de surface que certains phénomènes sémantiques plus complexes peuvent venir « parasiter ». Dans la perspective d'une analyse sémantique plus élaborée, elle gagnerait à être combinée avec une analyse discursive.

Cela permettrait de distinguer notamment au sein de la classe des prédicats verbaux différents niveaux de rattachement d'un contenu prédicatif en fonction de l'énonciateur ainsi que différents niveaux de prise en charge énonciative (modalité) et donc différents degrés de factualité. Ainsi, pour une phrase du type « In 2002, Russia declared it will eliminate its tactical nuclear weapons by the end of 2004 », le contenu prédicatif « it will eliminate its tactical nuclear weapons by the end of 2004 » serait considéré comme pris en charge non plus par l'énonciateur premier (nommément le journaliste) comme c'est actuellement le cas par défaut dans les systèmes d'annotation en rôles sémantiques mais par un énonciateur second, référent du sujet de la phrase à savoir la Russie. Pour une autre phrase comme « In 2002, Russia *claimed* it will eliminate its tactical nuclear weapons by the end of 2004 », la prise en compte du caractère non seulement énonciatif mais aussi modal du verbe, « claim » permettrait rendre compte de l'amenuisement du degré de factualité du même contenu prédicatif.

Un autre phénomène qui pourrait être pris en compte en complémentarité de l'analyse en rôles sémantiques serait la valeur axiologique de certains prédicats. Cette valeur peut être intrinsèque au prédicat (par exemple, le verbe « proliférer » qui dans son sens métaphorique revêt une connotation péjorative) ou actualisée au moment de l'énonciation.

Table des figures et tableaux

Figure 1 Représentation d'après la logique des prédicats d'une phrase construite autour d'un verbe transitif.....	5
Figure 2 Règles de réécriture et représentation syntaxique des phrases à la voix active et composées d'une proposition principale en construction transitive (schéma emprunté à (Kasper, 2008))	5
Figure 3 Exemple de règles d'association ou de « Linking » entre rôles sémantiques et position structurale	18
Figure 4 Co-indexation dans FrameNet des éléments de frame et de leur(s) réalisation(s) lexicale(s) dans les rubriques respectives Définition et FE et entre ces rubriques.....	23
Figure 5 Co-indexation des éléments de frame entretenant une relation d'héritage dans FrameNet grâce au système de surlignage	23
Figure 6 Déduction du sens de la relation d'héritage entre les éléments de frame grâce à la rubrique « frame-frame Relations » dans FrameNet.....	24
Figure 7 Classe FrameNet dont hérite « Intentionally_create » et contenant l'élément de frame générique « Agent »	24
Figure 8 Rubrique « Membres » de la classe Verbnet put_spatial-9.2.....	27
Figure 9 Rubrique « Roles » de la classe Verbnet put_spatial-9.2.....	27
Figure 10 Différents ordres d'apparition des rôles sémantiques en fonction de l'ordre syntaxique impliqué dans l'alternance propre à un frame dans la classe Verbnet put_spatial-9.2.....	28
Figure 11 Mode de représentation de l'héritage des rôles de la classe mère dans Verbnet	28
Figure 12 Diagramme des sous-ensembles de traductions communes aux dictionnaires bilingues pour l'adaptation de Verbnet au français.....	32
Figure 13 Alternance syntaxique de la classe anglaise dine-39.5	33
Figure 14 Verbes membres de la classe anglaise dine-39.5	33
Figure 15 Validation du frame de la classe française Verbnet put-9-1.2 à partir du filtrage des traductions	35
Figure 16 Verbes filtrés automatiquement après changement du mapping LVF – Verbnet de la classe Verbnet put-9-2.1	35
Figure 17 Exemple de fichier contenant une phrase au format Conll et servant d'entrée pour le programme d'annotation en rôles sémantique du CEA.....	38
Figure 18 Détail de l'algorithme associé au programme d'annotation en rôles sémantique du CEA d'après (Pradet et al., 2013).....	39
Figure 19 Sortie Conll du programme d'annotation en rôles sémantiques du CEA avec les étiquettes de Verbnet.....	40
Tableau 1 Décomposition du linking de l'exemple des figures 1 et 2	6
Tableau 2 Caractéristiques syntaxiques et morphosyntaxiques des principaux déclencheurs verbaux du frame « commercial transaction » en anglais.....	11
Tableau 3 Trois configurations syntaxiques possibles d'assignation des théta-rôles par le prédicat	13
Tableau 4 Matrice de validation des conditions de description de l'alternance locative en fonction du nombre de listes de rôles	18
Tableau 5 Evaluation du programme d'annotation en rôles sémantiques d'arguments préalablement identifiés développé par (Pradet et al., 2013)	42
Tableau 6 Evaluation du programme d'identification des arguments et d'annotation en rôles sémantiques développé par (Pradet et al., 2013)	42

Bibliographie

- Abdi R., Représentation OWL de la ressource lexicale LVF et son utilisation dans le traitement automatique de la langue, Mémoire de master, Université de Montréal, 2013, Accessible en ligne sur <<http://rali.iro.umontreal.ca/rali/sites/default/files/publis/ABDI-MSc.pdf>>. Consulté le 16.05.2014.
- Amsili P., Candito M., « Analyse sémantique automatique : semantic role labeling », Support de cours, 2012, Accessible en ligne sur <http://www.linguist.univ-paris-diderot.fr/~amsili/Ens13/pdf/SemM2_semantic_role_labeling-part1.pdf>. Consulté le 29.01.2014.
- Baker, C.-F., Fillmore C.-J., Lowe J.-B., « The Berkeley FrameNet project », ACL-COLING 98, 1998, Accessible en ligne sur <http://acl.ldc.upenn.edu/C/C98/C98-1013.pdf?origin=publication_detail>. Consulté le 14.05.2014.
- Blake, F. R., « A Semantic Analysis of Case ». *Language*, vol. 6, n° 4, 1930.
- Booij G., *Lexique et syntaxe en grammaire générative*, 1988, Accessible en ligne sur <<http://books.google.fr/books?id=00Ky4PkWsQ0C&printsec=frontcover&dq=Lexique+et+syntaxe+en+grammaire+g%C3%A9n%C3%A9rative&hl=fr&sa=X&ei=eqxkU9-hEYHeOcvagOAM&ved=0CDQQ6AEwAA#v=onepage&q=causatif%20BREAK&f=false>>. Consulté le 29.03.2014.
- Cori M., « Généralités sur le TAL », Support de cours de « Méthodologie de la recherche. Epistémologie du TAL », Master 2 Ingénierie linguistique, Université Paris Ouest Nanterre La Défense, 2013-2014.
- Ferret K., « Lexique et syntaxe », Support de cours de Théories syntaxiques, Licence 3 Sciences du langage, Université Paris III Sorbonne Nouvelle, 2011-2012.
- Filliozat P.-S., « PĀNINI (~V^e s. env.) », *Encyclopædia Universalis* [en ligne], Accessible en ligne sur <<http://www.universalis-edu.com/encyclopedie/panini/>>. Consulté le 25.04.2014.
- Fillmore C., « The Case for Case », Bach and Harms (Ed.): *Universals in Linguistic Theory*. New York: Holt, Rinehart, and Winston, 1968.
- Fillmore C., « Frame semantics and the nature of language », *Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech*. vol. 280, 1976.
- Hadouche F., Annotation syntaxico-sémantique des actants en corpus Spécialisé, Thèse de doctorat, Université de Montréal, 2011a, Accessible en ligne sur <https://papyrus.bib.umontreal.ca/xmlui/bitstream/handle/1866/5032/Hadouche_Fadila_HF_2011_these.pdf?sequence=2> Consulté le 29.03.2014.
- Hadouche F., Lapalme G., L'Homme M.-C., « Attribution de rôles sémantiques aux actants des lexies verbales », *Traitement automatique des langues TALN 2011*, 27 juin au 1^{er} juillet 2011, Avignon, 2011b, Accessible en ligne sur <http://www.lirmm.fr/taln2011/PDF_court/Hadouche_taln11_submission_192.pdf>. Consulté le 29.03.2014.
- Hamm F., « Frame Semantics », Patrick Colm Hogan (ed.), *The Cambridge Encyclopedia Of The Language Sciences*, Cambridge University Press, 2009, Accessible en ligne sur <http://www.uni-stuttgart.de/linguistik/sfb732/files/hamm_framesemantics.pdf>. Consulté le 07.04.2014.

- Hirschbühler P., Labelle M., *Syntaxe du français, l'universel et le particulier dans la langue*, Montréal, Québec: Université de Montréal, 1992, Accessible en ligne sur <<http://www.courseweb.uottawa.ca/LIN3710/hl/hirschbuhler-labelle-7.pdf>>.
- Kasper, a comparison of thematic role theories, Marburg University, 2008 , Accessible en ligne sur <<http://www.uni-marburg.de/fb09/dsa/mitarbeiter/kasper/Magisterarbeit> >.
- Levin B., *English Verb Classes and Alternations: A Preliminary Investigation*, University of Chicago Press, 1993.
- Maurel D., Notes de lecture, 2010.
- Mouton C., Ressources et méthodes semi-supervisées pour l'analyse sémantique de textes en français, Thèse de doctorat, Université Paris 11 – Paris Sud UFR d'informatique, 2010.
- Petruck Miriam R. L., « Frame Semantics », J - O . Östman , J. Verschueren, et J. Blommaert (eds.) *Handbook of Pragmatics*, Amsterdam & Philadelphia, 1996, Accessible en ligne sur <http://www.princeton.edu/~adele/LIN_106:_UCB_files/Miriam-Petruck-frames.pdf>. Consulté le 7.04.2014.
- Pradet Q., Danlos L., De Chalendar G., « Adapting VerbNet to French using existing resources », LREC 2014, Reykjavik, Iceland. Accessible en ligne sur <<http://quentin.pradet.me/articles/2014adapting/pradet2014adapting.pdf>>. Consulté le 20.04.2014.
- Pradet Q., De Chalendar G., Pujol G., « Revisiting knowledge-based Semantic Role Labeling », LTC'13, 2013, Accessible en ligne sur <<http://quentin.pradet.me/articles/2013srl/pradet2013revisiting.pdf>>. Consulté le 4.04.2014.
- Radimsky J., « Actants, arguments et rôles sémantiques : combien de niveaux d'analyse? », In: Tomasziewicz T., Vetulani G., *L'apport linguistique et culturel français à l'Europe : du passé aux défis de l'avenir*, Lask, Leksem, 2012, p. 97-103.
- Riegel, M, Pellat, J.C., Rioul, R., *Grammaire méthodique du français*. Paris: PUF, 2009.
- Ruppenhofer J., Ellsworth M., Petruck M. R. L., C. R. Johnson, J. Scheffczyk, *FrameNet II: Extended Theory and Practice*, International Computer Science Institute, Berkeley CA, 2010, Accessible en ligne sur <<https://framenet2.icsi.berkeley.edu/docs/r1.5/book.pdf>>. Consulté le
- Saint-Dizier P., « Rôles thématiques », *Sémanticopédie Dictionnaire de Sémantique*, 2006, Accessible en ligne sur <http://www.semantique-gdr.net/dico/index.php/R%C3%B4les_th%C3%A9matiques>. Consulté le 1.05.2014.
- Stowell T., *Origins of Phrase Structure*, Thèse de doctorat, MIT, 1981.