

Des descriptions linguistiques à leurs représentations

Yvette Yannick Mathieu

▶ **To cite this version:**

Yvette Yannick Mathieu. Des descriptions linguistiques à leurs représentations. Linguistique. Université Paris-Diderot - Paris VII, 2009. tel-00372133

HAL Id: tel-00372133
https://theses.hal.science/tel-00372133

Submitted on 31 Mar 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Mémoire d'Habilitation à diriger des Recherches

Des descriptions linguistiques à leurs représentations

Yvette Yannick Mathieu

27 janvier 2009 - Université Paris 7

Jury :

Nicholas Asher, Directeur de recherche au CNRS
Laurence Danlos, Professeur à l'Université Paris 7
Christiane Fellbaum, Research Professor, Princeton University (rapporteur)
Alain Kihm, Directeur de Recherche au CNRS (rapporteur)
Jean-Luc Minel, Professeur, Université Paris X
Paul Sabatier, Directeur de Recherche au CNRS (rapporteur)

TABLE DES MATIÈRES

INTRODUCTION ____ *3*

CHAPITRE 1. QUATRE ÉTUDES IMPORTANTES SUR LE LEXIQUE ____ *7*

1.1. LE LEXIQUE GRAMMAIRE ____ **7**

1.1.1. Approche du Lexique Grammaire ____ 7

1.1.2. Les verbes de sentiment et psychologiques ____ 11

1.1.3. Les tables du lexique grammaire comme ressource en TAL ____ 13

1.2. L'ÉTUDE DES VERBES DE BETH LEVIN ____ **15**

1.2.1. Les alternances ____ 15

1.2.2. Les 49 classes de Beth Levin ____ 17

1.2.3. Les verbes de sentiment et psychologiques ____ 18

1.2.4. La classification de Beth Levin comme ressource en TAL ____ 21

1.3. WORDNET ____ **24**

1.3.1. Organisation du lexique ____ 25

1.3.2. Les wordnets ____ 28

1.3.3. WordNet, Ontologie et TAL ____ 28

1.3.4. Wordnet et la classification de Beth Levin ____ 29

1.3.5. Exemple du verbe adore (Wordnet 2.1) ____ 30

1.3.6. En guise de synthèse ____ 31

1.4. LE DICTIONNAIRE DES VERBES FRANÇAIS DE DUBOIS ET DUBOIS-CHARLIER ____ **32**

1.4.1. Description du dictionnaire ____ 32

1.4.2. La classe des verbes psychologiques ____ 34

1.4.3. Le dictionnaire des verbes comme ressource en TAL ____ 34

CHAPITRE 2. SENTIMENTS, ÉMOTIONS, ET OPINIONS : DESCRIPTIONS LINGUISTIQUES ____ *37*

2.1. LES SENTIMENTS ET ÉMOTIONS ____ **37**

2.1.1. Le mot ____ 37

2.1.2. La proposition ____ 45

2.1.3. La phrase : formation de paraphrase par nominalisation ____ 47

2.1.4. Les sentiments dans un texte : traitements automatiques ____ 48

2.2. LES OPINIONS ____ **56**

2.2.1. Le mot ____ 56

2.2.2. Le segment et la phrase : Relations rhétoriques et évaluation des opinions ____ 60

2.2.3. Le texte : Évaluation et Annotation des expressions d'opinion ____ 61

2.3. BILAN COMPARATIF DE NOS DESCRIPTIONS AVEC CELLES DU LEXIQUE GRAMMAIRE, DE BETH LEVIN ET DE WORDNET **65**
2.3.1. Lexique grammaire 65
2.3.2. Les verbes de Beth Levin 66
2.3.3. WordNet 66
CHAPITRE 3. MODÈLES DE DONNÉES ET REPRÉSENTATIONS DES CONNAISSANCES ***69***
3.1. AVANT LES SENTIMENTS : *TRANSLEGS* ET *INTERFACILE* **69**
3.1.1. TRANSLEGS 69
3.1.2. INTERFACILE 73
3.1.3. Comparaison de TRANSLEGS et INTERFACILE 75
3.2. LES SENTIMENTS : *FEELING, FEELING-G* ET L'ONTOLOGIE DES SENTIMENTS **76**
3.2.1. FEELING 76
3.2.2. FEELING-G 81
3.2.3. Du lexique sémantique à l'ontologie 82
3.2.4. Annotation et navigation à travers les sentiments 83
3.3. REPRÉSENTATION DES EXPRESSIONS D'OPINION **91**
3.3.1. Représentation d'un segment d'opinion 91
3.3.2. Représentation discursive d'un texte 93
3.3.3. Représentation générale des opinions dans un texte 93
3.4. ÉVOLUTION DES REPRÉSENTATIONS DES CONNAISSANCES UTILISÉES **96**
CONCLUSION ***99***
BIBLIOGRAPHIE ***103***
TABLE DES FIGURES ***109***
TABLE DES TABLEAUX ***111***

INTRODUCTION

Bien souvent, la façon de considérer la langue est très différente du point de vue des informaticiens et de celui des linguistes. Si ces divergences tendent à s'amenuiser, elles étaient importantes quand nous avons commencé à nous intéresser à la linguistique et au TAL, au début des années 1980. Nous avons pu le vérifier lors de notre mémoire de DEA d'informatique fondamentale (option traitement de l'information, Université Paris 6, 1981) qui avait pour sujet « La correction automatique des fautes d'orthographe » dans un texte. Nous venions d'arriver au LADL dirigé par Maurice Gross, et lui avions demandé conseil pour la façon d'aborder ce problème. Mais lui et le professeur d'informatique responsable du DEA, avaient des avis totalement divergents au sujet du traitement de la langue. Pour le linguiste, bien sûr, la langue était l'objet d'étude, phénomène extrêmement complexe, qu'il faudrait simplifier (bien que difficilement simplifiable) pour un traitement automatique, alors que du point de vue de l'informaticien l'important était la réalisation informatique, la langue n'étant qu'une application comme une autre, ne présentant pas de difficultés particulières. Notre approche rejoint entièrement celle de Maurice Gross, mais nous avons tenté dans nos travaux d'avoir à la fois une approche linguistique et une approche informatique, tout en essayant de ne pas appauvrir la description linguistique lors de la formalisation et de l'implémentation.

Les recherches du LADL étaient essentiellement orientées vers la constitution du lexique grammaire du français, c'est pourquoi nous nous sommes très vite intéressée au lexique. Mais, jusqu'à notre travail de thèse, notre activité était essentiellement orientée vers le choix des représentations et leur implémentation. La linguistique et la sémantique lexicale en particulier sont devenues des éléments centraux de notre activité de recherche à partir de notre travail de thèse sur les verbes de sentiment, en 1994. Nous avons approfondi et complété cette étude en y intégrant les noms et les adjectifs ainsi que certaines propriétés

les caractérisant. Nous avons ainsi fait une description linguistique détaillée et formalisée de l'expression des sentiments en français, dont nous avons vérifié la validité par le développement du système *FEELING*. Des travaux connexes nous ont montré l'intérêt d'exploiter ces descriptions pour le repérage des expressions de sentiment dans des textes, et la nécessité de les représenter avec un format standard comme celui qui sera défini plus loin des « ontologies ».

Puis, nous avons élargi ce champ sémantique à un domaine voisin, celui des opinions. Ces deux champs sémantiques sont proches, ils relèvent tous deux de la sphère affective ou intellectuelle d'une personne, et se recoupent en partie car de nombreuses opinions sont l'expression de sentiments comme l'indignation, la colère ou l'admiration. Nous avons adopté une approche légèrement différente car nous avons d'abord établi une typologie sémantique des expressions d'opinions que nous avons ensuite associée à une analyse de la structure rhétorique des segments textuels contenant ces expressions, ce qui permet une étude détaillée des opinions dans les textes. Ces travaux sur les opinions sont menés en collaboration avec Farah Benamara et Nicholas Asher (IRIT, Toulouse).

Nous n'aborderons pas dans ce mémoire d'autres thématiques de recherche auxquelles nous nous sommes intéressée et qui ont contribué à nourrir notre réflexion comme l'étude des « formes en *–o* », par exemple *franco-québécois* ou *médico-légal*, en collaboration avec Blandine Courtois et André Dugas (Dugas *et al.* 1992, 1995), ou l'extraction automatique de néologismes (Mathieu 1998a, 1998b). Une part importante de nos recherches a été consacrée à la modélisation des connaissances ; c'est ainsi que nous nous sommes intéressée aux grammaires de construction (Mathieu 2003), développées en particulier par Charles Fillmore (1988).

Ce mémoire s'organise en trois parties. Dans la première partie nous présentons quatre approches sur le lexique qui nous semblent des étapes importantes parmi les nombreux travaux qui lui ont été consacrés. Ces études ont été réalisées avec des objectifs différents des nôtres, mais sont voisines par certains aspects de l'approche que nous avons adoptée dans nos descriptions. Pour une lecture plus facile, nous avons préféré, bien que cela puisse sembler parfois fastidieux, laisser des exemples complets plutôt que de les mettre en annexe.

Les deux parties suivantes illustrent bien les deux axes de recherche que nous avons eus constamment en parallèle lors de nos travaux : une approche de linguistique descriptive en sémantique lexicale, et une approche formelle en vue de traitements automatiques dans la représentation des connaissances. Nous expliciterons les différentes descriptions linguistiques que nous avons réalisées et leur complexification au fil du temps, et nous montrerons comment les représentations de ces connaissances, et les traitements et outils pour les exploiter, ont suivi une évolution parallèle. Cela nous amènera parfois à certaines redites, nécessaires pour la compréhension du lecteur.

Plutôt que de faire un exposé exhaustif de ces travaux, qui ont été publiés par ailleurs, nous en dégagerons des éléments permettant d'expliquer les problématiques auxquelles nous nous sommes intéressée dans notre parcours.

Dans la présentation de ce mémoire nous utilisons le « nous » en place du « je », excepté pour la partie consacrée d'une part au système INTERFACILE et d'autre part aux travaux sur les opinions, où il s'agit d'un « nous » collectif qui désigne respectivement Paul Sabatier et moi-même, et Nicholas Asher, Farah Benamara et moi-même.

CHAPITRE 1. QUATRE ÉTUDES IMPORTANTES SUR LE LEXIQUE

Nous présentons dans ce chapitre plusieurs approches importantes dans l'étude du lexique. Les trois premières (le Lexique grammaire de Maurice Gross et collaborateurs, les classes de verbes de Beth Levin et WordNet) sont des travaux qui nous ont influencée à divers titres ; bien qu'il tienne une place moindre dans nos travaux nous mentionnons également le dictionnaire des verbes de Jean Dubois et Françoise Dubois-Charlier car il s'inscrit parmi les descriptions importantes des verbes du français

.

1.1. LE LEXIQUE GRAMMAIRE

1.1.1. Approche du Lexique Grammaire

L'approche du Lexique-grammaire est essentiellement syntaxique, elle repose sur l'hypothèse qu'on ne peut formuler de règles de grammaire sans une description exhaustive des constructions et des distributions de tous les items lexicaux concernés. Ces travaux sont une application au français du système transformationnel de Zellig S. Harris (1971, 1990). Les citations extraites de Maurice Gross (1975) et de Jean-Paul Boons *et al.* (1976) sont notées MG et BGL respectivement.

Maurice Gross et ses collaborateurs ont commencé à développer le Lexique-grammaire au milieu des années 70 :

> « Notre intention a été d'aboutir à une description exhaustive du français par rapport aux propriétés que nous avons choisies et par rapport à un lexique donné. » MG page 21.

Les premières constructions étudiées ont été les phrases simples, c'est-à-dire avec un verbe (sémantiquement plein), puis les formes nominales et adjectivales, et enfin les

constructions figées. Le lexique grammaire des verbes est la classification de tous les verbes simples du français (environ 5 000 verbes divisés en 15 000 emplois) répartis en une soixantaine de tables qui répertorient leurs propriétés syntaxiques et distributionnelles (cf. Gross 1975, Boons *et al*. 1976, Guillet et Leclère 1992, entre autres).

Les items lexicaux sont regroupés en *classes* ou *tables*, selon les structures syntaxiques élémentaires qui les caractérisent, appelées « constructions (ou phrases) définitionnelles ». L'ensemble des verbes qui acceptent cette propriété constitue la table. Chaque table a un code numérique comme les tables de verbes à complétives numérotées de 1 à 19, ou les tables de constructions transitives locatives qui commencent par les numéros 36, 37 et 38. À chaque classe est associé un ensemble de propriétés pertinentes, distributionnelles et concernant la sélection des arguments (par exemple, le sujet est-il un substantif humain, non humain ou indifférencié ?) ou transformationnelles (par exemple, le verbe est-il passivable ?, peut-on extraposer le sujet phrastique ?). Ainsi, la table 36DT a pour phrase définitionnelle N_0 *V* N_1 *à* N_2 avec pour N_0 et N_2 des substantifs humains (N_0 est le sujet, *V* le verbe, N_1 et N_2. respectivement le 1er et le 2ème complément). Elle décrit un échange de N_1 entre deux humains N_0 et N_2. Tous les verbes (210) de cette table (*offrir, rembourser, vendre, voler*, etc.) ont cette caractéristique tel le verbe *voler* par exemple, car la phrase *Jean a volé 10 euros à Marie* est acceptable. Cette table a 22 propriétés, comme la possibilité (ou non) pour un verbe d'avoir la construction "croisée" N_0 *V* N_2 *de* N_1 : *Jean a volé Marie de 10 euros* (dont le sens est voisin mais non exactement synonyme de la construction définitionnelle).

Les tables sont représentées sous forme de matrice. Les lignes correspondent aux verbes et les colonnes à des propriétés syntaxiques :

> « à l'intersection de chaque rangée (verbe) et de chaque colonne (propriété syntaxique) d'une table figure obligatoirement une marque représentant une estimation de la grammaticalité, i.e. de la bonne ou mauvaise formation de la structure syntaxique (en gros la colonne) lorsqu'elle est complétée par l'insertion du verbe. » BGL page 50.

Cette marque est un signe + qui indique que le verbe a la propriété, et un signe – correspondant au cas contraire. Pour déterminer si une construction est grammaticale ou non, un jugement d'acceptabilité est utilisé. Mais il est important de noter que dans le cadre du Lexique grammaire, ce qui n'est pas « inacceptable » est « acceptable », ce qui revient à considérer comme grammaticaux les emplois douteux ou difficiles :

> « Nous avons donc opéré une première approximation en classant les formes de manière binaire : soit comme acceptables, soit comme inacceptables. En conséquence, ces classes comporteront des informations douteuses. » MG page 23.

> « Les exemples cruciaux étant de manière générale les séquences inacceptables, une première tactique a été, en cas d'hésitation sur l'entrée d'un élément dans une structure et d'absence d'une argumentation de niveau supérieur à celui du jugement empirique ponctuel, de considérer provisoirement cette structure comme bien formée. En effet, le risque de considérer comme mal

formée une structure qui ne l'est pas est pour nous plus grave que le risque inverse, inacceptable pour le puriste, d'accepter une structure « non recommandable ». BGL page 47.

« Dans l'ensemble donc, nous avons considéré comme acceptables des formes ressenties comme douteuses ou même comme inacceptables par de nombreux locuteurs ». MG page 24.

La méthode pour déterminer l'acceptabilité d'une phrase est l'introspection, avec parfois l'usage de dictionnaires :

« Le français est un français « standard », révélé essentiellement par introspection, et vérifié, pour certains exemples douteux, par enquête auprès de locuteurs d'appartenance socio-linguistiques voisines ; les grammaires et ouvrages de linguistique traditionnels ainsi que les dictionnaires nous ont également fournies de nombreux exemples. » MG page 21.

Dans le Lexique-grammaire, les propriétés ne sont pas « autonomes », elles doivent être prise en compte en perspective avec la propriété définitionnelles de la table:

« [...] en fait , les phrases construites sur les structures servant d'intitulés de colonnes ne sont jugées que par comparaison avec les structures définitionnelles de la table. » BGL page 50.

Les principales conventions adoptées pour la notation sont les suivantes :

« Les formes de base comportent des indices numériques qui permettent de repérer les différents groupes nominaux : N_0 *V* N_1 *Prép* N_2 . Les indices ont été déterminés à partir de propriétés syntaxiques. N_0 est toujours un sujet formel. Lorsqu'un verbe accepte deux compléments, leur ordre est défini au moyen de leur propriété de présence : facultative ou obligatoire.

[…] Rappelons que nous utilisons le symbole Ω pour nous référer au(x) complément(s) sans distinction de forme.[...]» MG pages 13, 15.

Les N_i peuvent avoir pour indice des symboles sémantico-syntaxiques qui donnent des indications sur le contenu du groupe nominal. Par exemple, N_i = *Nhum* indique que la classe du groupe nominal du *i*ème complément (ou du sujet si $i = 0$) est celle des substantifs humains, N_i = *N-hum* (ou *-Nhum*) que cette classe est celle des substantifs non humains, N_i = *Npc* caractérise une partie du corps, N_i = *Nnr* : indique que la position syntaxique *i* peut recevoir un substantif d'une nature sémantique quelconque, une complétive ou une infinitive, c'est-à-dire à la fois un nom propre, un nom commun humain ou abstrait, un verbe à l'infinitif ou une phrase ; on dit que le sujet ou le *i*ème complément est « non restreint ». *P* représente une phrase, et *Qu P* une complétive sans distinction de contenu. Une construction composée d'un verbe à l'infinitif et de ses compléments éventuels est notée *V Ω*. Les suffixes *–inf* (infinitif) ou *–ant* (participe présent) sont parfois utilisés, quand il est nécessaire de distinguer les deux formes. Les indices numériques des verbes *V* sont utilisés pour localiser la position du sujet de l'infinitive : V^0 *Ω*, V^1 *Ω*, etc.

Un verbe est noté *U* dans le cas où il s'agit d'un verbe qui peut ne pas sélectionner son sujet, comme *commencer* ou *continuer* :

« Un verbe *U* apparaissant dans une phrase de forme N_0 *U Prép* V^0*-inf* Ω peut ne pas imposer de contraintes sur N_0. Autrement dit, étant donnée une phrase quelconque N_0 V^0 Ω, il est possible d'y insérer un *U* sans que les relations entre N_0 et V^0 en soient affectées. » MG page 160.

Les parenthèses contenant plusieurs éléments séparés par le signe + indiquent une possibilité de choix entre ceux-ci. *E* est l'élément neutre par rapport à la concaténation, c'est-à-dire une séquence zéro. Ainsi, $N_0\ N_1 V\ (E + N_1)$ représente les deux constructions $N_0\ V$ et $N_0 V\ N_1$. Des transformations appliquées à une phrase apparaissent entre crochets, par exemple [passif].

L'organisation générale du Lexique-grammaire des verbes du français est donnée dans le tableau 1.

Constructions	
$(N_0$+*Qu P)* V *(E+Prép)*$(N_1$+ *Qu P) (E+ Prép* $(N_2$+*Qu P))*	Tables 1-19
$N_0\ V\ (E+\ N_1)\ (E$+*Prép* $N_2)$ → $N_0\ V\ (E$+*Prép N1)*	Tables 31, 33-35 (et 2,7,8)
→ $N_0\ V\ N_1$	Tables 32
→ $N_0\ V\ N1$ *Prép* $_2$	Tables 36-38

Tableau 1. Organisation générale des tables du Lexique grammaire des verbes

Maurice Gross a étudié les constructions complétives qu'il a réparties dans 19 tables. Le tableau 2 indique, pour chacune, le numéro, la définition de la table et une ou plusieurs phrases représentatives.

Numéro et définition des tables	**Exemples de phrases**
Table 1 *(*N_0 *U Prép* V^0 Ω*)*.	Pierre se dépêche d'amuser Paul
Table 2 *(*N_0 *V (E+* N_1 *)* V^0 Ω*)* verbes dits de « mouvement »	Il court voir Marie, Il franchit la rivière retrouver Marie
Table 3 *(*N_0 *V* N_1 V^1 Ω*)* causatifs de mouvement excepté *laisser*.	Paul envoie la balle rebondir contre le mur
Table 4 *(Qu P V* N_1*)*.	Que Luc vienne ennuie Paul.
Table 5 *(Qu P V Prép* N_1*)*. Sujet extraposable.	Que Paul ait pu faire cela dépasse mon imagination (cela dépasse mon imagination que)
Table 6 *(*N_0 *V Qu P)*	Je vois que Paul travaille
Table 7 *(*N_0 *V à ce Qu P)*	Paul tient à ce que tout soit en ordre

Numéro et définition des tables	Exemples de phrases
Table 8 (*N_0 V de ce Qu P)*	Paul est content de ce qu'il soit venu.
Table 9 *(N_0 V Qu P à N_2)*	J'ai démontré à Pierre que Paul était venu
Table 10 *(N_0 V Qu P Prép N_2).*	Paul a ancré dans son esprit que tu avais fait cela
Table 11 *(N_0 V N_1 à ce Qu P)*	Pierre accoutume Paul à ce qu'il vienne au cours
Table 12 *(N_0 V N_1 de V^1 Ω).*	Paul déteste Marie d'avoir fait cela.
Table 13 *(N_0 V N_1 de ce Qu P).*	Cette loi assure Max de ce que ses droits seront respectés
Table 14 *(N_0 V à ce Qu P Prép N_2).*	Max a collaboré avec Luc à ce que tout soit en ordre.
Table 15 *(N_0 V de ce Qu P Prép N_2).*	Paul s'explique de cela auprès de Marie.
Table 16 *(N_0 V Prép$_1$ N_1 Prép$_2$ N_2)*	Max conclut que le résultat est correct de ce que les deux calculs concordent.
Table 17 *(Il V (E + Prép$_1$ ce) Qu P (E + Prép$_2$ N_2)).*	Il semble à Max que Luc est parti
Table 18 *(N_0 V Prép$_1$ N_1 Prép$_2$ N_2 (E + pour) que P subj)*	Max a reçu de Luc comme garantie qu'il aurait une prime
Table 19 (Qu P V N_1Prép N_2).	Qu'il parte discrédite Max auprès de Luc

Tableau 2. Tables des verbes à constructions complétives

1.1.2. Les verbes de sentiment et psychologiques

Les verbes qui relèvent du champ sémantique qui nous intéresse (les sentiments et états psychologiques) sont contenus principalement dans deux classes : la table 4, de propriété définitionnelle *Qu P V N_1*, c'est-à-dire la possibilité d'un sujet phrastique, dans laquelle sont les verbes de type *irriter* comme dans *Que Luc vienne irrite Paul*, et la table 12 définie par la construction *N_0 V N_1 de V^1 Ω,* par exemple *Paul déteste Marie d'avoir fait cela*. Un exemple extrait de la table 4 est donné dans le tableau 3. Notre description des verbes de sentiment a été établie à partir de ces deux classes (voir plus loin §2.1.1.).

N0= :¨Nhum	N0= :¨Nnr	N0= :¨le fait Qu P	N0= :¨V1-Ω	**verbe**	V concret	N0 V	Adjectif = ant	Adjectif = able	Adjectif = eux	Adjectif =(E+at)eur	N1= :¨ Nhum	N1= :¨-Nhum	Le fait Qu P	N1 se V de ce que P	N1 se V auprès de Nhum de ce que P	N1 est Vpp de ce que P	[passif par]	[passif de]	N0 V N1 contre Nhum
+	+	+	+	calmer	+	+	+	-	-	-	+	-	-	-	-	-	+	-	-
-	+	+	+	captiver	-	+	+	-	-	-	+	-	-	-	-	+	+	-	-
+	+	+	+	cataloguer	+	+	-	-	-	-	+	-	+	-	-	-	+	-	-
+	+	+	+	chagriner	-	+	+	-	-	-	+	+	-	+	-	+	+	-	-
-	+	+	+	chambouler	+	+	-	-	-	-	+	+	-	-	-	+	+	-	-
+	+	+	+	charmer	-	+	-	-	-	-	+	+	-	-	-	+	+	-	-
+	+	+	+	charpenter	+	+	-	-	-	-	+	+	-	-	-	-	+	-	-
+	+	+	+	chatouiller	+	+	-	-	-	-	+	+	-	-	-	-	+	-	-
-	+	+	+	chavirer	+	+	+	-	-	-	+	+	-	-	-	+	+	-	-
-	+	+	+	chiffonner	+	+	+	-	-	-	+	+	-	+	-	+	+	-	-
-	+	+	+	choquer	-	+	+	-	-	-	+	+	-	+	-	+	+	-	-
+	+	+	+	cingler	+	+	-	-	-	-	+	+	-	-	-	-	+	-	-
+	+	+	+	civiliser	-	-	-	-	-	+	+	+	-	-	-	-	+	-	-
+	+	+	+	claquer	-	+	+	-	-	-	+	-	-	-	-	-	+	-	-
-	+	+	+	classer	+	+	+	-	-	-	+	-	-	-	-	-	+	-	-
+	+	+	+	coincer	+	+	-	-	-	-	+	+	-	-	-	-	+	-	+
+	+	+	+	combler	+	+	-	-	-	-	+	+	-	-	-	-	+	-	-
-	+	+	+	commotionner	-	+	+	-	-	-	+	-	-	-	-	-	+	-	-
+	+	+	+	compléter	-	+	-	-	-	-	+	+	-	-	-	-	+	-	-
+	+	+	+	compromettre	-	+	+	-	-	-	+	+	-	-	-	-	+	-	-

Tableau 3. Extrait de la table 4 du Lexique grammaire (Gross 1975)

1.1.3. Les tables du lexique grammaire comme ressource en TAL

Les tables du Lexique grammaire n'ont pas été élaborées dans l'optique du Traitement automatique de la langue, mais leur formalisation en matrices rend possible leur emploi par un logiciel. Selon Maurice Gross, cette description en classes de verbes partageant une même construction définitionnelle et acceptant (ou non) des propriétés de structure permet une approche nouvelle pour l'analyse syntaxique :

> « Une éventuelle analyse automatique prendrait alors une forme nouvelle. [...]. La procédure suggérée par notre grammaire est différente. Des consultations de dictionnaires sont effectuées pour les verbes de la phrase à analyser, elles déterminent des ensembles de contextes possibles pour chaque verbe. L'exploration des contextes réels (e.g. l'examen des substantifs et de leurs relations avec les verbes), permettent d'effectuer un choix parmi les contextes possibles et donc d'éliminer ceux qui sont incorrects. » MG p.221

Néanmoins, leur utilisation est complexe, car comme nous l'avons indiqué, la valeur binaire (+,-) des propriétés autorise des acceptabilités parfois douteuses. De plus, les tables contiennent beaucoup d'informations très détaillées, mais compactes c'est-à-dire dont une grande partie n'est compréhensible qu'avec des connaissances implicites et il n'y a pas d'indication dans la description sur la procédure qui a permis de les construire.

À l'intérieur d'une table, nous avons vu qu'une propriété est considérée comme acceptée s'il existe au moins un emploi possible du verbe dans la construction. Mais aucune précision n'est donnée sur cet emploi. Par exemple, dans la table 4, si nous considérons la possibilité du passif introduit par la préposition « de », notée N_1 *est Vpp de* N_0, que nous avons également étudiée, nous montrerons au §2.1.2.4, que l'acceptabilité de ce passif est contrainte par la nature du N_1. Ainsi, pour le verbe *déconcerter* par exemple, si le sujet (à l'actif) est un substantif strictement humain, ce passif n'est pas possible car si les phrases *Marie/Le comportement de Marie déconcerte Luc* et *Luc est déconcerté du comportement de Marie* sont acceptables, **Luc est déconcerté de Marie* ne l'est pas. Par ailleurs si le sujet est une infinitive, ce passif est toujours possible pour tous les verbes de la table 4, ce qui est dit dans (Gross 1975) mais pas indiqué dans la table.

Donc, malgré l'acceptabilité d'un emploi indiquée par la propriété définitionnelle ou par un « + » pour toute construction de la table, l'emploi n'est pas toujours possible car il est contraint par la nature syntactico-sémantique des arguments. Pour des applications de TAL, en génération automatique par exemple, cette caractérisation des arguments est cruciale.

Malgré la complexité d'emploi des tables du lexique grammaire, plusieurs travaux les utilisent comme ressource lexicale, après les avoir transformées et reformatées selon leurs besoins spécifiques. C'est le cas en particulier de Laurence Danlos et ses collaborateurs, dans leur projet de constitution d'un lexique morphologique et syntaxique du français. Ce lexique électronique, baptisé « L*efff* » (Lexique des Formes Fléchies du Français), est en cours d'élaboration depuis 2004 (Danlos 2007). Il a été construit initialement avec des

informations sur les formes verbales induites principalement à partir de corpus, il est enrichi progressivement avec l'ajout d'autres catégories et des informations syntaxiques, à partir des données du lexique grammaire. L'objectif est de permettre son utilisation dans les applications dans le domaine du TAL, quel que soit le cadre théorique adopté.

Un lexique syntaxique, « Synlex », a également été développé par Claire Gardent *et al.* (2006), Falk *et al.* (2007). Il a été construit de façon semi-automatique à partir d'un extrait des tables de verbes du lexique grammaire disponibles et complété manuellement :

> « Seules 60% des tables du LADL étant disponibles, nous avons complété manuellement le lexique extrait des tables disponibles avec environ 2 000 verbes et leurs cadres de base. » Gardent et al. 2006 page 3.

Ils ont également proposé une méthode de conversion des tables du LADL en un format approprié pour les applications du TAL.

1.2. L'ÉTUDE DES VERBES DE BETH LEVIN

Beth Levin (1993) a examiné 3 000 verbes anglais et a montré comment certains modèles syntaxiques sont parallèles à certains modèles sémantiques. Elle part de l'hypothèse que le comportement d'un verbe, et particulièrement l'expression et l'interprétation de ses arguments, est guidé par son sens. Les citations sont extraites de Beth Levin (1993), noté BL dans la suite.

1.2.1. Les alternances

Elle a tout d'abord examiné un large ensemble d'alternances syntaxiques, avec parfois l'indication sémantique des verbes qui acceptent l'alternance en question. Les alternances sont nommées, décrites en texte libre, puis exemplifiées avec des phrases acceptables ou non. Elles portent principalement :

1) sur la modification de la transitivité du verbe : transitif <-> intransitif. Il y a dix alternances principales :

- avec l'objet du verbe qui devient le sujet de la construction intransitive et alternance des rôles sémantiques (celui de l'objet de la forme transitive devient celui du sujet de la forme intransitive) : *The butcher cuts the meat* ↔ *The meats cuts easily* (6 alternances) ;

- avec conservation du même sujet du verbe *Mike ate the cake* ↔ *Mike ate* (1 alternance) et l'objet non exprimé (9 alternances) ;

- alternances conatives ; L'objet du verbe dans la variante transitive devient l'objet de la préposition dans un syntagme préposionnel dont la tête est *at* (et quelquefois d'autres prépositions comme *against, on*). *I pushed the table* ↔ *I pushed at/on/against the table* (1 alternance) ;

- alternances causatives: impliquent des verbes avec des usages transitifs et intransitifs, avec l'emploi transitif grossièrement paraphrasé en « cause to V-intransif ». deux alternances principales : l'alternance causative/inchoative et alternance action induite *Jane broke the cup, the cup broke.* (3 alternances) ;

- alternances avec suppression de la préposition : la forme transitive apparaît dérivée de la forme intransitive avec suppression de la préposition *They skated along the canals* ↔ *They skated the canals* (2 alternances) ;

- alternances substance/source : avec les verbes d'émission de substances *Heat radiates from the sun , the sun radiates heat* (1 alternance) ;

2) pour les verbes transitifs, sur les arguments à l'intérieur du syntagme verbal VP. Il y a quatorze alternances principales :

- l'alternance dative : *Bill sold a car to Tom* ↔ *Bill sold Tom a car* ;

- l'alternance bénéfactive: *Martha carved a toy for the baby ↔ Martha carved the baby a toy* ;

- l'alternance locative : *Jack sprayed paint on the wall ↔ Jack sprayed the wall with paint* (5 alternances) ;

- les alternances de création et de transformation : *Martha carved a toy out of the piece of wood ↔ Martha carved the piece of wood into a toy* (4 alternances);

- les alternances réciproques : *the car collided with the bicycle ↔ the car and the bicycle collided* (6 alternances) ;

- l'alternance de satisfaction (fulfilling) : *the judge presented a prize to the winner ↔ The judge presented the winner with a prize* ;

- l'alternance d'impression : *the jeweller inscribed the name on the ring ↔ the jeweller inscribed the ring with the name* ;

- l'alternance with/against : *Brian hit the stick against the fence ↔ Brian hit the fence with the stick* ;

- l'alternance *through/with* : *Alison pierced the needle through the cloth ↔ Alison pierced the cloth with a needle* ;

- l'alternance *"blame"* : *Mira blamed the accident on Terry ↔ Mira blamed Terry for the accident* ;

- l'alternance *"search" : Melissa searched the papers for a clue ↔ Melissa searched for a clue in the papers* ;

- l'alternance de mise en relief du possesseur de la partie du corps : *Selina touched the horse on the back ↔ Selina touched the horse's back* ;

- l'alternance de mise en avant du possesseur ou de l'attribut : *I admire his courage ↔ I admire him for his courage* ;

- l'alternance "as": *The president appointed Smith press secretary ↔ The president appointed Smith as press secretary* ;

3) les alternances sujet "oblique" avec changement du nombre de syntagmes nominaux associés au verbe : il y a un syntagme nominal de moins dans une forme que dans l'autre. Les dix principales alternances sont :

- l'alternance avec le temps en sujet : *The world saw the beginning of a new era in 1492 ↔ 1492 the beginning of a new era* ;

- l'alternance avec une force naturelle en sujet : *I dried the clothes in the sun ↔ The sun dried the clothes* ;

- l'alternance avec un instrument en sujet : *David broke the window with a hammer ↔ The hammer broke the window* ;

- l'alternance avec une cause abstraite en sujet : *He established his innocence with the letter ↔ The letter established his innocence* ;

- l'alternance où l'objet localisé devient sujet (oblique) : *I filled the pail with water* ↔ *Water filled the pail* ;

- l'alternance où le lieu est sujet (oblique) : *We sleep five people in each room* ↔ *Each room sleeps five people* ;

- l'alternance avec un contenant en sujet : *I incorporated the new results into the paper* ↔ *the paper incorporated the new results* ;

- l'alternance avec un matériel brut en sujet : *She baked wonderful bread from that whole wheat flour* ↔ *That whole wheat flour bakes wonderful bread* ;

- l'alternance avec une somme d'argent en sujet : *I bougth (you) a ticket for $5* ↔ *$5 will buy (you) a ticket* ;

- l'alternance avec une source en sujet : *The middle class will benefit from the new tax laws* ↔ *The new tax laws will benefit the middle class* ;

4) la réflexivité (2 alternances) ;

5) les passifs (4 alternances) ;

6) les alternances qui impliquent des sujets « postverbaux » (2 alternances) ;

7) d'autres constructions (9 alternances) ;

8) et enfin les verbes sans alternances, qui demandent des diathèses particulières, par exemple le passif obligatoire *It is rumored that he left town /**[1]*They rumored that he left town* (6 alternances) ;

1.2.2. Les 49 classes de Beth Levin

Dans une seconde étape de son étude, Beth Levin a ensuite identifié 49 classes de verbes cohérentes sémantiquement, comme les verbes de contact, de perception, de jugement, etc. La liste complète de ces classes est donnée dans le tableau 4. Chaque case du tableau décrit la sémantique des verbes de la classe.

Les verbes de chaque classe ont un comportement semblable par rapport à un sous-ensemble des alternances étudiées précédemment. Ce sous-ensemble est constitué de propriétés caractéristiques acceptées et non acceptées, exemplifiées par des phrases avec un verbe considéré comme représentatif de la classe. Beth Levin associe également à chaque classe un ensemble de propriétés partagées par tous ou quelques verbes de la classe :

> « Although any single property of a class of verbs will in itself not be very informative, the conjunction of properties shown by a class of verbs may well be more revealing , since it will reflect the entire set of meaning components shared by the class members. […]

[1] Le caractère « * » devant une phrase indique qu'elle est inacceptable.

Verbs classes arise because a set of verbs with one or more shared meaning components show similar behavior. » BL page 17.

« A representative verb is then used to exemplify the characteristic properties of the class, including argument-taking properties, behavior with respect to diathesis alternations, and occasionally morphological properties (related nominals or adjectives). These properties have been chosen to illustrate what is distinctive to the members of that class. » BL page 22.

Placement	Enlever	Envoi et portage	Exercer une force	Changement de possession
Apprentissage	Attraper et garder	Dissimulation	Lancer	Contact par impact
Contact avec un objet pointu	Contact	Coupure	Mélanger, assembler et attacher	Séparation et désassemblage
Coloration	Création d'image	Création et transformation	Engendrer	Naissance d'animaux
Perception	États psychologiques	Désir	Jugements	Appréciation
Recherche	Interaction social	Communication	Sons produits par des animaux	Ingestion
Impliquant le corps	Toilette, soin du corps	Meurtre	Émission	Destruction
Changement d'état	Logement	Existence	Apparition, disparition	Mouvement interne du corps
Position spatiale	Mouvement	Évitement	Raccourcissement ou prolongation de la durée d'un événement	Mesure
Aspectuels	Weekend	Saison	Avec des compléments prédicatifs	

Tableau 4. Classes des verbes anglais par Beth Levin (1993)

Un verbe qui a plusieurs sens apparaît dans plusieurs classes, par exemple *arrange* est dans la classe « Put verbs » et dans la classe « Build verbs ».

1.2.3. Les verbes de sentiment et psychologiques

Les verbes qui relèvent du champ sémantique qui nous intéresse (les sentiments et les états psychologiques) sont contenus dans deux classes : la classe des verbes d'États psychologiques, qu'elle nomme Psych-Verbs, et la classe des verbes de Désir. La classe des verbes psychologiques est sous divisée en 4 classes : le type *amuse,* caractérisé par la position objet de la personne qui éprouve le sentiment, appelée l'expérienceur, (221 verbes), le type *admire* dont l'expérienceur est en position sujet, avec indication de la polarité (45 verbes : 26 positifs et 19 négatifs), le type *marvel* qui contient des verbes intransitifs avec une préposition et dont l'expérienceur est le sujet, comme *marvel at* (s'émerveiller de) ou *groove on* (s'amuser) (79 verbes) et le type *appeal* qui sont des verbes intransitifs avec une préposition et dont l'expérienceur est le complément, comme *appeal to* (plaire à) (5 verbes).

Pour chacune de ces classes, et pour chaque type, nous allons décrire les propriétés qu'elle a considérées. Les intitulés de propriétés, les commentaires et les exemples anglais sont extraits de Levin (1993). Nous n'avons pas repris la numérotation de Beth Levin, qui est une numérotation séquentielle des exemples et des propriétés. Nous avons numéroté les propriétés avec pour chacune un numéro unique.

Pour les verbes psychologiques, avec l'expérienceur en position objet, elle a étudié 15 propriétés pour les verbes du type *amuse* (amuser), et 2 pour les verbes comme *appeal*.

Les propriétés pour le type *amuse* sont :

1. The clown amused the children (construction de base)

2. Causative alternations (most verbs):

The clown amused the children

**The children amused (at the clown)*

3. Middle alternation (most verbs):

The clown amused the little children

little children amuse easily

4. PRO-Arb Object alternation

That joke never fails to amuse little children

That joke never fails to amuse

5. Possessor Subject Possessor-Attribute Factoring Alternation (transitive)

The clown amused the children with his antics

The clown's antics amused the children

9. The clown was amusing to the children (most verbs)

10. an amusing joke

12. Extraposition of sentential complements

That the clown had a red nose amused the children

It amused the children that the clown had a red nose

14. Choice of prepositions in the "passive" depends on the verb

The children were amused at/by/with the clown

15. Resultative Phrase : *That movie bored me silly/to tears*

17. Derived nominals has “passive” interpretation only

The children’s amusement(at the clown)

**The clown’s amusement of the children*

19.*-er* Nominal (some verbs) : *Baffler, disturber*, …

20*.-able* Adjectives : **amusable, *surprisable, excitable, irritable, upsettable*

La classe *Appeal* contient 5 verbes intransitifs. L’expérienceur est en position objet avec une préposition: *niggle at, grate on, jar on, appeal on, matter on.* Les propriétés sont :

1. This painting appeals to Malinda (construction de base)

13. *Passive

**Malinda is appealed by this painting*

Beth Levin a classé les verbes psychologiques avec l’expérienceur en position sujet, en deux sous-classes : les verbes de type *admire*, divisés en positifs et négatifs, caractérisés par 9 propriétés, et les verbes de type *marvel* avec deux propriétés.

Les propriétés pour le type *admire* sont :

1. The tourists admired the paintings (construction de base)

3.*Middle alternation

6. Possessor Object Possessor-Attribute Factoring Alternation

I admire his honesty

I admire him for his honesty

7. Attribute Object Possessor-Attribute Factoring Alternation

I admire his honesty

I admire the honesty in him

11. Sentential Complements (some verbs)

The children liked that the clown had a red nose

12. Extraposition of sentential complements (some verbs)

16. Derived nominals has “active” interpretation only

The children’s enjoyment of the movie

* *The movie's enjoyment by the children*

18. **As* alternation

I admire him as a teacher

**I admire him a teacher*

19.-*er* Nominal

20.-*able* Adjectives : *detestable, enjoyable, hatable, likable*

Les verbes comme *marvel* sont des verbes intransitifs avec une préposition. Les propriétés étudiées sont :

1. *Megan marveled at the beauty of the Grand Canyon* (construction de base)

13. Passive (some verbs)

La deuxième classe, les verbes de désir**,** est constituée de deux sous classes : les verbes transitifs comme *want* et les intransitifs, de type *long*. Comme le remarque Beth Levin :

> « These verbs are sometimes included among the psych-verb that take an experiencer subject, that is, the admire-type and the marvel-type psych-verbs ». BL page 194.

Les propriétés des verbes de type *want* sont :

1. *Dorothy needs new shoes* (construction de base)

6. Possessor Object Possessor-Attribute Factoring Alternation

7. *Attribute Object Possessor-Attribute Factoring Alternation

8. **Dorothy is needing new shoes*

13. ?Passive

18. **As* alternation

Les verbes de type *long* ont une préposition obligatoire comme *long for* or *lust after/for*. Les propriétés considérées sont :

1. *Dana longs for a sunny day* (construction de base)

8. *Dana is longing for a sunny day*

13. ?Passive

1.2.4. La classification de Beth Levin comme ressource en TAL

Les classes sémantiques de verbes contiennent une description très détaillée des relations syntactico-sémantiques entre un verbe et ses arguments, ce qui en fait une

ressource intéressante pour des applications informatiques. Mais cette étude a été faite dans une optique purement linguistique sans aucun objectif de traitement automatique, il n'est donc pas étonnant que l'on ne puisse pas l'utiliser telle quelle comme ressource informatique.

Un problème majeur est que cette description n'est pas homogène à divers titres. Nous énumérons plusieurs exemples de cette hétérogénéité (cette liste n'est pas exhaustive).

- Le format de l'énoncé des propriétés est multiple. En effet, Beth Levin décrit une propriété soit par un énoncé, par exemple « *Middle alternation* », soit par un couple de phrases qui montre l'alternance entre la construction de base et la propriété, comme « *The clowns amused the little children/ little children amused easily* », soit uniquement par une phrase qui l'exemplifie comme « *an amusing joke* ».

- La description est faite au niveau de la classe, et exemplifiée avec un verbe « représentatif », mais aucune indication n'est mentionnée ou pas assez précise pour pouvoir généraliser à tous les verbes. Par exemple, la propriété est supposée être acceptée (ou non) par tous les verbes de la classe, mais quelquefois elle ne l'est que par la plupart : dans ce cas la mention « *most verbs* » est indiquée derrière la propriété, sans aucune autre indication des verbes concernés, comme « *Middle alternation (most verbs)*».

De même certaine listes données sont partielles, par exemple : « *-able Adjectives : *amusable, *surprisable, excitable, irritable, upsettable ».* Qu'en déduire pour la possibilité de formation d'adjectifs à partir des autres verbes de la classe ? Quand une propriété est indiquée explicitement, elle ne l'est pas toujours de façon précise comme « *choice of prepositions in the « passive » depends on the verb : The children were amused at/by/with the clowns* », on comprend que pour le verbe *amuse* les trois constructions sont possibles, mais, là encore, qu'en est il des autres verbes de la classe ?

- La nature des arguments n'est pas prise en compte. Ainsi les propriétés « *Possessor Object Possessor-Attribute Factoring Alternation* » et « *Attribute Object Possessor-Attribute Factoring Alternation* » sont exemplifiées par « *I admire his honesty »* / *« I admire him for his honest* » et /« *I admire the honesty in him »* respectivement. Mais si le complément est d'une nature différente ces deux dernières constructions peuvent devenir difficilement acceptables comme le montrent les exemples suivants : « *I admire his haircut* », *« ?I admire him for his haircut* » et « **I admire the haircut in him* ». Cette caractérisation des arguments, complexe, est importante pour le TAL, dans une optique de génération par exemple.

- Certaines propriétés sont implicites. Par exemple, le passif est supposé être accepté par tous les verbes. Cette construction n'est indiquée que si elle est peu (classe *Appeal* et *Desire*) ou pas possible (classe *Marvel*). On pourrait la déduire de la propriété déjà évoquée « *choice of prepositions in the « passive » depends on the verb* » pour les verbes comme *amuse,* mais pas pour les verbes comme *admire* (elle est implicite dans cette classe).

- Enfin, de nombreuses propriétés sont indiquées dans le commentaire de la classe comme la position syntaxique de l'expérienceur et de l'objet (ou la cause) du sentiment ou la transitivité des verbes; par exemple, pour les Psych-verbs :

> « They are transitive verbs whose object is the experiencer of the emotion and whose subject is the cause of the change in psychological state. » BL page 191.

Pour un usage en TAL de ce travail, il serait donc nécessaire de le compléter, pour avoir des informations sur **chaque** verbe, et de le reformuler. L'indication de l'existence (ou non) de propriétés de plusieurs façons différentes implique qu'il faut les écrire dans une forme unifiée et formalisée, les propriétés implicites doivent être exprimées, etc. Une classification d'un sous-ensemble des verbes du français a été faite par Patrick Saint-Dizier (1996) en s'appuyant sur l'étude de Beth Levin, mais dans une optique de TAL.

1.3. WORDNET

La base de données WordNet offre une large couverture lexicale de l'anglais. Elle est très utilisée car c'est une ressource à la fois libre et assez bien documentée. Son développement manuel a commencé en 1985, puis une première version électronique à été construite en 1993 jusqu'à Wordnet 2.1 en 2005. La base est en constante évolution. (Miller 1990, Miller et Fellbaum 1991, Fellbaum 1998, 1999). C'est un réseau lexical qui couvre la grande majorité des noms, verbes, adjectifs et adverbes, organisés en réseaux sémantiques distincts. Son contenu est indiqué dans le tableau 5.

Réseau	Formes	Synsets	Paires mot-sens
Noms	117798	82115	146312
Verbes	11529	13767	25047
Adjectifs	21479	18156	30002
Adverbes	4481	3621	5580
TOTAL	155287	117659	206941

Tableau 5. Contenu du réseau lexical WordNet (juillet 2008)

Sa conception est basée sur la notion de réseau sémantique organisé de façon hiérarchique, en utilisant la relation d'inclusion qui lie, par exemple, des triplets comme « animal », « poisson », et « saumon ». Les unités dans WordNet sont des ensembles de synonymes appelés « synsets ». Ce sont des groupes de mots ou de phrases qui expriment un même concept. Christiane Fellbaum a étudié et décrit le lexique des verbes. Nous la citons (1999, noté CF dans la suite):

> « Des différences de sens entre les membres d'un synset se montrent dans différentes restrictions de sélection. Par exemple, « rise » (monter) et « fall » (tomber / descendre) peuvent choisir comme argument des entités abstraites comme « temperature » (température) et « prices » (prix), mais leurs proches synonymes « ascend » (monter) et «descend » (descendre) ne peuvent pas sélectionner ces noms. D'ailleurs, tous les proches synonymes n'ont pas les mêmes propriétés syntaxiques : on peut « give », « hold, » ou « throw» « a party » (organiser une soirée), mais le déplacement du datif n'est possible qu'avec « give » :
>
> *(1) We gave /threw/ held a party for John.*
>
> *(2) We gave / *threw / *held John a party. Nous avons organisé une soirée pour Jean.*
>
> Les synonymes ont été choisis largement en examinant les dictionnaires traditionnels, où la synonymie est une façon habituelle de définir le sens des mots.

Les synsets sont reliés par des arcs étiquetés qui représentent des relations lexicales et sémantiques. Les relations lexicales lient des mots tandis que les relations sémantiques (ou conceptuelles) lient des synsets. WordNet n'a que des relations paradigmatiques, de sorte que les quatre catégories lexicales principales (N, V, Adj, Adv) sont traitées séparément. » CF pages 27-28.

À chaque synset est associé une description du sens qu'il représente et des exemples d'usage. Les synsets n'expliquent pas ce que sont les concepts, ils en posent l'existence : :

« These synonym sets (synsets) do not explain what the concepts are; they merely signify that the concepts exist. People who know English are assumed to have already acquired the concepts, and are expected to recognize them from the words listed in the synset. » Miller *et al*., 1993, pages 5-6.

1.3.1. Organisation du lexique

Deux types de relations sont définis dans le lexique, les « relations lexicales » entre les formes (les mots), et les « relations sémantiques » entre les sens (les synsets).

Les relations lexicales sont des relations de synonymie, d'antonymie et des relations morphologiques.

Les relations sémantiques sont des relations hiérarchiques qui relient les synsets entre eux. Ces relations dépendent de la catégorie des mots contenus dans les synsets.

1.3.1.1. Les noms

Les synsets de noms sont organisés de façon hiérarchique, à partir d'une racine unique (nommée Entity), puis 25 nœuds fils décrivant des champs sémantiques différents : (Act/action/activity), (Animal/fauna), (Artifact), (Attribute, Property), (Body, Corpus), (Cognition, Knowledge), (Communication), (Event, Happening), (Feeling, Emotion), (Food), (Group, Collection), (Location, Place), (Motive), (Natural Object), (Natural Phenomenon), (Plant, Flora), (Possession), (Process), (Quantity, Amount), (Relation), (Shape), (State, Condition), (Substance), (Person, Human being), (Time).

Les relations fondamentales qui relient les synsets de noms sont l'hyperonymie (est une sorte de : *un chien est une sorte d'animal*) et la méronymie (est une partie de : *une roue est une partie de voiture*). La figure 1 montre les relations d'hyperonymie du substantif « doberman » fournies par WordNet 2.1. Rappelons que la racine unique de la hiérarchie des noms est **entity**, les relations d'hyperonymie sont ensuite notées « -- » ; ainsi, l'extrait suivant (fin de la figure 1) « *working dog, watchdog, guard dog pinscher --Doberman pinscher --* ***Doberman*** », signifie : un Doberman *est une sorte de* Doberman pinscher qui *est une sorte de* pinscher qui *est une sorte de* working dog, watchdog, guard dog.

(Noun.top) entity --physical entity -- object, physical object--living thing, animate thing --organism, being --animal, animate being, beast, brute, creature, fauna -- **(noun.animal)** chordate--vertebrate, craniate --mammal, mammalian --placental, placental mammal, eutherian, eutherian mammal --carnivore --canine, canid --dog, domestic dog, Canis familiaris --working dog, watchdog, guard dog--pinscher -- Doberman pinscher -- **Doberman**

Figure 1. La relation d'hyperonymie pour « Doberman » (WordNet 2.1).

1.3.1.2. Les adjectifs

Les adjectifs sont divisés en deux catégories : les adjectifs descriptifs (*grand, intelligent*) et les adjectifs relationnels. Les adjectifs descriptifs sont structurés par des relations d'antonymie (*grand, petit*) et de similarité. Aucune gradation n'est représentée dans WordNet. Les adjectifs relationnels sont reliés à un nom, par exemple *electrical* relié à *electricity* ou *maternal* à *mother.* Ils n'ont pas d'antonymie directe mais une indirecte via leur similarité sémantique avec des adjectifs qui ont un antonyme. Les adjectifs de couleur sont traités séparément.

1.3.1.3. Les verbes

Le lexique verbal est divisé en 15 domaines sémantiques, dont nous donnons la liste avec des exemples de verbes dans le tableau 6. Un verbe polysémique peut se trouver dans différents domaines comme *to hit* (frapper, atteindre) qui relève de 7 domaines (communication, compétition, consommation, contact, interactions sociales, mouvement et perception).

Les synsets de verbes sont reliés entre eux par les relations hiérarchiques d'hyperonymie et de méronymie. D'autres relations existent également, comme la « troponomie », qui signifie *V1 c'est V2 d'une certaine façon* comme ***jogger*** qui est ***courir*** *pour faire de l'exercice* ou ***crier*** qui est ***parler*** *fort*, l'implication lexicale quand on peut dire « *quelqu'un V1* » implique logiquement « *quelqu'un V2 »* comme *ronfler* qui implique *dormir*, mais pas l'inverse, et la relation de cause qui lie deux synsets ou deux verbes, comme *donner* et *recevoir*, tels que l'un dénote une cause et l'autre le résultat.

Les verbes (et non les synsets) sont reliés entre eux par des relations de synonymie et d'antonymie :

> « La plupart des verbes antonymes sont des verbes statiques ("live/die" (vivre/mourir), "differ/equal" (différer/égaler), "wake/sleep" (se réveiller/dormir), etc.) ou des verbes dénotant un

changement d'état ("grow/shrink" (grandir/rétrécir), "blacken/whiten" (noircir/blanchir), etc.) qui peut être exprimé par des attributs. » CF page 32.

Domaines sémantiques	**Exemples de verbes**
Changement	*grandir, chauffer*
Cognition	*penser, juger, douter, analyser*
Communication	*dire, demander, chanter*
Compétition	*attaquer, sauter*
Consommation	*manger, boire, ingérer*
Contact	*toucher, frapper,*
Creation	*coudre, cuisiner, peindre*
Émotion	*aimer, haïr, craindre*
États	*être avoir*
Interactions sociales	*élire, voter*
Météorologie	*pleuvoir, neiger, geler*
Mouvement	*se promener, voler, nager*
Perception	*voir, entendre, sentir*
Possession	*acheter, vendre, posséder*
Soins du corps	*Se laver, se peigner, panser*

Tableau 6. Domaines sémantiques des verbes de WordNet

À chaque synset du lexique verbal est associé un ensemble de « phrases représentatives **»** qui indique quel type de sujet est choisi par les verbes dans le synset, ainsi que les propriétés syntaxiques et sémantiques des compléments des verbes**.**

1.3.1.4. Les adverbes

Les adverbes sont reliés entre eux par des relations de synonymie et d'antonymie, et aux adjectifs quand ils en sont dérivés (*sad, sadly*). Les synsets contiennent également des exemples d'usage.

1.3.2. Les wordnets

WordNet est en cours de développement pour de nombreuses langues citons, parmi d'autres, Arabic WordNet[2] développé pour l'arabe, HindiWordnet[3] pour le hindi, ou le projet *EuroWordNet*[4] qui est une base de données multilingue constituée de lexiques organisés sur le modèle de WordNet pour plusieurs langues européennes (allemand, néerlandais, italien, espagnol, français, tchèque et estonien). L'association *The Global WordNet Association*[5] est une plate-forme de discussion et permet la connexion aux différents Wordnet existants.

Dans le domaine des sentiments, deux ressources sont en cours de réalisation à partir de Wordnet : SentiWordnet et Wordnet-Affect. Dans SentiWordnet, Esuli *et al.* (2006) ont assigné à chaque synset de WordNet un triplet (négativité, positivité, objectivité) ayant chacun un score de sentiment, la somme des 3 scores étant toujours égale à 0. Wordnet-Affect (Strapparava et Valitutti 2004) a été construit à partir d'un ensemble de synsets extraits de WordNet, synsets relevant de concepts affectifs. La hiérarchie de WordNet a été conservée, et des informations détaillant le contenu d'un point de vue affectif ont été ajoutées.

1.3.3. WordNet, Ontologie et TAL

Pour répondre à plusieurs critiques faites à WordNet, ses concepteurs ont apporté un certain nombre d'améliorations, notamment pour son utilisation avec une approche ontologique, ou dans le domaine du TAL. Une ontologie est considérée ici comme un ensemble de concepts organisés hiérarchiquement auxquels sont associés des attributs qui caractérisent le concept. Les relations d'hyperonymie qui hiérarchisent les synsets de WordNet sont de même type que les relations qui relient les concepts entre eux dans les ontologies ; il n'est donc pas étonnant que l'on considère parfois WordNet comme une ontologie. Mais WordNet n'a jamais été conçu comme une ontologie, même si, comme le souligne Georges Miller et Christiane Fellbaum (2007), quelques modifications ont été apportées pour le rendre plus proche des « bonne pratiques » ontologiques :

> « WordNet was originally constructed with 25 unique beginners, rather that a single root node. Most ontologists would agree that it is sensible to have a single class of all entities -real, abstract or imagined. Repeated requests to merge the 25 initial trees were finally met and WordNet version 2.1 provides a single unique beginner, labeled "entity." Second, WordNet did not distinguish between Types (classes) and Instances. So "A nation IS A KIND OF political unit" and "Spain IS A KIND OF nation" were represented in the same way. But nations are a class and Spain is an instance of that class. By working through all the nouns in WordNet, this distinction was finally drawn [MH06]. […] As ontologies tend to confine themselves to higher-level concepts, a mapping to a lexical resource is

[2] http://www.globalwordnet.org/AWN/
[3] http://www.cfilt.iitb.ac.in/wordnet/webhwn/
[4] http://www.illc.uva.nl/EuroWordNet/
[5] http://www.globalwordnet.org/

desirable as it extends the concepts covered by the ontology down into the "\leaves" of the hierarchies. The Suggested Upper Merged Ontology (SUMO, [NP03]) has been fully linked to WordNet as well as to wordnets of other languages. » page 210.

Une partie du succès de WordNet vient de son utilisation comme base de données lexicales dans le domaine du TAL, mais une critique souvent formulée concerne l'absence de liens entre les réseaux sémantiques des différentes catégories : verbes, noms, etc. Pour répondre à cette remarque, des liens morphosémantiques entre noms et verbes ont été ajoutés :

« But WordNet's effectiveness for WSD[6] is limited by the sparsity of its arcs. We tried to address the lack of cross-part of speech arcs at least in part by including "morphosemantic links" ([FM03]). For example, the verb *interest* is connected to the noun *interest* and the adjective *interesting*. Since *interest* and *interesting* each have several meanings -not all of them related to an adjective sense- the links had to be entered manually among the appropriate senses. » Miller et Fellbaum (2007) page 211.

Le manque de solution efficace pour répondre à la polysémie des mots est également souvent mentionné par les utilisateurs. En effet, quand un mot est associé à plusieurs synsets, il n'est pas aisé de distinguer entre ses différents sens. Pour aider les utilisateurs à désambiguïser les mots, un corpus[7] construit manuellement et contenant des annotations sémantiques a été associé aux synsets de Wordnet :

« Currently, the words in the definitions (or "glosses") that are part of each synset are manually disambiguated against WordNet. In this way, a user who wishes to better understand a gloss can refer to the synsets for more information on the nouns, verbs, adjectives, and adverbs in that gloss. In addition, we have created a semantically annotated corpus (the glosses) that can be used to train machine learning systems for WSD. » Miller et Fellbaum (2007) page 211.

1.3.4. Wordnet et la classification de Beth Levin

Les démarches de Beth Levin et de Christiane Fellbaum (dans WordNet) pour l'étude des verbes sont assez différentes mais ont néanmoins quelques points en commun relevés par Christiane Fellbaum :

« Bien que découvrir des classes sémantiques en examinant la syntaxe soit une approche très différente de celle de WordNet -- qui cherche à découvrir l'organisation sémantique par des relations sémantiques et lexicales -- les deux efforts ont en commun beaucoup de résultats. De nombreuses classes de Levin ayant les mêmes propriétés syntaxiques, sont dans WordNet des hiérarchies créées uniquement selon des critères sémantiques. De plus, les phrases qui indiquent les propriétés de sous-catégorisation des verbes et leurs restrictions sélectionnelles révèlent des irrégularités parallèles à celles qu'on trouve chez Levin. ». CF page 37.

[6] Word Sense Disambiguation
[7] http://wordnet.princeton.edu/glosstag

1.3.5. Exemple du verbe *adore* (Wordnet 2.1)

The verb **adore** has 1 sense (first 1 from tagged texts)

1. (1) adore -- (love intensely; "he just adored his wife")

Synonymes

----------par fréquences

adore -- (love intensely; "he just adored his wife")

=> love -- (have a great affection or liking for; "I love French food"; "She loves her boss and works hard for him")

----------par similarités

adore -- (love intensely; "he just adored his wife")

⇨ love -- (have a great affection or liking for; "I love French food"; "She loves her boss and works hard for him")

Termes coordonnés

adore -- (love intensely; "he just adored his wife")

-> love -- (have a great affection or liking for; "I love French food"; "She loves her boss and works hard for him")

=> love -- (be enamored or in love with; "She loves her husband deeply")

=> care for, cherish, hold dear, treasure -- (be fond of; be attached to)

=> dote -- (shower with love; show excessive affection for; "Grandmother dotes on her the twins")

=> adore -- (love intensely; "he just adored his wife")

Hyperonymes (**adore** est une façon de …)

adore -- (love intensely; "he just adored his wife")

⇨ love -- (have a great affection or liking for; "I love French food"; "She loves her boss and works hard for him")

Troponymes (différentes manières de…)

adore -- (love intensely; "he just adored his wife")

=> idolize, idolise, worship, hero-worship, revere -- (love unquestioningly and uncritically or to excess; venerate as an idol; "Many teenagers idolized the Beatles")

=> drool over, slobber over -- (envy without restraint)

Formes associées par dérivation

adore -- (love intensely; "he just adored his wife")

RELATED TO->(noun) adorer#1

⇨ admirer, adorer -- (someone who admires a young woman; "she had many admirers")

Exemples de phrases

adore -- (love intensely; "he just adored his wife")

EX: Sam cannot adore Sue

EX: Sam and Sue adore the movie

Usage : adore used as a verb is very rare (polysemy count = 1)

1.3.6. En guise de synthèse

- WordNet est un Réseau lexical ;
- la constitution du lexique a été faite à partir du « Brown corpus » puis enrichie avec différents thésaurus et lexiques ;
- Le réseau ne contient pas des mots mais des Synsets (ensembles de synonymes) ;
- les noms, verbes, adjectifs et adverbes sont organisés en réseaux sémantiques distincts, mais des liens morphosémantiques entre noms et verbes ont été ajoutés ;
- La hiérarchie des noms a une racine unique (nommée **Entity**) ;
- Les nœuds du réseau (le synsets) sont reliés par différentes relation hiérarchiques : l'hyperonymie (est une sorte de), la méronymie (est une partie de), et, pour les verbes, la « troponomie » qui signifie *V1 c'est V2 d'une certaine façon,* l'implication lexicale et la relation de cause ;
- il y a une relation de synonymie et d'antonymie entre mots, pas entre synsets ;
- ces descriptions ont été implémentées et sont directement utilisables par des humains et par des outils informatiques.

1.4. LE DICTIONNAIRE DES VERBES FRANÇAIS DE DUBOIS ET DUBOIS-CHARLIER

Le dictionnaire des verbes français a été élaboré par Jean Dubois et Françoise Dubois-Charlier (1997). Leur objectif et la méthode utilisée sont décrits dans l'introduction de leur ouvrage : (la référence est donnée par François *et al.*, 2007).

> « La classification syntaxique des verbes français repose sur l'hypothèse qu'il y a adéquation entre les schèmes syntaxiques de la langue et l'interprétation sémantique qu'en font les locuteurs de cette langue […]. Le schème syntaxique est défini, d'une part, par la nature des constituants de la phrase, leurs propriétés et leurs relations, et, d'autre part, par les mots du lexique qui entrent dans les types de constituants définis […]. L'objet est donc de faire l'inventaire le plus complet possible de ces schèmes syntaxiques. »

L'approche adoptée est assez voisine de celle du Lexique grammaire car le comportement des verbes est étudié au sein de la phrase élémentaire.

1.4.1. Description du dictionnaire

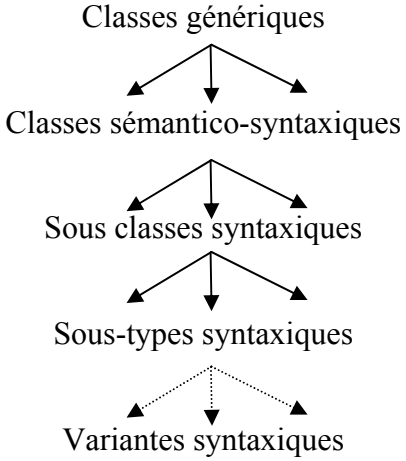
Jean Dubois et Françoise Dubois-Charlier ont classé 12 310 verbes français. Un tiers de ces verbes faisant l'objet de plusieurs entrées, le dictionnaire contient 25 610 entrées verbales différentes. Ces verbes ont été répartis en quatorze classes génériques, parmi lesquelles treize se rapportent à des domaines sémantiques (verbes de communication, de don, etc.) et une classe contient les auxiliaires. (cf. Tableau 7).

Classes génériques	Classes sémantico-syntaxiques
C Communication	**C1** s'exprimer par un son, une parole, **C2** dire/demander qc **C3** montrer qc, **C4** figuré de C1 et C2
D Don, privation	D1 donner qn à qn/qc, D2 donner qc à qn/qc D3 figuré de D2
E Entrée, sortie	E1 (faire) sortir/venir de qp ou aller qp, sujet hum E2 figuré de F1 E3 (faire) sortir/venir de qp ou aller qp, sujet non-animé E4 figuré de E3
F Frapper, toucher	F1 frapper ou toucher qn, F2 figuré de F1 F3 frapper ou toucher qc, F4 figuré de F3
H États physiques et comportements	H1 être/mettre qn dans état physique H2 avoir tel comportement, activité H3 être/mettre qc dans tel état, H4 figuré de H3
L Locatifs	L1 être ou mettre qn qp, L2 figuré de L1 L3 être ou mettre qc qp, L4 figuré de L3
M Mouvement sur place	M1 faire/imprimer mouvement, hum, M2 figuré de M1 M3 faire/imprimer mouvement, non-animé M4 figuré de M3
N Munir, démunir	N1 munir/démunir qn de qc, N2 figuré de N1 N3 munir/démunir qc de qc, N4 figuré de N3
P Verbes psychologiques	P1 avoir tel sentiment, telle pensée P2 faire avoir tel sentiment, P3 manifester telle pensée sur qn/qc

Classes génériques	Classes sémantico-syntaxiques
R Réalisation, mise en état	R1 mettre en état le corps de qn, R2 créer qn, un personnage R3 fabriquer qc, mettre qc en un certain état, R4 figuré de R3
S Saisir, serrer, posséder	S1 saisir, arrêter, serrer qn, S2 prendre, choisir, abandonner qn S3 saisir, garder qc ; tordre, tirer qc, S4 figuré de S3
T Transformation, changement	T1 lier, unir qn à ; détacher qn de, T2 figuré de T1 T3 lier, unir qc à ; détacher qc de, T4 figuré de T3
U Union, réunion	U1 lier/unir qn à ; détacher qn de, U2 figuré de U1 U3 lier/unir qc à ; détacher qc de, U4 figuré de U3
X Verbes auxiliaires	X1 auxiliaires, semi-aux., modaux, X2 impersonnels X3 existence, X4 inchoatifs, résultatifs

Tableau 7. Classes génériques et syntaxico-sémantiques des verbes français (Dubois et Dubois–Charlier 1997)

Chacune de ces classes est divisée en 3 ou 4 classes sémantico-syntaxiques qui décrivent les propriétés du sujet et du complément des verbes (être vivant / non-animé et emploi propre ou emploi figuré). Ces classes sont elles-mêmes constituées de sous-classes syntaxiques qui, à leur tour, sont sous-divisées en sous-types syntaxiques qui contiennent éventuellement des variantes syntaxiques. La description est donc une arborescence à 5 niveaux :



Ainsi, la classe générique des verbes psychologiques (classe P) contient 2074 entrées, réparties en trois classes sémantico-syntaxiques (P1 à P3, cf. Tableau 7).

La sous-classes P1 contient 1159 entrées de type « avoir tel sentiment, telle activité consciente», comporte dix sous-classes syntaxiques E1a à E1j.

La sous-classe syntaxique P1a contient 211 verbes de type « être dans tel état psychologique ou avoir tel sentiment », et comporte douze sous-types : les intransitifs (*Jean boude*[8]), les pronominaux (*Jean s'éclate au concert*), les pronominaux intransitifs

[8] Les exemples de phrases sont extraits de Dubois et Dubois et Charlier (1997).

(avec 2 variantes : *Jean colère, Jean se colère à voir cette lâcheté*) les pronominaux avec factitifs (*Jean se braque dès la première remarque, ce reproche a braqué Jean*), les intransitifs pronominaux (*Jean culpabilise, Jean se culpabilise, cette séparation culpabilise Jean*), les pronominaux intransitifs avec factitif (*Jean angoisse, Jean s'angoisse, cette prédiction angoisse Jean*), les pronominaux avec factitif à sujet humain et instrumental pouvant devenir le sujet de phrase (*Paul se décourage facilement, ces critiques le découragent),* les pronominaux avec locatifs *(Jean se plaît à la campagne)*, les intransitifs avec un complément en *pour* (*Jean craint pour sa vie),* les pronominaux avec un complément en *pour* (*Jean se passionne pour la musique baroque, Jean est passionné pour la musique*), et enfin les formes « *être* participe passé » avec manière *bien/mal* (*Paul est mal luné ce matin)* ou sans manière (*Son rire est forcé*).

À chaque classe générique est associé un ensemble de schémas syntactico-sémantiques appelés *opérateurs* qui caractérisent cette classe. Par exemple, il y a 6 opérateurs pour la classe des verbes psychologiques : *sent* (avoir tel sentiment), *f.sent* (donner tel sentiment), *ger.mens* (avoir telle activité consciente), *percep (*avoir telle perception), *percep.mens* (avoir tel savoir) et *scrut* **(**appliquer son esprit à quelque chose)..

À chaque entrée lexicale, comme *s'amuser* de la classe P1a, sont associés une description du sens (*égayer,divertir*), des dérivés nominaux (*ment, -teur*) et une ou plusieurs phrase(s) qui exemplifie(nt) l'emploi du verbe *(On s'a~ avec des riens. Un rien a~ P. On a~ P avec un rien)*.

1.4.2. La classe des verbes psychologiques

La classe des verbes psychologiques, notée **P**, contient 2074 entrées verbales. La sémantique des verbes qui figurent dans cette classe est plus étendue que celle considérée habituellement dans les travaux sur les verbes psychologiques. En effet, cette classe comprend les verbes de sentiment mais également tous les verbes qui expriment une activité de la conscience et de la pensée, comme *oublier, penser, critiquer*, etc. Nous indiquons dans le tableau 8 l'organisation générale des verbes psychologiques dans le dictionnaire. Le nombre d'entrées et un exemple sont donnés pour chaque sous-classe.

1.4.3. Le dictionnaire des verbes comme ressource en TAL

Une version électronique du dictionnaire a été implémentée dans un tableur *(EXCEL)* et dans une base de données relationnelle *(ACCESS)* [9]. Un projet est en cours (Denis Le Pesant, Modyco - Université Paris X, Paul Sabatier, LIF - Université et Max Silberztein, Laseldi - Université de Franche-Comté) pour reformater cette version et l'intégrer à des logiciels de TAL existants comme *Nooj* (Koeva *et al.* 2007).

[9] http://www.modyco.fr/

Classes sémantico-syntaxiques	Sous classes syntaxiques	Nombre d'entrées	Exemple
***Avoir tel sentiment, telle activité consciente* (P1)**	Être dans tel état psychologique avoir tel sentiment (P1a)	211	*bouder*
	Manifester physiquement sentiment de joie, de douleur (P1b)	34	*ricaner*
	Avoir telle manifestation physique due à tel sentiment, avoir tel sentiment pour/contre (P1c)	214	*blanchir*
	Avoir tel sentiment portant sur qque chose ou poussant à qque chose (P1d)	70	*parier*
	Avoir telle activité mentale (P1e)	180	*méditer*
	Avoir telle activité mentale dirigée sur tel objet (P1f)	71	*penser*
	Prendre quelque chose comme objet d'activité mentale (P1g)	89	*oublier*
	Avoir l'esprit tendu vers (P1h)	97	*penser*
	Avoir tel sentiment portant sur qque chose (P1i)	102	*concevoir*
	Avoir tel sentiment à l'égard de qqu'un (P1j)	91	*adorer*
***Faire avoir tel sentiment à qqu'un, l'augmenter ou le diminuer* (P2)**	Induire tel sentiment (P2a)	438	*apaiser*
	Donner tel sentiment à qqu'un (P2b)	18	*plaire*
	Augmenter, diminuer, remplir tel sentiment chez qqu'un (P2c)	80	*encourager*
***Manifester telle connaissance, telle analyse* (P3)**	Diriger son esprit, son activité réflexive sur qqu'un (P3a)	71	*analyser*
	Diriger son esprit, son activité réflexive sur qque chose (P3b)	136	*critiquer*
	Avoir ou ne pas avoir la connaissance de qque chose, qqu'un (P3c)	172	*connaître*

r

Tableau 8. Organisation des verbes psychologiques dans le dictionnaire des verbes français de Dubois et Dubois-Charlier (1997)

CHAPITRE 2. SENTIMENTS, ÉMOTIONS, ET OPINIONS : DESCRIPTIONS LINGUISTIQUES

Nous venons de présenter quatre études qui ont été des étapes importantes dans la description du lexique. Nous montrerons comment certaines de ces études ont influencé nos travaux, et en particulier le Lexique grammaire qui a été le point de départ de notre étude sur les verbes de sentiment. Nous avons tenté de montrer que la description linguistique approfondie d'un domaine sémantique particulier peut être utilisée pour repérer et analyser les éléments de ce champ sémantique dans des textes. Nous avons ainsi étudié le domaine des sentiments, travaux que nous décrirons dans une première partie, puis celui des opinions que nous aborderons dans une deuxième partie. Le fil conducteur de cet exposé sera le passage de l'étude du mot, à la prise en considération du mot dans une proposition, un segment (que nous définirons) ou une phrase et enfin à l'étude d'un texte et de plusieurs textes.

2.1. LES SENTIMENTS ET ÉMOTIONS

Nous nous sommes intéressée à la façon dont les sentiments sont exprimés dans la langue avec l'objectif de parvenir à une description linguistique intuitivement plausible, et donc de rendre compte de la relation de synonymie, ou plus généralement de la sensation du locuteur que certains mots relèvent du même champ sémantique (la « tristesse », par exemple) alors que d'autres relèvent d'affects différents (la « tristesse » *vs* la « joie »).

2.1.1. Le mot

Il existe de très nombreuses définitions des mots « sentiment » et « émotion », selon les niveaux de langage, populaire ou scientifique, selon les disciplines, ou encore en fonction

du point de vue théorique adopté. Pour notre étude, nous avons retenu ce que l'on peut *éprouver* ou *ressentir,* et nous parlerons indifféremment de sentiment ou d'émotion. Pour rendre compte du champ sémantique de ces mots de sentiment, nous avons considéré les émotions en général qui ont, depuis longtemps, suscité l'intérêt des philosophes et des psychologues. On peut distinguer plusieurs approches dans l'étude et la catégorisation des émotions :

1) de nombreux auteurs définissent des émotions de base, comme par exemple Descartes (1649/1988) ou Philip N. Johnson-Laird et Keith Oatley (1989). Dans son *Traité sur les passions de l'âme*, Descartes, étudie six passions simples : l'admiration, l'amour, la haine, le désir, la joie et la tristesse, primitives dont sont dérivées toutes les autres passions. Voici un exemple de définition :

« Art.86. La définition du désir. La passion du désir est une agitation de l'âme causée par les esprits qui la dispose à vouloir pour l'avenir les choses qu'elle se représente être convenables. Ainsi on ne désire pas seulement la présence du bien absent, mais aussi la conservation du présent ; et de plus l'absence du mal, tant de celui qu'on a déjà que de celui qu'on croit pouvoir recevoir au temps à venir. »

Philip N. Johnson-Laird et Keith Oatley ont établi une taxinomie à partir de 590 mots pour décrire les émotions. Leur classification comprend cinq rubriques : le bonheur, la tristesse, la peur, la colère, le dégoût. Chaque rubrique est à son tour classée en six catégories : les émotions de base, les émotions relationnelles, les émotions qui ont une cause, les émotions qui sont cause d'autres émotions (ce sont en général des verbes) et enfin les émotions complexes ;

2) d'autres proposent des modèles dimensionnels qui partent de l'hypothèse que toutes les réactions affectives peuvent être décrites à partir de plusieurs dimensions, en général deux : la valeur (plaisir/déplaisir ou affect négatif/affect positif) et le dynamisme (activation/tranquillité). Parmi eux, citons les travaux de Charles E. Osgood *et al*. (1957, 1979), et ceux de James A. Russel et Geraldine A. Pratt (1980) ;

3) enfin une troisième approche considère une émotion comme le résultat d'une évaluation cognitive que fait une personne des interactions entre ses croyances, ses espoirs etc. et son environnement. Citons en particulier le modèle d'Andrew Ortony *et al*. (1988), dans lequel les émotions peuvent être issues soit de l'évaluation des conséquences d'événements, de l'action d'agents ou de la perception d'un objet. Par exemple, un évènement permettant de réaliser un but déclenche de la joie ; une action de l'agent contraire à ses principes déclenche de la honte ; la perception d'un objet peut déclencher du dégoût suivant les préférences de l'agent. Les auteurs spécifient dans ce modèle les conditions de déclenchement de 22 émotions types.

Certains auteurs se sont intéressés au vocabulaire des émotions et à la façon dont le langage est utilisé pour décrire des états émotionnels ou des attitudes. C'est le cas par exemple de Joel R. Davitz (1969) qui a fait une description systématique de 50 termes d'émotions, ou de J. R Martin et P.R.R. White (2005) qui ont fait une taxinomie du

langage utilisé par des locuteurs anglais pour exprimer des attitudes et des engagements. Ils distinguent trois catégories d'attitudes : les émotions, les jugements et les appréciations. Leurs travaux s'inscrivent dans l'approche de la linguistique systémique fonctionnelle (Michael A. K. Halliday 1994/2004).

En linguistique, l'étude des sentiments a été abordée dans le cadre de travaux plus généraux, en sémantique lexicale par exemple dans la recherche de primitifs sémantiques par Anna Wierzbicka (1993), ou dans la classification des verbes de l'anglais faite par Beth Levin (1993) dont nous avons parlé au §1.2.

Les sentiments ont fait l'objet d'un grand intérêt dans différents formalismes grammaticaux. Les travaux d'Adriana Belleti et Luigi Rizzi (1988)], Ray Jackendoff (1990), Danielle Leeman (1995), Nicolas Ruwet (1993, 1994, 1995) Danièle Van de Velde (1999) et Jan Van Voorst (1992), parmi d'autres, en témoignent.

Enfin, plusieurs grandes ressources lexicales informatisées contiennent des informations sur les sentiments et les émotions. C'est le cas de deux d'entre elles très utilisées : Wordnet, que nous avons décrit au §1.3., et Framenet (2008). La base de données Framenet est organisée sous forme de réseau. Elle contient 10 000 unités lexicales, dont 6 100 annotées. Chaque item lexical est relié à un patron sémantique ou frame et associé à un ensemble de mots qui évoquent ce même frame. Il y a actuellement 825 frames exemplifiés par 135 000 phrases annotées.

Dans notre étude, notre hypothèse est que la façon dont nous conceptualisons le monde se reflète dans la langue, dans les interrelations établies entre les mots par le système linguistique, et que le sentiment de synonymie repose sur le fait que les mots ressentis comme relevant du même champ notionnel ont des comportements formels similaires. Notre lexique comporte un millier de mots exprimant des sentiments et des émotions, tels que des noms comme *amour* ou *peur*, des verbes comme *aimer, effrayer*, ou des adjectifs comme *amoureux* ou *jaloux*. Nous avons également répertorié des expressions plus ou moins « figées » et familières, exprimant un sentiment, comme « *avoir les chocottes* » (avoir peur) ou « *boire du petit lait* » (jubiler).

Nous proposons une classification dans laquelle ces mots sont répartis en classes sémantiquement homogènes. Pour déterminer ces différentes classes, nous avons fait appel à la notion de synonymie telle que la définit Igor Mel'čuk (1992) :

> « Dans notre approche, la synonymie des phrases est prise comme une notion intuitive de départ. Nous posons que la capacité de juger si deux phrases données sont équisignifiantes (= ont le même sens) fait partie de la compétence linguistique des locuteurs. [...] nous définissons la notion de " sens " à partir de la notion de "même sens", cette dernière étant considérée comme acquise. ». page 11.

Nous avons commencé cette étude des sentiments par le cas des verbes dits « psychologiques », en partant des verbes de la table 4 Lexique grammaire, (Gross 1975), qui est définie, comme nous l'avons vu au §1.1.2., par un sujet non restreint et un complément humain. La table 4 contient un grand nombre de verbes psychologiques, mais

nous en avons écarté certains car ils ne correspondent pas à notre critère d'interprétation (N_1 ressent un sentiment déclenché par N_0). Ainsi, nous n'avons pas retenu comme psychologiques les verbes qui acceptent le complément *aux yeux de Nhum*, qui implique un jugement extérieur, comme dans les exemples suivants :

Sa lâcheté déshonore Marie aux yeux de Max

**Sa lâcheté irrite Marie aux yeux de Max*

Ce complément sépare la classe de *déshonorer* (non psychologique) de celle de *irriter* (psychologique) car *Max* juge que la lâcheté de *Marie* déshonore *Marie*, mais *Marie* (N_1) peut ne rien ressentir. À partir de ces critères nous avons sélectionné près de 400 verbes auxquels nous avons ajouté les verbes de sentiment comme *adorer* (environ 100). Puis nous avons classé ces verbes en classes homogènes sémantiquement.

2.1.1.1. Les classes sémantiques des sentiments

Ces classes ont été construites à partir de notre propre intuition à l'aide de dictionnaires, puis nous les avons soumises pour vérification, à un grand nombre de locuteurs qui sont pour la plupart des « experts » (des chercheurs en linguistique), mais également des non spécialistes. Ce premier classement a été fait au milieu des années 90, période où le recours à de grands corpus électroniques n'existait pas encore, la méthode de travail était alors l'introspection.

Cette classification a été revue et étendue aux noms, adjectifs et quelques expressions plus ou moins figées, et contient actuellement 40 classes. Pour cela, nous avons étudié ces mots en contexte dans des grands corpus électroniques comme la base Frantext [FRA 08], et dans différents textes disponibles sur le Web (articles de presse, blogs, critiques, moteur de recherche, etc.). Une classe a été supprimée, **Endurcir,** une autre a changé de polarité (**Émouvoir** de « + » est devenue neutre), d'autres ont été ajoutées. Les verbes réflexifs n'ont pas été pris en compte.

Cette classification n'est pas totalement représentative de la complexité de la langue française car chaque mot a un sens particulier, et il y a très peu de vrais synonymes dans la langue. S'il y a un large consensus pour regrouper ensemble des mots tels que *énerver, exaspérer* et *irriter*, d'autres regroupements donnent lieu à des avis divergents, mais pour une formalisation en vue d'un traitement informatique, des regroupements sont nécessaires. De plus la langue évolue, et des modifications de sens apparaissent fréquemment. La langue du Web est également différente de la langue des romans, car certains textes sont écrits par des internautes qui ne respectent pas la langue « standard ».

Nous distinguons trois catégories de classes selon leur valence (ou polarité) :

1-) Les classes de polarité négative qui décrivent un sentiment plutôt désagréable, comme la peur ou la colère. On en dénombre vingt-trois, telles **Peur**, **Tristesse**, **Colère**, etc.

2-) Les classes de polarité positive qui décrivent un sentiment plutôt agréable, comme l'amour ou l'amusement. Il y a quatorze classes de ce type, telles **Amour**, **Intérêt**, **Passion**, etc.

3-) Les classes de mot sans polarité indiquée, soit qui décrivent un sentiment qui n'est souvent ni désagréable ni agréable comme la classe **Indifférence**, soit avec une polarité qui sera indiquée par le contexte : les classes **Émouvoir** et **Étonnement**. Par exemple, dans la phrase « *La hauteur de la Tour Eiffel surprend les touristes* »*,* le sentiment ressenti par les touristes n'a pas de polarité explicitement exprimée alors que dans la phrase « *La trahison de son mari a surpris Marie* » ou « *La vue de tous ces morts causés par la bombe a bouleversée Marie* », on comprend que Marie éprouve un sentiment assez pénible et négatif. Par convention, nous noterons « neutre » une polarité ni positive ni négative. Précisons toutefois que les mots d'indifférence sont le plus souvent de polarité négative et les mots d'émotion de polarité plutôt positive. La liste des classes est donnée dans le tableau 9.

Polarité négative				
Amertume	Colère	Déception	Dédain	Dégoût
Déprime	Déranger	Désapprobation	Déconcerter	Douleur
Effarement	Ennui	Haine	Honte	Indignation
Intimidation	Jalousie	Obsession	Offense	Peur
Remords	Tracas	Tristesse		
Polarité positive				
Amour	Amusement	Apaisement	Émerveillement	Excitation
Fascination	Intérêt	Joie	Orgueil	Passion
Pitié	Satisfaction	Soulagement	Stimulation	
Sans polarité				
Émouvoir	Étonnement	Indifférence		

Tableau 9. Classes sémantiques des sentiments

La distribution des mots dans ces classes n'est pas homogène. En effet, certains sentiments sont décrits par très peu de noms et d'adjectifs, comme la honte ou le remords, et par aucun verbe.

À ces mots et à ces classes, nous avons associé un ensemble de propriétés syntaxiques et sémantiques dont nous allons décrire un sous-ensemble, on trouvera dans (Mathieu 1999, 2000 et 2005) l'ensemble des descriptions.

2.1.1.2. Relations entre classes

Les classes sémantiques sont reliées entre elles par des relations de sens, d'intensité et d'antonymie, représentées par des graphes simples. Le graphe d'intensité est un graphe

orienté selon l'intensité du sentiment exprimé et ressenti. Nous avons représenté une partie de ce graphe dans la Figure 2.

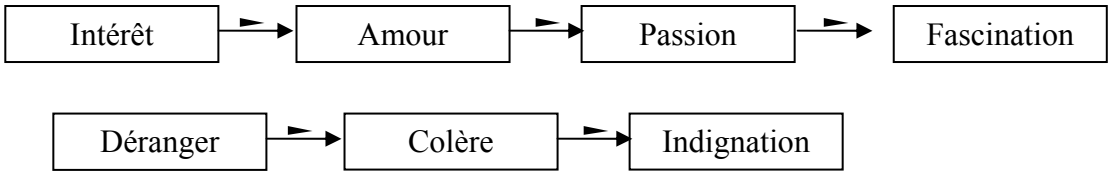


Figure 2. Exemples de relations d'intensité entre classes sémantiques

Les classes sémantiques sont reliées par des arcs qui sont étiquetés par des indications de degré d'intensité, représentés par le symbole « ▸ ». Ainsi, un accroissement d'intensité de l'intérêt sur les éléments de la classe **Intérêt** (*intérêt, attraction*, etc.), est décrit par les mots de la classe **Amour** (*affection, amour,* etc.), un sentiment d'amour plus fort est reflété par la classe **Passion** (*passion, excitation*, etc.), et un sentiment de passion plus intense est reflété par les mots de la classe **Fascination** (*fascination, subjuguer*, etc.). De la même façon, il y a une gradation de l'intensité entre les classes **Déranger** (*importuner*, *gêner,* etc.), **Colère** (*exaspérer, irriter,* etc.) et **Indignation** (*indigner*, *scandaliser*, etc.).

Le graphe des relations de sens est un graphe orienté dont les arcs sont étiquetés par des traits sémantiques. Il contient divers types d'information. D'une part, il représente les proximités sémantiques de certaines classes, indiquées par des arcs étiquetés avec des traits sémantiques spécifiques. Ainsi, dans l'exemple de la figure 3, la classe **Intérêt** (*attirance, attirer, intérêt*, etc.) est reliée à la classe **Excitation** (*émoustiller, troubler,* etc.) par l'ajout d'un désir sexuel.

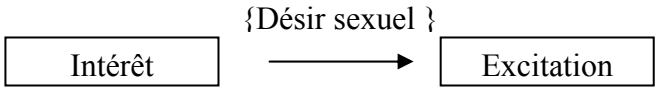


Figure 3. Exemple de relation de sens entre classes sémantiques (1/3)

D'autre part, dans ce graphe est indiqué comment la sémantique de certaines classes est construite à partir de celles de plusieurs autres ; un exemple est donné dans la figure 4, avec la classe **Amertume** (*aigrir, amertume*, etc.) qui est un mélange de tristesse et de rancœur due à une déception (Le Petit robert 2001).

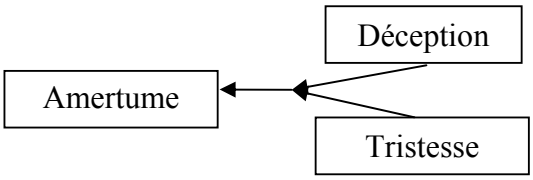


Figure 4. Exemple de relation de sens entre classes sémantiques (2/3)

Enfin, ce graphe contient également le fait qu'une modification de sens, indiquée par la spécification d'un trait sémantique particulier, peut faire basculer la polarité d'une classe « neutre ». C'est le cas dans l'exemple de la figure 5, où la classe **Étonnement** (qui contient des mots d'étonnement comme *étonner*, *surprendre*, *stupéfaction*, etc.) sans polarité explicite devient une classe de polarité positive, la classe **Émerveillement** *(éblouir* ou *émerveiller, etc.*)**,** si l'étonnement est mêlé d'admiration ; à l'inverse, elle devient une classe de polarité négative si l'étonnement est mêlé d'effroi (classe **Effarement :** *consterner, effarer*, etc.)

Les relations d'antonymie caractérisent des mots ayant un sens opposé. Ainsi, la classe **Colère** est antonyme de la classe **Apaisement**, c'est-à-dire que chaque verbe, nom ou adjectif de la première classe est antonyme avec au moins un verbe, nom ou adjectif respectivement de la seconde classe et vice versa.

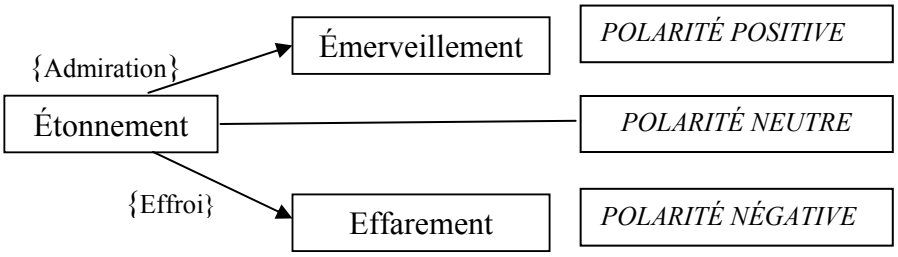


Figure 5. Exemples de relation de sens entre classes sémantiques (3/3)

Une classe peut avoir plusieurs antonymes et être antonyme de plusieurs classes. La relation d'antonymie n'implique pas une inversion de polarité ; ainsi, les classes **Apaisement** et **Stimulation** sont toutes deux de polarité positive. Un extrait du graphe des antonymes est donné dans la figure 6. L'antonymie entre deux classes est indiquée par un arc et le symbole « ↔ **».**

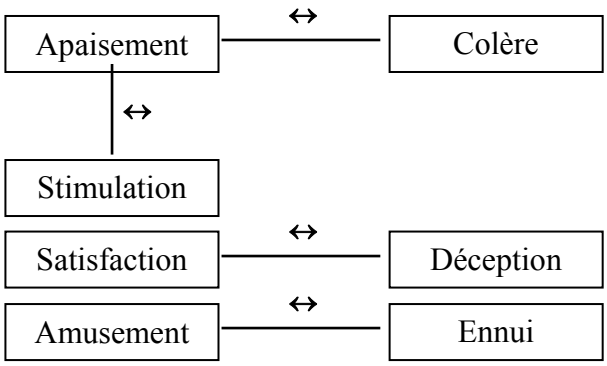


Figure 6. Exemples de relation entre classes sémantiques

2.1.1.3. Propriétés des items lexicaux

Nous avons étudié un ensemble de propriétés linguistiques d'une part pour les verbes, d'autre part pour les emplois de noms et d'adjectifs.

2.1.1.3.1. Intensifieur

L'hypothèse que nous avons proposée pour les verbes est qu'il y a des verbes « neutres » et des verbes qui ont un intensifieur incorporé. Considérons les phrases (1) et (2) :

(1) *Les fantômes terrifient Luc*

(2) *Les fantômes effrayent beaucoup Luc*

Le sens de ces phrases est très proche, ce qui pourrait indiquer que le verbe *terrifier* est synonyme de *effrayer beaucoup* et contient l'intensifieur *beaucoup*. Ainsi, *effrayer* et *terrifier* sont dans la même classe (**Peur**), mais *terrifier* est plus fort qu'*effrayer*. De la même façon, *exaspérer* et *irriter* sont dans la même classe (**Colère**), mais *exaspérer* est plus fort qu'*irriter*. Nous avons défini une propriété [Intensifieur] qui décrit cette différence d'intensité, dont la valeur est « Médian » pour *irriter* et *effrayer*, et « Fort » pour *exaspérer* et *terrifier*. Nous avons étendu cette propriété des verbes aux autres catégories de mots (nom et adjectif).

2.1.1.3.2. Emploi métaphorique des verbes

Certains verbes ont uniquement un sens psychologique, comme *aimer*, alors que d'autres, tels *irriter* ou *éblouir*, ont deux sens, un qui est « concret » ou « propre » (phrases 3 et 4) et un psychologique par métaphore (phrases 5 et 6) :

(3) *Le soleil irrite Marie* (sa peau)

(4) *La lumière éblouit Marie* (ses yeux)

(5) *Paul irrite Marie* (par son comportement)

(6) *L'intelligence de l'orateur éblouit Marie*

Cette particularité d'existence ou non d'un emploi non psychologique pour un verbe, est décrite par la propriété [Emploi non psychologique], dont la valeur est « non » pour le verbe *aimer* et « oui » pour le verbe *irriter*.

2.1.2. La proposition

Certaines propriétés caractérisent les constructions dans lesquelles l'item lexical peut apparaître, comme la position syntaxique de l'expérienceur, l'agentivité du sujet du verbe, etc.

2.1.2.1 Expérienceur

On appelle « expérienceur » la personne qui ressent l'émotion ou le sentiment, et « objet » la cause ou l'objet de ce sentiment. Ainsi, dans les phrases (7) et (8), l'expérienceur est *Marie* et l'objet est *Paul* :

(7) *Marie aime Paul*

(8) *Paul effraye Marie*

Ces structures diffèrent par la position syntaxique de l'expérienceur. Dans (7), l'expérienceur est le sujet syntaxique *Marie* et le complément *Paul* est l'objet de son amour. Dans (8), l'expérienceur est le complément d'objet *Marie*, et le sujet *Paul* est la cause de sa frayeur. La valeur de cette propriété [Expérienceur] est « Sujet » pour les verbes comme *aimer*, et « Complément » pour les verbes comme *irriter*.

2.1.2.2. Agentivité du sujet du verbe

De nombreux verbes de sentiment, comme le verbe *effrayer*, ont un sujet qui peut avoir une interprétation agentive. La phrase *Luc effraye Marie* a deux interprétations possibles : Luc effraye Marie de façon volontaire, intentionnellement, ou il l'effraye sans le vouloir, par son comportement ou son apparence ou quelque chose d'autre. Cette propriété est notée [Sujet agentif] (« sujet actif » dans Mathieu 2000), elle a pour valeur « oui » pour le verbe *effrayer* et « non » pour les verbes *tracasser* et *obséder,* par exemple.

2.1.2.3. Structure et sélection des arguments du verbe

Quand l'expression qui décrit celui ou celle qui éprouve le sentiment (l'expérienceur) est en position de complément, il peut toujours être explicitement une personne, comme *Marie* dans *Paul irrite Marie,* mais il peut également, pour certains verbes, être parfois une expression métonymique se référant à une personne. Nous avons distingué trois catégories d'expressions, selon la distance de la métonymie par rapport à la personne : ce peut être *(i)* une métonymie totale, une partie du corps ou « partie de l'âme », comme *cœur* ou *esprit, (ii)* une métonymie partielle, un nom de sentiment ou de qualité comme *colère*, *vanité*, etc., ou *(iii)* aucune métonymie, tout autre nom comme, par exemple, *espoirs, convictions*, etc.

Certains verbes sélectionnent les noms de la première catégorie tel *briser*, par exemple dans *Marie a brisé le cœur de Luc*, certains verbes sélectionnent en plus les noms de la seconde catégorie comme *apaiser* ou *calmer* comme dans *La berceuse a apaisé/ calmé la colère du bébé*, certains verbes acceptent également les noms de la troisième catégorie comme *satisfaire* dans *Les promesses de Marie ont satisfait les espoirs de Luc*. Enfin, certains verbes comme *tracasser* ou *dégoûter*, acceptent parfois des compléments non strictement humains, mais avec difficulté.

L'association de cette caractérisation du complément et de la propriété indiquant un possible emploi métaphorique permettra, par exemple, de reconnaître un sens « concret » dans *Marie a brisé les doigts de Luc* et une interprétation psychologique dans *Marie a brisé le cœur de Luc.*

2.1.2.4. Quelques contraintes dans les constructions passives

Nous considérons ici les verbes comme *irriter*, c'est-à-dire dont l'expérienceur est en position de complément. À la forme passive, le complément d'agent est introduit par la préposition *par*, comme dans la phrase *Paul est calmé par les paroles de Marie*. Le passif de nombreux verbes, *étonner* par exemple, peut se former également avec la préposition *de,* comme dans *Paul est étonné de l'attitude de Marie*. Mais pour certains verbes cela n'est pas possible : on ne peut pas dire **Paul est calmé des paroles de Marie* ou **Paul est intéressé de Marie/ de la politique.* Nous notons cette propriété [N_1 *est Vpp de* N_0] avec pour valeur « non » pour les verbes *calmer* et *intéresser* et « oui » pour le verbe *étonner*.

Les caractéristiques sémantiques du sujet contraignent les possibilités de formation du passif avec la préposition *de*. Ainsi, si le sujet est un nom humain, le passif n'est pas toujours possible. Par exemple les phrases à l'actif *Luc déconcerte Marie* et *Le comportement de Luc déconcerte Marie* et la phrase passive *Marie est déconcertée du comportement de Luc* sont acceptables, mais *Marie est déconcertée de Luc* est plus difficile. Il en est de même si le sujet est un substantif concret. En effet, les phrases *Marie est déconcertée de ce tableau/tissu* sont difficiles, voire inacceptables.

La grande majorité des verbes de sentiment ont une forme passive avec le complément d'agent effacé. C'est le cas du verbe *calmer*. Ainsi on peut dire *Paul calme Marie*, *Marie est calmée par Paul* et *Marie est calmée*. Cependant, un petit groupe de verbes interdit cet effacement, comme *tenailler* par exemple, car si *Les remords tenaillent Marie* et *Marie est tenaillée par les remords* sont des phrases acceptables, on ne peut dire *Marie est tenaillée.* On nomme cette propriété [passif avec agent obligatoire]. Elle prend la valeur « oui » pour *tenailler* et « non » pour *calmer*.

Enfin, les quatre cinquièmes de ces verbes à agent obligatoire sont des emplois figurés de verbes dont l'emploi propre autorise un complément d'objet humain. Le passif sans agent n'est alors possible que par référence à l'emploi propre. Par exemple *Marie est brûlée* peut être le passif de *Les flammes ont brûlé Marie* qui est un emploi propre mais non de *La*

passion brûle Marie. De même *Max est rasé* peut être le passif de *Le coiffeur rase Max* mais beaucoup plus difficilement de *Ce discours interminable rase Max*.

2.1.2.5. Force des sentiments

La force d'un sentiment peut être intensifiée, comme nous l'avons vu, par l'emploi d'un mot de même sémantique mais avec une intensité plus forte par exemple *peur/terreur*, mais également par l'utilisation d'adjectif ou de verbe particulier. Ainsi, les noms de sentiment peuvent être intensifiés par les adjectifs dits « intensifs » comme *grand, immense*, *profond*, habituellement utilisés pour dénoter l'intensité (*un amour profond*), mais aussi avec des adjectifs plus spécifiques à certains noms de sentiment comme *fou, ivre, mort (être fou de joie, ivre de colère, mort de peur*) et également des adjectifs de couleur : on peut être *vert (de jalousie, de rage, de colère, de peur), rouge de colère, blanc (de colère, de peur), être jaune de jalousie, blême (de colère, de rage), pâle de colère, avoir une peur bleue, une colère noire,* etc. Ces adjectifs, et en particulier ceux de couleur, sont utilisés le plus souvent pour des sentiments à polarité négative dont ils augmentent la force.

Certains verbes indiquent une modification de l'intensité de sentiment, en général une augmentation, exprimée par une manifestation physique du corps tout entier : *déborder de joie, trembler de peur/de rage, crever de jalousie,* ou bien d'une partie seulement, souvent le visage : *écumer de rage*, *blêmir de peur, rougir/blanchir de colère* [LEE 91]. L'intensité exprimée par les verbes de sentiment peut également être modifiée, souvent par l'ajout d'adverbe ou de locution adverbiale : *aimer passionnément/à la folie*. Enfin, des adverbes comme *peu, très*, etc. modifient la force d'adjectifs ou de verbes : *il est très amoureux, très inquiet*, etc.

2.1.3. La phrase : formation de paraphrase par nominalisation

Les phases avec un verbe de sentiment comme *Luc aime Marie* ou *La hauteur de la tour Eiffel étonne Jean* peuvent être paraphrasées en remplaçant le verbe de sentiment par « éprouver le sentiment », ainsi *éprouver de l'étonnement* a le même sens qu'*être étonné*, et *éprouver de l'amour* peut remplacer (dans certains cas) *aimer*. On pourra dire *Luc éprouve de l'étonnement devant la hauteur de la tour Eiffel*, et *Jean éprouve de l'amour pour Marie*. La valeur des prépositions introduites (*devant, pour*) dépend (dans notre description) de la classe du verbe, et de la nature du sujet (un humain, un objet concret, etc.). La phrase *Léa éprouve (du dégoût + de la passion) pour la lecture* est possible, mais les phases suivantes sont plus difficiles : *Léa éprouve du dégoût / de la passion devant la lecture* ou *Jean éprouve de l'amour devant Marie* ou *Luc éprouve de l'étonnement pour la hauteur de la tour Eiffel*. Précisons que cette paraphrase par nominalisation n'est pas toujours possible car certains verbes n'ont pas de nom associé, comme *interloquer* par exemple.

2.1.4. Les sentiments dans un texte : traitements automatiques

Notre objectif, dans nos premiers travaux sur les sentiments, était de faire une description linguistique, formalisée mais sans buts applicatifs. Et, si la réalisation d'un système informatique (*FEELING*, cf description détaillée §3.2.) nous a permis de montrer le bien-fondé, ainsi que leurs limites, des représentations que nous proposions, elle n'était pas le but initial de nos descriptions. La formalisation que nous avons adoptée pour décrire ces connaissances sur les sentiments est dans un format qui nous est propre, constitué de types hétérogènes de représentations. Cette disparité, due à la complexité des données à représenter, rend difficile leur emploi par des logiciels qui ne seraient pas *ad hoc* comme *FEELING*. D'un point de vue applicatif, on pourrait considérer ces connaissances comme des données accumulées mais inertes, « passives », ne pouvant être utilisées pour une quelconque application.

Plusieurs éléments ont contribué à modifier notre perspective :

- un large consensus s'est fait au sein de la communauté linguistique et du TAL pour adopter des formalismes standardisés de représentation des données, comme les ontologies, ce qui permet de mettre les connaissances à la disposition du plus grand nombre. Nous avons pu ainsi représenter une partie de notre description des sentiments sous forme d'une ontologie (cf. description §3.2.3. dans la suite), ce qui rend ces connaissances accessibles aussi bien à des utilisateurs humains qu'à des outils informatiques, nous avons ainsi transformé ces données « passives » en données « actives » ;

- l'augmentation de la capacité de traitement et de stockage des ordinateurs et l'abondance de documents électroniques disponibles dans tous les domaines et dans des formes très variées (articles de presse, romans, forums, blogs, bases de données, etc.) a rendu possible divers traitements de corpus (recherches d'occurrences, vérification d'hypothèses, etc.), au niveau d'un document ou de plusieurs documents ;

- enfin, étant donné la nature du support du texte, qui de papier est devenu numérique, et grâce à l'arrivée d'une nouvelle génération de plateformes informatiques, comme Navitexte (Couto et Minel 2006, 2008, Couto à paraître), qui proposent des outils divers dédiées à ce nouveau support, des problématiques nouvelles sont apparues dans l'étude d'un texte ou de plusieurs textes. En effet, la lecture d'un texte via un écran d'ordinateur n'est pas la même activité que lire un texte sur un support papier et sa version électronique offre de nombreux modes de lecture. Une lecture linéaire, bien sûr, mais également une lecture selon des modalités que le lecteur définit. Pour un roman, par exemple, on peut s'intéresser à des personnages particuliers, se focaliser sur la chronologie des événements, etc.

Pour permettre ces différentes lectures, il faut d'abord enrichir le texte, le « décorer », avec des annotations sur les morceaux de textes (mot, syntagme phrase, paragraphe, etc.) que l'on estime pertinents pour ce que l'on veut étudier, avec des annotations syntaxiques et/ou sémantiques.

Une fois cette annotation faite, on peut, en utilisant NaviTexte par exemple, effectuer divers traitements sémantiques sur ce texte, comme du filtrage avec certaines conditions, ou parcourir ce texte avec des objectifs précis.

C'est ce que nous avons proposé dans le domaine des sentiments, en faisant appel à l'ontologie que nous avons construite pour annoter des expressions de sentiment dans un roman, puis en définissant des parcours de lecture du roman, une navigation textuelle, en fonction de ces sentiments (leur nature, leur force, les personnages qui les éprouvent, etc.)

Le cadre que nous avons adopté est celui développé dans les travaux de Javier Couto et Jean-Luc Minel pour qui la navigation textuelle :

> « telle qu'elle est conçue, considère que l'utilisateur parcourt un texte en exploitant, sans que cette démarche soit pour lui nécessairement explicite, des informations linguistiques qui y sont présentes : des marqueurs discursifs, des groupes nominaux portant sur un thème, des verbes liés à l'information recherchée, des phrases qui comportent une sémantique particulière (phrases conclusives, récapitulatives, définitoires, causales, etc.), l'organisation de différents unités textuelles... » Couto à paraître.

Le texte que l'on veut parcourir doit donc d'abord être annoté, avec des informations à la fois syntaxiques et sémantiques sur les sentiments. Nous montrons un exemple d'annotation ci-dessous.

2.1.4.1. Annotation du texte

Le texte que nous avons annoté est un extrait de *Madame Bovary* de Gustave Flaubert (1857/2001). Les expressions de sentiment sont annotés avec six attributs : la polarité du sentiment exprimé (négative, neutre ou positive), la catégorie syntaxique du mot de sentiment annoté (nom, verbe ou adjectif), la classe sémantique du sentiment, son niveau d'intensité (Intensifieur Faible, Médian ou Fort), la personne qui ressent le sentiment (Experienceur) et son objet ou sa cause (Objet). L'annotation est représentée dans la figure 7.

(1)- Partie I, Chap.IX, p.116.
Elle se sentait, d'ailleurs, plus irritée de lui. [Polarité **Négative**, Catégorie **Verbe**, Classe **Irritation**, Intensifieur **Médian**, Experienceur **Emma**, Objet **Charles**].

(2)- Partie II, Chap.V, p.168.
Elle était amoureuse de Léon. [Polarité **Positive**, Catégorie **Adjectif**, Classe **Amour**, Intensifieur **Médian,** Experienceur **Emma**, Objet **Léon**]

(3)- Partie II, Chap.IX, p.225.
- Oh ! Vous êtes bon ! Dit-elle. - Non, je vous aime voilà tout ! [Polarité **Positive**, Catégorie **Verbe**, Classe **Amour**, Intensifieur **Médian,** Experienceur **Rodolphe**, Objet **Emma**]

(4)- Part.II, Chap.IX, p.234.
C'était Rodolphe. […] ! - Je t'aime ! [Polarité **Positive**, Catégorie **Verbe**, Classe **Amour**, Intensifieur **Médian**, Experienceur **Emma**, Objet **Rodolphe**]

(5)- Part.II, Chap.X, p.242.
Il [Rodolphe] la subjuguait. [Polarité **Positive**, Catégorie **Verbe**, Classe **Fascination**, Intensifieur **Fort**, Experienceur **Emma**, Objet **Rodolphe**]

(6)- Part.II, Chap.X, p.245.
Elle se demanda même pourquoi donc elle exécrait Charles. Polarité **Négative**, Catégorie **Verbe**, Classe **Haine**, Intensifieur **Fort,** Experienceur **Emma**, Objet **Charles**]

(7)- Part.II, Chap.XI, p.258.
Charles […].
-Assieds-toi, dit-elle, tu m'agaces !. [Polarité **Négative**, Catégorie **Verbe**, Classe **Irritation**, Intensifieur **Médian** Experienceur **Emma**, Objet **Charles**]

(8)- Part.II, Chap.XII, p.261.
Plus elle se livrait à l'un, plus elle exécrait l'autre. [Polarité **Négative**, Catégorie **Verbe**, Classe **Haine**, Intensifieur **Fort,** Experienceur **Emma**, Objet **Charles**]

Figure 7. Annotation d'un extrait de *Madame Bovary*

Pour chaque phrase, les résultats de l'annotation sont mis entre crochets. La localisation des extraits dans le roman (Partie, Chapitre, page) est indiquée devant chaque phrase.

Le processus d'annotation est actuellement semi-automatique. Il se fait en plusieurs étapes : dans une première étape, le texte source est soumis à un analyseur syntaxique (Bourigault 2002) qui, associé à des heuristiques « ad hoc », produit une annotation morphosyntaxique et les relations entre les différentes parties du discours comme verbe et arguments ou nom et adjectifs.

Puis, dans une deuxième étape, l'ontologie sera utilisée pour ajouter automatiquement des propriétés, telle que l'intensité du sentiment ou les relations entre les différentes parties du discours, comme la détermination de la personne qui éprouve le sentiment ou de l'objet du sentiment.

Enfin, dans l'étape 3, une annotation manuelle résout certaines difficultés, comme la résolution d'anaphores, et produit le texte final annoté.

La figure 8 montre l'annotation de l'exemple (3) de la figure 7. Dans cet exemple, la partie gauche de la figure indique les différentes étapes appliquées au texte source, la partie

droite indique le texte modifié. Le résultat de l'étape 1 : « **je vous aime**, Pro|je|1| Pro|vous|2| VCONJ|aimer|aime| sujet 1 cod 2 » indique que « je » et « vous » sont des pronoms (Pro), de position 1 et 2 respectivement dans la phrase, « aime » est le verbe conjugué « aimer », dont le sujet est en position 1 et le complément d'objet en position 2.

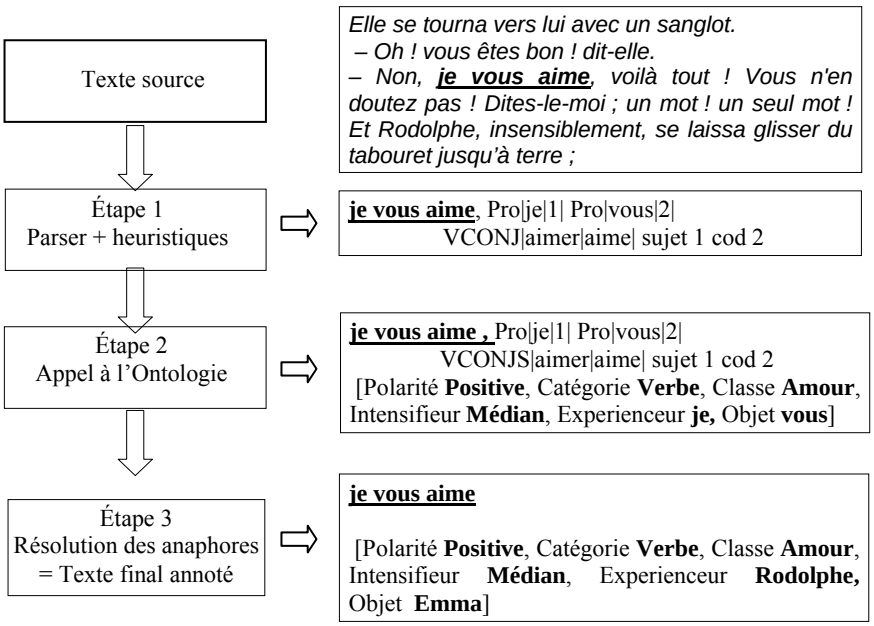


Figure 8. Processus d'annotation

Dans l'étape 2, l'appel à l'ontologie permet d'indiquer la Polarité, la Catégorie, et la classe sémantique de l'expression du sentiment , ici le verbe « aimer », le niveau d'intensité de cet amour (Intensifieur), l'Expérienceur « je **»,** et l'Objet du sentiment « vous **».**

Enfin, dans l'étape 3, l'Expérienceur « je » est remplacé par « Rodolphe », et l'Objet du sentiment « vous **»** est remplacé par « Emma ».

La première phase pour l'annotation sémantique des sentiments dans un texte est le repérage des expressions de sentiment. Ensuite, plusieurs solutions sont possibles :

1) Indiquer qu'on a repéré un sentiment, mais n'introduire aucune autre information. Cela suppose un appel à l'ontologie chaque fois qu'on s'intéressera à un critère particulier. D'un point de vue informatique, il y a une économie de la représentation, mais une dégradation des performances due à des appels répétés à l'ontologie ;

2) indiquer qu'on a repéré un sentiment et ajouter toutes les informations le concernant qui sont contenues dans l'ontologie. Le texte annoté contient ainsi toutes les connaissances connues, et donc il n'y aura qu'un appel à l'ontologie. Cette solution est très performante,

néanmoins elle présente le désavantage de la perte de lisibilité du texte et d'une grande redondance dans les annotations ajoutées ;

3) une solution médiane : indiquer qu'on a repéré un sentiment et inclure, par un appel à l'ontologie, un certain nombre d'informations que l'on estime pertinentes pour répondre à la majorité des interrogations du lecteur. Cela permet de ne faire appel ensuite à l'ontologie que dans certains cas particuliers. C'est l'option que nous avons choisie dans l'exemple précédent. Si, pour un sentiment repéré (l'amour dans cet exemple), on s'intéresse (uniquement) à sa catégorie grammaticale, sa classe sémantique, à la polarité, aux personnages qui éprouvent le sentiment, à l'objet et à l'intensité de ce sentiment, l'annotation du texte est suffisante. En revanche, si l'on veut étudier les sentiments proches sémantiquement mais de force plus ou moins grande, les synonymes, antonymes, etc., un nouvel appel à l'ontologie est nécessaire.

2.1.4.2. Navigation textuelle

Comme nous l'avons expliqué, cette annotation des sentiments est utilisée pour la navigation à travers un texte avec des critères spécifiés.

Plusieurs parcours de lecture sont possibles à travers les exemples (1) à (8) de la figure 7. À partir du début du roman, un parcours qui s'intéresse à ce « que ressent Emma pour Charles ? » suivra les séquences 1, 6, 7, 8; un autre « que ressent Emma pour Rodolphe ? » commence à la phrase 4, puis va à la phrase 5. Pour représenter ces différents parcours, nous avons implémenté le roman annoté dans la station de travail informatique NaviTexte. La figure 9 représente l'écran qui affiche l'extrait annoté de la figure 7.

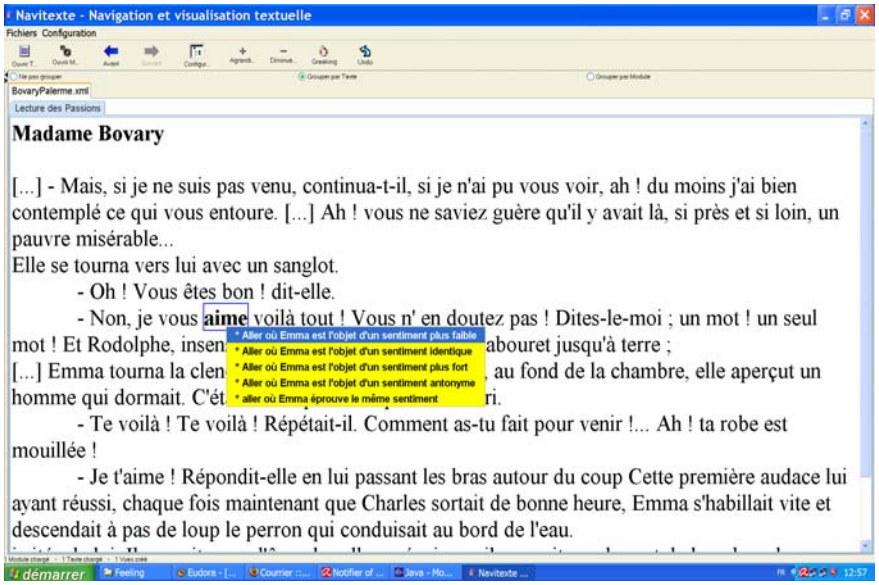


Figure 9. Exemple d'écran *NaviTexte*

Ici, le lecteur s'intéresse à un sentiment particulier : l'amour. À partir d'une phrase où Emma est l'objet d'un amour (ici celui de Rodolphe), 5 parcours sont possibles dans la suite du roman. Ces parcours sont indiqués dans le rectangle bleu et jaune associé au verbe « aime » (en gras dans le texte) et reproduit ci-dessous pour une meilleure lisibilité :

- Aller où Emma est l'objet d'un sentiment plus faible
- Aller où Emma est l'objet d'un sentiment identique
- Aller où Emma est l'objet d'un sentiment plus fort
- Aller où Emma est l'objet d'un sentiment antonyme
- Aller où Emma éprouve le même sentiment

Le lecteur peut par exemple souhaiter aller vers une phrase (1) où Emma est l'objet d'un sentiment plus faible, (2) aller où Emma est l'objet d'un sentiment (amour) identique, (3) aller où Emma est l'objet d'un sentiment plus fort, (4) aller où Emma est l'objet d'un sentiment antonyme (comme la haine), ou enfin (5) aller vers une phrase où c'est Emma et non plus Rodolphe, qui éprouve un sentiment d'amour. Dans cet exemple, la navigation proposée est vers l'avant, mais il est possible aussi de combiner différentes conditions (intensité, expérienceur, etc.) avec l'orientation souhaitée (en avant ou en arrière).

2.1.4.3. Filtrage sémantique

L'annotation des sentiments dans un texte permet également de définir des filtres sémantiques sur ce texte. Nous citons Jean-Luc Minel (à paraître) pour qui le filtrage sémantique :

> « se concrétise par l'existence d'un certain nombre d'éléments :
>
> - une représentation unique et abstraite d'un texte [...]. Un langage symbolique de description de ces unités doit aussi être proposé pour exprimer cette représentation.
>
> - un jeu d'annotations qui peuvent être associées aux unités et aux relations qui composent le texte.
>
> - une sémantique associée à ces annotations, celle-ci pouvant prendre la forme d'une description formelle comme dans une ontologie ou plus descriptive comme celle d'une représentation visuelle interprétable par un lecteur humain.
>
> - un langage qui permette de construire un filtre, c'est-à-dire qui possède une puissance d'expression nécessaire à l'élaboration de conditions de filtrage sur les éléments annotés du texte ainsi que sur leurs relations, elles aussi éventuellement annotées.
>
> - des spécifications des représentations visuelles qui seront présentées au lecteur sous forme d' « écrit d'écran », [...].
>
> - un langage qui permette de formaliser la navigation textuelle [...] ces connaissances incluent la modélisation des besoins spécifiques d'un lecteur pour un type de support d'affichage donné. »

Certaines des propriétés que avons étudiées sont utilisables pour l'annotation et le filtrage des sentiments dans les textes (que l'on peut considérer comme un sous-ensemble des annotations) et permettent par exemple de :

1. reconnaître l'expression d'un sentiment, même si le mot de sentiment est employé avec un sens métaphorique comme dans la phrase *Marie a brisé le cœur de Luc* qui indique une profonde souffrance morale de *Luc*, alors que dans *Marie a brisé les doigts de Luc,* ce dernier ressent une souffrance physique. Il faut également repérer les expressions comme « *en avoir gros sur l'estomac* » (être très triste, avoir beaucoup de chagrin) ou « *avoir les foies* » (avoir peur), *voir rouge* (être en colère) ;

2. comprendre et analyser sémantiquement cette expression : est-ce de l'amour, de la peur, de la joie, etc. ? Le contexte sera utilisé pour résoudre un problème que l'on rencontre fréquemment, celui de la polysémie. En effet, un même item lexical peut référer à plusieurs sentiments, par exemple le verbe *ennuyer* qui a trois interprétations possibles : *(i)* causer du souci (classe **Tracas**) comme dans *Les mensonges de son fils ennuient/tracassent Paul (ii)* remplir d'ennui, lasser (classe **Ennui**), par exemple dans *Le discours fleuve de l'orateur ennuie/rase l'assistance* ou *(iii)* importuner, causer du dérangement (classe **Dérangement**) qui est son sens dans la phase *Le bruit assourdissant des pelleteuses (ennuie/dérange) Paul* ;

3. déterminer quel est l'opposé et éventuellement la décomposition du sentiment exprimé, par exemple *l'émerveillement* est un mélange d'admiration et d'étonnement (Le Petit Robert 2001) ;

4. reconnaître qui éprouve le sentiment et quel est son objet. Par exemple dans le cas d'une phrase avec un verbe de sentiment, comme nous l'avons vu précédemment, la reconnaissance du sujet et du complément du verbe n'est pas suffisante, car si dans *Jean aime Marie,* c'est effectivement le sujet *Jean* qui éprouve de l'amour pour le complément *Marie*, dans la phrase *Les cris de Marie effrayent Luc,* le sujet du verbe est *Les cris* qui sont l'objet de la frayeur de *Luc*, c'est donc ce dernier qui éprouve un sentiment de frayeur, alors qu'il est le complément du verbe *effrayer* ;

5. repérer que la force d'un sentiment est amplifiée ou diminuée, et évaluer le degré plus ou moins fort de cette intensité. Comme nous l'avons vu, la force des sentiments peut être indiquée de plusieurs façons : par l'ajout d'adverbe (*peu*, *très*), par l'emploi d'un mot de même sémantique mais avec une intensité plus faible ou plus forte par exemple *peur* vs *terreur,* ou par des expressions qui, le plus souvent, expriment une très grande force du sentiment comme *blêmir*/*être vert de peur,* ou *écumer de rage*.

Comme nous l'avons souligné précédemment, la perspective de nos recherches s'est déplacée au cours du temps et nous sommes passée de travaux orientés « description » à des travaux orientés « application » (sans pour autant négliger la description). Cette évolution nous a amené à travailler sur le texte. Cet intérêt, ajouté à celui de mesurer la

généricité de nos représentations, nous a conduite d'une part à compléter notre ontologie des sentiments par une description des actes de paroles comme *vociférer* : « Parler en criant et avec colère » (Le Petit Robert 2001) ou *fulminer* : « Se laisser aller à une violente explosion de colère, se répandre en menaces, en reproches » (Le Petit Robert 2001), et d'autre part à élargir le champ sémantique de nos recherches à un domaine voisin, celui des opinions.

2.2. LES OPINIONS

Nous proposons une étude des expressions d'opinion dans les textes, en français et en anglais. Ce travail est en cours d'élaboration, il est fait en collaboration avec Nicholas Asher et Farah Benamara (Asher *et al.* 2008). Nous avons abordé cette étude avec une méthodologie qui était voisine au départ de celle adoptée dans l'étude des sentiments car nous avons recherché des régularités de sens entre les expressions d'opinions et pris en compte certaines constructions afin d'identifier des classes sémantiques homogènes. L'approche est légèrement différente cependant car l'objectif étant une analyse et une évaluation globales des opinions dans un (ou plusieurs) texte, nous nous sommes intéressés aux différents éléments qui concourent à leur évaluation, et en particulier, à la façon dont les propositions contenant les expressions d'opinion sont reliées entre elles à l'intérieur d'un discours.

De nombreux travaux cherchent à établir la part de subjectivité ou d'objectivité d'un texte (Pang *et al.* 2002) (Turney 2002), à déterminer la polarité et la force des opinions exprimées dans un document (Yu et Hatzivassiloglou 2003) (Kim 2004), (Wilson *et al*. 2006) et/ou repèrent celui qui exprime l'opinion (Bethard *et al.* 2004), (Kim et Hovy 2005), (Choi *et al.* 2005). Certains, comme (Whitelaw *et al.* 2005], utilisent la classification des attitudes et des engagements de (Martin et White 2005). L'approche du projet MPQA (Multi-Perspective Question Answering), développé par Janice Wiebe *et al.* à l'Université de Pittsburg (Wiebe *et al*. 2005], est différente car leur objectif est de fournir un large corpus de textes de journaux annotés manuellement, en repérant un ensemble d'expressions relevant de la sphère privée (opinions, sentiments, émotions, évaluations, spéculations) et la source de ces expressions.

2.2.1. Le mot

Il y a de très nombreuses définitions de la notion d'opinion, comme par exemple :

> « Jugement, avis, sentiment qu'un individu ou un groupe émet sur un sujet, des faits, ce qu'il en pense » Larousse
>
> « - Jugement, en bien ou en mal, qu'on porte d'une personne ou d'une chose.
>
> - Sentiment qu'on se forme des choses. » Littré
>
> « Manière de penser sur un sujet ou un ensemble de sujets, jugement personnel que l'on porte sur une question, qui n'implique pas que ce jugement soit obligatoirement juste. » TLFi
>
> « Manière de penser, de juger; attitude de l'esprit qui tient pour vraie une assertion; assertion que l'esprit accepte ou rejette (généralement en admettant une possibilité d'erreur). » Petit Robert

Plutôt qu'en adopter une précise, nous avons préféré étudier comment un contenu intellectuel ou affectif sur un sujet donné est explicitement et lexicalement exprimé.

Notre approche pour catégoriser les opinions utilise les recherches en sémantique lexicale de Beth Levin que nous avons citées au §1.2., et celles d'Anna Wierzbicka (1987).

Anna Wierzbicka classe approximativement 270 verbes anglais qui se réfèrent à des actes de paroles répartis en 37 classes selon leur signification. Cette classification est donnée dans le tableau 10.

Accuser	Admettre	Affirmer	Appeler	Attaquer	Avertir	Baptiser
Blâmer	Conclure	Confirmer	Conseiller	Déclarer	Demander	Deviner
Dire	Discuter	S'exclamer	Informer	Interdire	Interroger	Laisser Entendre
Louer	Se moquer	Offrir	Ordonner	Pardonner	Parler	Permettre
Se Plaindre	Promettre	Récapituler	Remarquer	Remercier	Répondre	Réprimander
Souligner	Soutenir					

Tableau 10. Classes des verbes de parole anglais par Anna Wierzbicka (1987)

Nous avons classé les expressions d'opinion en quatre catégories :

1) les expressions de Reportage qui permettent de relater ou d'introduire les opinions des autres ou les siennes propres, et qui fournissent une évaluation du degré d'implication à la fois de la personne qui exprime l'opinion et de son objet, comme le verbe *estimer* dans la phrase « *les routiers français estiment souffrir d'une fiscalité désavantageuse comparé à leurs rivaux européens* »[10] ;

2) les expressions de Jugement, qui expriment des évaluations normatives d'objets et d'actions, à l'intérieur desquelles on peut distinguer des jugements reliés aux normes sociales, par exemple les verbes *approuver* et *critiquer* dans « *Laurence Parisot approuve la réforme mais critique la méthode* »[11], et des jugements reliés à des normes personnelles comme « *C'est un pur chef d'œuvre* »[12] ;

3) les expressions de Conseil, qui enjoignent de faire ou penser quelque chose, par exemple « *À voir et même revoir*»[13] ;

4) les expressions de Sentiment, qui expriment un sentiment ou une émotion ressentie par une personne, comme « *J'ai adoré ce film* »[14].

Cette classification est représentée dans le tableau 11. Dans la catégorie ***Reportage*** les opinions sont le plus souvent exprimées dans les compléments des verbes utilisés pour rapporter un discours ou les opinions des autres. Ces verbes transmettent le degré d'implication de celui qui émet l'opinion pour l'opinion qu'il présente, et certains fournissent, au moins indirectement, un jugement de celui-ci. La polarité de l'opinion

[10] Libération, 16 juin 2008
[11] Libération, 10 juin 2008
[12] Allocine.fr, 8 mai 2008
[13] Télérama.fr, 21 août 2008
[14] Télérama.fr, 9 janvier 2008

(positive, négative ou « neutre ») est donnée par les compléments du verbe ou le contexte. Cette catégorie est constituée de trois groupes selon le degré d'implication et de véracité concernant l'information contenue dans les compléments : **Informer/Soutenir, Dire/Remarquer** et **Penser/Supposer.**

Catégories	Classes	Exemples
Reportage	Informer	*informer, notifier, révéler*
	Soutenir	*insister, prétendre, soutenir*
	Dire	*dire, annoncer, rapporter*
	Remarquer	*commenter, observer, remarquer*
	Penser	*penser, considérer*
	Supposer	*présumer, supposer, imaginer*
Jugement	Blâmer/Louer	*blâmer, critiquer, condamner, louer, approuver*
	Appréciation	*bon, brillant, mauvais, nul*
Conseil	Recommander	*recommander, déconseiller*
	Suggérer	*suggérer, proposer*
	Espérer	*espérer*
Sentiment	Colère/Apaisement	*colère, irritation, apaiser*
	Étonnement	*étonner, stupéfier*
	Amour/Fascination	*fasciner, captiver*
	Haine/Déception	*démoraliser, décevoir*
	Peur	*effrayer, peur*
	Offense	*choquer, outrer*
	Tristesse/Joie	*heureux, triste*
	Ennui/Divertissement	*ennui, distraction*
	Émouvoir	*bouleverser, émouvoir*

Tableau 11. Classification des expressions d'opinion

Le premier groupe **Informer/ Soutenir** est constitué de deux classes de verbes qui introduisent une information que l'auteur estime vraie et établie (classe **Informer**) ou pour laquelle l'émetteur de l'opinion est fortement engagé (classe **Soutenir**). Les verbes de la classe **Informer** communiquent une information au moyen de leurs présuppositions que l'auteur prend pour établies. Les verbes de ce groupe (*révéler, ...*) présupposent la véracité de leurs compléments. Ainsi, il serait incohérent de dire « *Paul a révélé qu'il était malade mais il n'était pas malade* », alors qu'il ne l'est pas d'utiliser un verbe de la classe **Soutenir** (*soutenir, prétendre*, etc.) et de dire « *Paul soutient/prétend qu'il était malade mais il n'était pas malade* ». La véracité de l'information introduite par les verbes de la

classe **Informer** est plus forte que celle introduite par ceux de la classe **Soutenir**. D'autre part, les verbes de la classe **Informer**, contrairement à ceux de **Soutenir,** n'impliquent pas nécessairement un fort engagement de la part celui qui émet l'opinion pour son contenu, il peut révéler quelque chose inconsciemment par exemple.

Le second groupe **Dire/Remarquer** contient également deux classes : **Dire** et **Remarquer**. Contrairement aux verbes d'**Informer**, ces verbes n'indiquent rien sur le point de vue ou sur l'engagement de l'auteur au sujet de l'opinion émise. La distinction entre ces deux classes porte sur le statut de l'information transmise par l'opinion. Les compléments des verbes de la classe **Remarquer** sont un discours subsidiaire dans le texte. Ils fournissent le contexte, l'arrière plan ou des commentaires.

Enfin, le dernier groupe **Penser/Supposer** introduit une opinion avec un certain degré de subjectivité. L'auteur exprime sa pensée sur quelque chose sans être sûr de sa véracité. Les verbes de la classe **Penser** expriment le fait que l'auteur a un fort engagement pour l'opinion qu'il introduit (le complément du verbe) alors que les verbes de la classe **Supposer** expriment un engagement plus faible. La véracité de l'information introduite par des verbes de **Penser** est plus forte que l'information introduite par des verbes de **Supposer** et le lecteur aura plus confiance concernant P dans « *Il pense que P* » plutôt que dans « *Il imagine que P* ».

La seconde catégorie ***Jugement*** contient des mots qui expriment une estimation de quelque chose ou quelqu'un. Cette catégorie contient des noms, des verbes et des adjectifs. Les expressions de jugement transmettent une polarité binaire de l'opinion (positive ou négative). Nous avons considéré deux classes: les jugements se référant à un système de normes sociales et les jugements se référant à un système de normes personnelles. La classe **Blâmer/Louer** exprime un jugement moral sur la responsabilité ou le comportement de quelqu'un, alors que la seconde classe **Appréciation** implique une appréciation positive ou négative sur quelqu'un ou sur quelque chose.

Les expressions de la troisième catégorie ***Conseil*** enjoignent le lecteur à adopter une certaine voie, un certain type d'action ou d'opinion. Celui qui émet l'opinion se place en position d'autorité par rapport au lecteur et renforce son avis avec d'autres arguments pour le convaincre. Nous avons considéré trois classes : les verbes de la classe **Recommander** qui expriment une bonne ou mauvaise opinion et essayent de convaincre avec force; les verbes de la classe **Suggérer** qui sont utilisés par l'émetteur d'opinion pour dire ce qu'il suggère ou spécule sans en être absolument certain; et enfin, les verbes de la classe **Espérer** qui expriment le sentiment général que certains désirs seront satisfaits. Les expressions d'opinion de **Recommander** sont plus fortes que celles de **Suggérer** et **Espérer**, ces dernières étant les plus faibles.

Les mots de la dernière catégorie, ***Sentiment,*** expriment une attitude par rapport à quelque chose, généralement basée sur le sentiment ou l'émotion plutôt que sur le raisonnement. Ils ont une polarité et une force. Nous distinguons ici entre des sentiments positifs, exprimés par les mots des classes (ou sous-classes) **Apaisement, Divertissement,**

Joie, Amour et **Fascination,** des sentiments négatifs exprimés par les mots des classes **Colère, Ennui, Offense, Tristesse, Peur, Haine** et **Déception**, et des sentiments de polarité « neutre » dans les classes **Étonnement** et **Émouvoir**.

2.2.2. Le segment et la phrase : Relations rhétoriques et évaluation des opinions

La structure rhétorique est un élément important pour comprendre les opinions transmises par un texte. L'analyse de différents corpus a montré que les relations de discours en modifient le sens et la force. Considérons, par exemple, les textes suivants :

T1: [*je suis d'accord avec vous*]a *même si je suis choqué*.

T2: *Achetez le DVD,* [*vous ne le regretterez pas*]b

Les opinions dans T1 et T2 sont positives mais le contraste introduit par l'adverbe *même* en T1 diminue la force de l'opinion exprimée dans (a), alors que l'explication donnée fournie par (b) en T2 augmente la force de la recommandation.

2.2.2.1. Annotation des segments d'opinion

Nous avons utilisé nos quatre catégories d'opinion pour étiqueter les expressions d'opinion à l'intérieur d'un segment de discours. Nous définissons un segment de base comme une proposition contenant une expression d'opinion ou une séquence de propositions (reliées à un segment exprimant l'opinion par une relation de discours) exprimant une relation rhétorique avec un segment exprimant une opinion. Les segments complexes sont composés de segment de base reliés par des relations de discours. Les segments sont alors connectés entre eux en utilisant un petit ensemble de relations de discours (Asher et Lascaridès 2003). Nous avons segmenté des propositions conjointes en propositions séparées, par exemple, *le film est beau et puissant* est considéré comme deux segments : *le film est beau* et *le film est puissant*.

Actuellement l'annotation est faite manuellement. Pour chaque segment d'opinion, les informations que nous indiquons sont : la catégorie sémantique à laquelle il appartient, la modalité associée, celui qui émet l'opinion, le sujet ou thème de l'opinion, et le déclencheur lexical qui a permis d'identifier le segment d'opinion. La modalité dépend de la catégorie sémantique de l'expression d'opinion : elle exprime le degré de préférence, ou d'engagement (avec sa force) ou une polarité (avec sa force) ; Nous la décrivons plus en détail au §3.3.

2.2.2.2. Relations rhétoriques

Une relation rhétorique entre deux segments a and b est notée **NomRelation**(a,b). Nous avons utilisé quatre types de relations rhétoriques **Contraste**, **Correction**, **Explication**/**Élaboration** (réunies dans une relation appelée **Support**) et **Résultat**. Les relations **Contraste** et **Correction** indiquent une différence d'opinion ; **Contraste**(a, b)

implique que a et b sont vrais tous les deux, mais certaines implications de l'un sont contredites par l'autre, alors que **Correction**(a, b) dénote une opposition forte et implique que b est vrai quand a est faux. Pour repérer les relations de **Contraste** dans les textes, nous utilisons des marqueurs de discours spécifiques, tels que *néanmoins*, *bien que*, *mais*, *même si*, etc. et les marqueurs *contredire, rétorquer, objecter, démentir, protester, nier*, etc. pour les relations de **Correction**. La relation **Support** est la fusion des relations appelées généralement **Explication** et **Élaboration,** car nous avons considéré que ces relations sont toutes deux utilisées pour soutenir les opinions. Les relations **Explication**(a, b), marquées dans les textes avec *parce que*, *car*, etc. indiquent que b fournit une raison (suffisante) pour a. La relation **Élaboration**(a, b), marquée avec *par exemple, en particulier*, etc. implique que b donne plus de détails sur ce qui est exprimé dans a. La relation **Résultat**(a, b), notée par des marqueurs comme *donc, par conséquent ainsi*, etc., indique que b est une conséquence ou un résultat de a. Enfin, **Continuation**(a, b) signifie que a et b forment une partie d'un tout, d'une thématique plus large, et relient deux constituants à l'intérieur de la même phrase. Ainsi, il y a trois segments d'opinion dans chacune des phases suivantes :

T3: *Même [si le produit est excellent]a, [le design est très basique]b, [ce qui est décevant dans cette marque]c.*

T4: *[[C'est bouleversant]d, [triste]e]g et [beau]f en même temps.*

Dans T3, il y a une relation **Contraste** entre a et b, ce contraste introduit le segment c. La structure rhétorique de la phrase est **Résultat**(**Contraste**(a,b),c). Dans T4, il y a une **Continuation** entre d et e, et un **Contraste** entre g et f. La structure rhétorique est **Contraste**(**Continuation**(d, e), f).

2.2.3. Le texte : Évaluation et Annotation des expressions d'opinion

De nombreux éléments doivent être pris en compte pour faire une évaluation globale des opinions exprimées dans un texte, comme sa structure rhétorique ou le mode des verbes. Nous avons pu le vérifier en analysant des textes électroniques de genres différents, en français et en anglais.

2.2.3.1. Évaluation des opinions dans un texte

Une fois les structures de traits établies (cf. §3.3.1.), nous recherchons les relations de discours qui indiquent comment sont combinés les différents segments.

Dans certains cas, la relation **Contraste** modifie la force de l'opinion alors que **Continuation** renforce sa polarité.

Par exemple, dans la phrase T4 déjà vue : [[*C'est bouleversant*]d, [*triste*]e]g *et* [*beau*]f *en même temps*, le segment d appartient à la classe **Émotion** de la catégorie ***Sentiment*** qui

a une polarité neutre, alors que le segment e appartient à la catégorie **Tristesse** de ***Sentiment*** qui a une polarité négative. Un ensemble de règles dédiées est utilisé pour combiner les structures de traits de plusieurs segments et permettent d'établir que la polarité du segment g est négative. Ces règles sont en cours d'élaboration et ne sont pas aisées à établir. En effet, est-il possible de déterminer si c'est la tristesse ou l'émotion qui prédomine quand on écrit « *bouleversant, triste* » ? De nombreux éléments entrent en compte, comme l'ordre des mots qui influe peut-être.

La relation **Résultat** indique la polarité ou l'opinion dans le second argument. Pour la relation **Contraste**, nous avons distingué deux cas : si les deux arguments sont des segments d'opinion, alors le **Contraste** affaiblit la polarité du premier argument : *J'ai aimé ce film mais il était un peu long* ; si l'un des arguments est en relation rhétorique avec l'autre argument, alors le **Contraste** renforce la polarité de l'opinion, comme dans [[*Je suis athée*]*, mais* [*je soutiens ce curé totalement*]][15].

Nous avons également pris en compte les impératifs, les conditionnels et les négations pour l'évaluation des opinions dans un texte. Les négations sont traitées en inversant la polarité de l'expression de l'opinion dans son périmètre (champ, portée). Les conditionnels sont complexes à interpréter. Dans certains cas, pour les opinions relevant du ***Conseil*** par exemple, ils peuvent inverser l'opinion. Ainsi la phrase *Si vous voulez gaspiller votre argent, achetez ce film* ne sera pas interprétée comme recommandant l'achat de ce film, mais au contraire comme recommandant de ne pas l'acheter. Mais les conditionnels peuvent également renforcer la recommandation comme dans *Si vous voulez passer un bon moment, Allez voir ce film*.

Une représentation discursive du texte est faite à l'aide de graphes, que nous décrivons au §3.3.2.

2.2.3.2. Annotation des expressions d'opinion dans des textes

De façon à vérifier le bien-fondé de notre approche, nous avons étudié la distribution des expressions d'opinion dans des textes de genres différents, en anglais et en français, puis nous avons déterminé la structure rhétorique des segments composant ces textes. Nous avons annoté 3 types de corpus en ligne, chacun de style et d'audience différents : critique de films par des internautes, courrier des lecteurs et articles de presse. Nous avons sélectionné au hasard 40 articles de chaque catégorie et annoté près de 500 segments pour le français ; pour l'anglais 15 articles ont été étudiés et 170 segments annotés. Cette annotation a été faite manuellement, notre objectif est de la rendre automatique.

Pour chaque corpus, un segment a été annoté seulement s'il contient explicitement une expression d'opinion, c'est-à-dire un mot qui appartient à notre lexique, ou si ce segment contient une relation rhétorique qui le relie à un segment qui contient une expression

[15] La Dépêche du Midi, 25 avril 2007

d'opinion. Quand un segment est identifié, l'annotateur définit sa représentation sémantique telle que nous l'avons décrite, puis relie les segments de discours en utilisant l'ensemble des relations rhétoriques que nous avons identifiées.

L'annotation de critiques de film a été très aisée. Le style d'écriture est simple et bref, généralement un paragraphe, et les expressions d'opinion sont principalement des noms et des adjectifs. L'exemple qui suit en est assez représentatif : « *J'ai adoré ce film. Denzel Washington en flic calme et zen est parfait, Owen en gangster des temps moderne, froid et impénétrable est formidable, quant à Jodie Foster, en mercenaire très classe, elle est fascinante. Ce film est très bien réalisé, il a de l'action et le rythme est dans l'ensemble soutenu (une ou deux longueurs cependant). Spike Lee joue parfaitement bien sur la psychologie de ses personnages, mystère et une touche d'humour sont au rendez vous! A voir absolument!!!* »[16]. Nous avons trouvé en moyenne 5 segments par critique. Les chroniqueurs résument leur opinion générale, puis donnent leur opinion sur chaque point particulier comme la mise en scène, le jeu des acteurs, etc. et concluent leur critique avec des expressions de recommandations (« *allez voir ce film* ») et quelquefois des expressions de suggestion (« *ce film devrait gagner un oscar* »). Les opinions exprimées sont essentiellement des sentiments et des appréciations, aussi bien en français qu'en anglais.

Le Courrier des lecteurs contient souvent un titre qui introduit la lettre puis un mélange de faits et d'opinion, sur plusieurs paragraphes, comme dans cet exemple : « *NADAL, le rouleau compresseur. Pour ma part, je pense que Rafaël Nadal est aujourd'hui invincible sur terre battue. Pour commencer, ce jeune espagnol a été impressionnant tout au long de l'année dernière sur terre battue, de par sa qualité physique et un mental d'acier.[...] Seule la pression pourrait mettre un terme à cette série ou un Federer des grands jours* »[17]. Les mots sont aussi bien des noms, des adjectifs que des verbes. Comme pour les critiques de films, les opinions sont principalement des appréciations et des sentiments. Par contre, les lettres en anglais contiennent plus d'expressions de jugement en fonction de normes sociales et moins d'expressions de sentiment qu'en français. La distribution des jugements**,** appréciation, espoirs et reportage est la même en français et en anglais.

Enfin, les opinions dans la presse contiennent essentiellement des discours rapportés. Les articles sont longs et composés de plusieurs paragraphes. Comme nous n'avons annoté que les segments contenant ou en relation, via les relations rhétoriques, avec des expressions d'opinion, notre annotation couvre une faible partie du document. Les articles de presse sont les plus difficiles à annoter car l'écriture est plus libre, et ils contiennent beaucoup de structures introduites par des verbes de **Reportage**, ainsi que de nombreuses négations. Il y a une grande similitude de distribution des catégories en français et en anglais : beaucoup d'expressions de jugement et peu de sentiments.

[16] Allociné.fr, 4 mai 2006

[17] La Dépêche du midi, 27 avril 2007

La comparaison des différentes annotations faites pour un même corpus a montré qu'il y avait un large consensus pour la détermination des segments et la répartition des expressions d'opinion dans les catégories. Les principales sources de divergence entre annotateurs ont été dans le choix de certaines relations rhétoriques et dans la distinction entre les jugements relatifs aux normes sociales (classe **Blâmer/Louer**) et les jugements relatifs aux normes personnelles (classe **Appréciation**). Ces désaccords montrent la complexité de l'étude des opinions et la nécessité de prendre en compte d'autres phénomènes. Nous en étudions certains comme par exemple l'emploi de verbes modaux, du conditionnel ou du subjonctif ainsi que la présence d'adverbes qui peuvent modifier l'intensité ou la force de l'opinion émise comme dans « *J'ai **vraiment** adoré ce film* », « *Je vous le recommande **vivement*** ». ou « *Un film **absolument** génial* ».

Nous poursuivons notre travail de recherche afin de produire une annotation automatique des expressions d'opinion, et fournir une évaluation de l'opinion globale d'un texte. Ces annotations permettront ainsi de se déplacer dans un texte, ou dans des textes, selon des critères choisis par exemple l'opinion elle-même, la personne qui l'émet, ou encore son sujet.

Cependant de nombreuses difficultés sont encore à résoudre. D'une part, la prise en compte d'autres éléments que les mots comme constituants signifiants dans un texte. De nombreux auteurs comme Edgar Poe, Gustave Flaubert ou André Breton ont utilisé les italiques ou les guillemets pour donner un sens particulier à un mot[18]. Dans le cas de l'expression des opinions, une modification de leur intensité et de la conviction de celui qui l'émet sont données par l'utilisation de la ponctuation, de la casse des caractères, ou la graphie d'une prosodie. On le voit dans les critiques de films avec l'emploi de majuscules ou de points d'exclamation qui renforce la recommandation dans « *Je vous le recommande VIVEMENT* »[19], « *Excellent film [...] Surtout ne le ratez pas !!!* »[20], et dans la répétition de mots ou le découpage en syllabes avec des majuscules pour refléter une intonation dans « *J'ai A-DO-RE ! Tout tout tout! Le scénario, la mise en scène, le montage, les acteurs, le suspens! Excellent!* »[21].

D'autre part, les opinions ne sont pas toujours exprimées de façon littérale, le recours à l'humour et à l'ironie par exemple, est fréquent. Ainsi, interprétée dans son contexte, la phrase : « *Ce projet est une vraie réussite »* signifie que ce projet est, en fait, un vrai fiasco, et « *On peut remercier le Conseil régional d'avoir voté ce texte »* exprime une opinion très négative sur ce vote qui s'avère une grosse erreur.

[18] Le lecteur intéressé pourra se reporter à l'étude de Julien Gracq sur l'emploi des italiques dans l'œuvre d'André Breton [GRA 72]

[19] Allociné.fr, 4 mai 2006

[20] Allociné.fr, 29 mai 2006

[21] Allociné.fr, 1er novembre 2006

2.3. BILAN COMPARATIF DE NOS DESCRIPTIONS AVEC CELLES DU LEXIQUE GRAMMAIRE, DE BETH LEVIN ET DE WORDNET

2.3.1. Lexique grammaire

Nous sommes partie du cadre théorique du Lexique grammaire pour notre étude des verbes de sentiment (et en particulier pour les verbes « psychologiques »), mais nous nous en sommes éloignée par l'importance que nous avons accordée aux descriptions sémantiques, objet central de nos travaux. Nous avons constaté que la description des tables ne rend pas compte du lien sémantique (évident pour le locuteur) entre des verbes qui n'ont pas nécessairement les mêmes propriétés syntaxiques et, pour cette raison, nous avons adopté le principe heuristique inverse : que donne une description syntaxique délimitée au préalable par l'intuition sémantique ? Nous avons regroupé les verbes en classes sémantiquement homogènes puis nous avons étudié un ensemble de propriétés qui décrivent des phénomènes syntaxiques et sémantiques, et recherché des régularités entre ces propriétés et les classes.

Nous avons gardé quelques propriétés des verbes de la table 4, que nous avons renommées pour certaines d'entre elles. Les propriétés dont nous avons conservé le libellé sont : la possibilité (ou non) de la construction pronominale *N1 se V de ce que P*, l'existence ou non d'un adjectif relié morphologiquement au verbe par une suffixation en *–able (Adjectif = able)* ou une suffixation en *–ent (Adjectif = ant)*, et la construction passive N_1 *est Vpp de ce que P*. Nous avons renommé la possibilité du passif introduit par la préposition *par [passif par]* par **«** N_1 *est Vpp par* N_0 », et de celui introduit par la préposition *de [passif de]* par **«** N_1 *est Vpp de* N_0 ». Enfin, nous avons appelé « emploi non psychologique » la propriété *V concret* qui indique si un verbe est ou non un emploi métaphorique d'un verbe concret.

Pour certains verbes, nous avons des acceptabilités différentes à certaines propriétés. C'est le cas par exemple pour le verbe *choquer* dont le passif en *de* est notée inacceptable alors que nous l'acceptons, acceptabilité que nous avons vérifiée grâce à l'accès à des corpus : *« ... et j'avais été presque* ***choqué de*** *la violence de son ton »*[22].

C'est le cas également quand un verbe a différentes interprétations, et apparait alors dans différentes classes dans notre classification, mais accepte la même construction syntaxique *Qu P V* N_1 et donc fait partie de la table 4. Ainsi le verbe *charmer* qui a deux sens (Le Petit Robert 2001) (1) « Captiver par un attrait puissant; plaire par son charme. » et (2) « Causer une grande joie à ». Les constructions passives *[passif de]* et N_1 *est Vpp de ce que P* sont jugées inacceptables dans la table 4, ce qui correspond à l'emploi (1) alors que l'emploi (2) accepte ces passifs comme le montrent ces deux extraits : « *La bergère*

[22] PROUST Marcel, *À la recherche du temps perdu. 19. Le Temps retrouvé*, 1922, p. 737.

charmée de *ses bontés »*[23]*, « Camarade, je suis* ***charmée de*** *ce que vous venez de me dire »*[24].

2.3.2. Les verbes de Beth Levin

L'étude de Beth Levin sur les verbes a été faite dans une perspective purement linguistique, avec pour objectif la mise en évidence d'une correspondance entre modèles syntaxiques et modèles sémantiques. Notre perspective (le traitement automatique des langues) et nos objectifs (une description sémantique détaillée et formalisée) sont très différents dans notre étude sur les sentiments. Néanmoins, notre approche est voisine sur plusieurs aspects. Notre hypothèse de départ est la même, nous considérons que le comportement syntaxique d'un verbe est guidé par son sens et notre classification présente quelques similarités avec la sienne. En effet, nous avons distingué deux grandes catégories de verbes selon la position de l'expérienceur, et certaines propriétés que nous avons étudiées ne concernent que l'une ou l'autre de ces deux catégories. Nous avons également pris en compte la sélection et la sémantique des arguments du verbe, et certaines propriétés que nous avons étudiées l'ont été également par Beth Levin dans sa classe Psych-Verbs et celle des verbes de Désir, comme la propriété 1, qui est la construction de base avec indication de la position de l'expérienceur, la propriété 13 qui est la possibilité ou non de formation du passif, la propriété 14 qui indique que le choix des prépositions du passif dépend du verbe, mais nous avons considéré qu'il dépend aussi de la nature du sujet et du complément, et la propriété 20, similaire en français, qui indique une possible dérivation morphologique pour former des adjectifs en *–able*. Nous avons également pris en compte la propriété 19, la possibilité de nominalisation, avec différentes formations comme les suffixations en *–ation* (adorer/adoration) ou *-ement* (étonner/étonnement), et les conversions (passionner/passion).

Nous nous écartons de cette analyse dans la mesure où nous ne nous sommes pas intéressée à la langue en général (le français) mais à un domaine sémantique particulier, et notre lexique n'est pas limité aux verbes mais nous l'avons étendu aux noms et aux adjectifs dont nous avons étudiés des propriétés particulières. De plus, étant donné notre projet de départ (une description sémantique) nous avons construit un ensemble de relations sémantiques entre les classes et entre les mots, comme les relations d'intensité ou d'antonymie. Enfin, nous avons proposé des formalismes pour représenter ces connaissances.

2.3.3. WordNet

Nos descriptions sur les sentiments présentent certaines similitudes avec le réseau WordNet. En effet, notre lexique est également constitué de noms, verbes et adjectifs, mais

[23] *Les Colinettes par Madame d'Auneuil*
[24] TZARA Tristan, *Théâtre (1912-1924)*, 1924, p. 530.

pas d'adverbes. Nous avons construit des classes homogènes sémantiquement, mais, à la différence des synsets, elles regroupent des mots de catégories différentes, avec des propriétés qui dépendent de ces catégories. Les classes et les mots de sentiments sont liés par des relations, certaines similaires à celles de WordNet comme la synonymie et l'antonymie, mais nous n'avons pas défini d'hyperonymie ni de troponymie. Nous avons défini des relations d'antonymie entre classes, alors qu'elles sont entre les mots dans WordNet. Ces relations ne sont pas définies de la même façon. Par exemple, dans notre description *irriter* et *adorer* ont des antonymes, respectivement *calmer* et *haïr*, alors que *Irritate* (irriter) ou *Adore* (adorer) n'en n'ont pas dans Wordnet. De plus, nos relations d'antonymie sont réciproques, *détester* est antonyme d'*aimer* et vice versa, alors que dans Wordnet, *love* a pour antonyme *hate* qui a pour synonyme *detest*, mais si *hate* a bien pour antonyme *love*, on n'a pas la relation directe d'antonymie entre *detest* et *love*. Ainsi, si on interroge WordNet 2.1. (2008), on obtient les résultats indiqués dans la figure 10. La question de l'utilisateur est en en caractères gras italiques, WordNet propose un menu avec le choix ***Antonyms*** que l'utilisateur a choisi. Pour *detest*, le choix ***Antonyms*** n'est pas proposé.

Search Word : ***love***

love -- (have a great affection or liking for; "I love French food"; "She loves her boss and works hard for him")

Antonym of hate

=> hate, detest

Search Word : ***hate***

hate, detest -- (dislike intensely; feel antipathy or aversion towards; […])

Antonym of love

Search Word : ***detest***

hate, detest -- (dislike intensely; feel antipathy or aversion towards; […])

Figure 10. Exemples d'antonymie dans WordNet

Les relations ou propriétés d'intensité n'existent pas en tant que telles dans WordNet, mais sont parfois indiquées comme relation de troponymie, par exemple une des relations troponymiques donnée par WordNet 2.1. pour *terrify* (terrifier), est « *frighten greatly* (effrayer beaucoup », qui correspond à notre propriété intensifieur avec la valeur **Fort** pour *terrify* et **Médian** pour *frighten.*

Nous avons étudié des relations morphosyntaxiques aussi mais seulement si elles sont liées au sens c'est-à-dire si elles décrivent ou caractérisent un sentiment ou un état psychologique ; ainsi nous ne prenons pas en compte la relation entre *amuser* et *amuseur* contrairement à WordNet qui, dans un champ sémantique légèrement différent, relie *entertain* (divertir) à *entertainer* (personne qui procure du divertissement*)*.

Nous avons pris en compte la sémantique des arguments des verbes, mais alors que dans WordNet des exemples de phrases sont construits en respectant cette caractérisation des arguments, dans notre description elle est explicitement indiquée par des propriétés. Ainsi, les verbes de type *irriter* sont exemplifiés par des phrases où ils ont un complément d'objet humain dans WordNet (*The bad news will irritate him, The performance is likely to fascinate Sue*), alors que dans notre description ils acceptent la propriété « objet humain ».

Enfin WordNet décrit la langue dans sa totalité alors que nous ne nous sommes intéressée qu'à un domaine sémantique particulier.

CHAPITRE 3. MODÈLES DE DONNÉES ET REPRÉSENTATIONS DES CONNAISSANCES

Le problème de la modélisation des données linguistiques dans l'optique de leur traitement par une machine a toujours été au centre de nos travaux. D'abord dans *TRANSLEGS*, une plate-forme de comparaison de lexiques grammaires de langues différentes, puis dans *INTERFACILE*, un système de compréhension du langage qui permet de donner des explications sur un sujet donné et qui reformule les questions de l'utilisateur mal comprises, et enfin dans nos travaux en sémantique lexicale, en particulier dans le domaine des sentiments et des opinions avec les systèmes *FEELING* et *FEELING-G*. Cela explique notre souci de construire une description entièrement explicite et non-ambiguë, utilisable par un système informatique.

3.1. AVANT LES SENTIMENTS : *TRANSLEGS* ET *INTERFACILE*

3.1.1. TRANSLEGS

TRANSLEGS est une station de travail linguistique qui permet, d'une part, l'étude du lexique grammaire d'une langue et, d'autre part, la comparaison de lexiques grammaires de langues différentes (Elia et Mathieu 1986, Mathieu 1989). L'application a été faite pour l'italien et le français, dans le sens italien vers français, pour la comparaison des verbes à un complément. La base contient 9 tables italiennes, une trentaine françaises, environ 2300 verbes, et de 20 à 36 propriétés selon les tables. *TRANSLEGS* a été réalisé en collaboration avec le professeur Annibale Elia de l'Université de Salerne (Italie).

Comme l'explique ce dernier (Elia 1984), des correspondances ont été établies entre classes, entre verbes et entre propriétés :

« [...] les différences générales entre les structures du français et de l'italien impliquent une certaine prudence théorique quant à la comparaison entre les deux langues. Cependant, ils n'interdisent pas de mettre en évidence des similarités de comportement des deux langues sur des fragments de lexique qui, à notre avis, sont fort voisins ». Elia page 160.

Les structures définitionnelles des classes du lexique grammaire italien et de celles du lexique grammaire français ont été mises en correspondances, elles sont indiquées dans le tableau 12.

Classes italiennes (préfixées *it*)	Classes françaises
41	4
42	5
43, 43B	6
44, 44B	7
45, 45B, 46	8
47, 47B	9

Tableau 12. Correspondances entre les tables de verbes à complétives italiennes et françaises

À chaque verbe italien a été associé son équivalent français, c'est-à-dire non seulement sa traduction ou son équivalent le plus proche, mais également la table du lexique grammaire français à laquelle celui-ci appartient. Un extrait de la table italienne **it41** est donné dans le tableau 13, la figure 11 montre un exemple de comparaison des Lexiques grammaires italien et français.

N_0=:¨Nhum	N_0=:¨Nnr	N_0=:¨Ch F	N_0=:¨ V^1 Ω	italien	Prep	N_1=:¨ Nhum	Ppv = gli	N_1=:¨-Nhum	N_1=:il fatto Ch F	a N_1 ?	in N_1 ?	(di + da) N_1 ?	Ppv = (ci + vi)	del fatto Ch F	N_0 V	N_1 ¨di V^1-Ω	N_0 V N^0	français	Table française
-	+	-	+	accadere	a													arriver	5
-	+	+	+	addirsi	a													convenir	5
-	+	+	+	affiorare	loc													affleurer	35ST
+	+	+	+	aggradare	a													agréer	5
				agire	su													agir	5
				albergare	loc													demeurer	5
				aleggiare	loc													voleter	35L
				andare	a													aller	5
				apparire	a													apparaître	17
				apparire	loc													apparaître	5

Tableau 13. Extrait de la table italienne it42

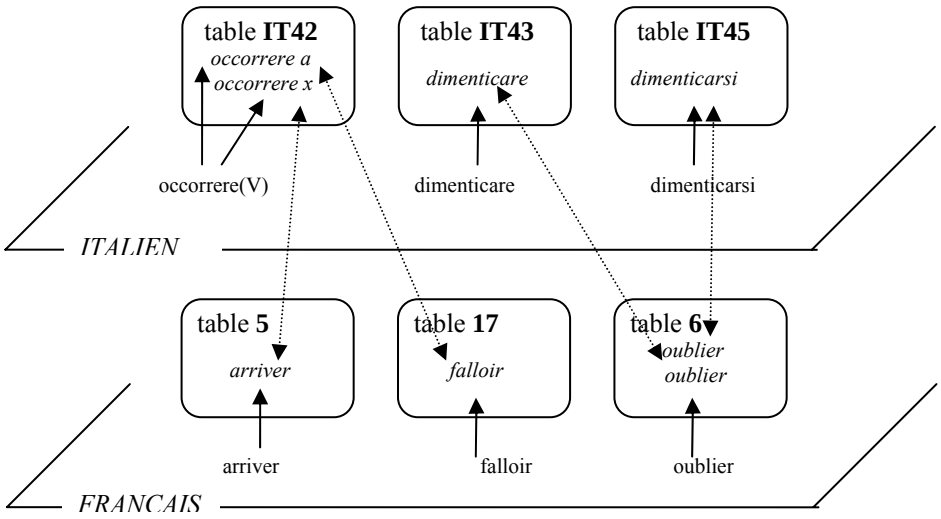


Figure 11. Exemple de comparaison des Lexiques grammaires italien et français

Les correspondances ne sont pas bijectives. En effet, une propriété ou une classe peuvent ne pas avoir de correspondant, ou en avoir plusieurs ; lorsqu'une propriété n'a pas de correspondant le système doit proposer des propriétés de substitution formellement voisines. Ainsi la propriété *V de ce qu P* n'existe pas en italien, le système propose la propriété *V del fatto Ch F* . Un exemple avec le verbe *douter* est donné dans la figure 12. Les critères de choix des propriétés de substitution peuvent à tout moment être modifiés par l'utilisateur.

ITALIEN		FRANÇAIS
dubitare Ch F	<---------->	*douter qu P*
dubitare del fatto Ch F	<---------->	*douter du fait qu P*
	<---------->	
0	<---------->	*douter de ce qu P*

La ligne « <----------> » indique le choix d'une équivalence par proximité formelle

Figure 12. Propriétés de substitution du verbe *douter*

Le système *TRANSLEGS* est constitué de deux modules : un module monolingue qui permet l'étude des propriétés syntaxiques d'une langue (français ou italien ou...), et un module bilingue qui est un outil d'aide à la traduction, actuellement italien-français.

L'interface de *TRANSLEGS* avec l'utilisateur est constituée d'un ensemble d'écrans affichant les choix d'interrogation possibles de la base de connaissances. L'utilisateur peut soit consulter le lexique grammaire d'une langue, soit faire une comparaison entre les lexiques grammaires de deux langues, dans les deux cas à partir d'un verbe, d'une propriété ou d'une table choisis dans une liste proposée par le système. Il n'y a donc pas de véritable « dialogue » entre *TRANSLEGS* et l'utilisateur, les choix de ce dernier sont entièrement dirigés par le système.

Quand l'utilisateur interroge *TRANSLEGS*, il ne travaille pas sur une représentation fixe des lexiques grammaires : il étudie une vue temporaire qui n'existe pas en tant que telle dans les données et que *TRANSLEGS* construit dynamiquement au moment de la consultation. L'exemple de la comparaison d'un verbe italien et de son correspondant français est donné dans la figure 13.

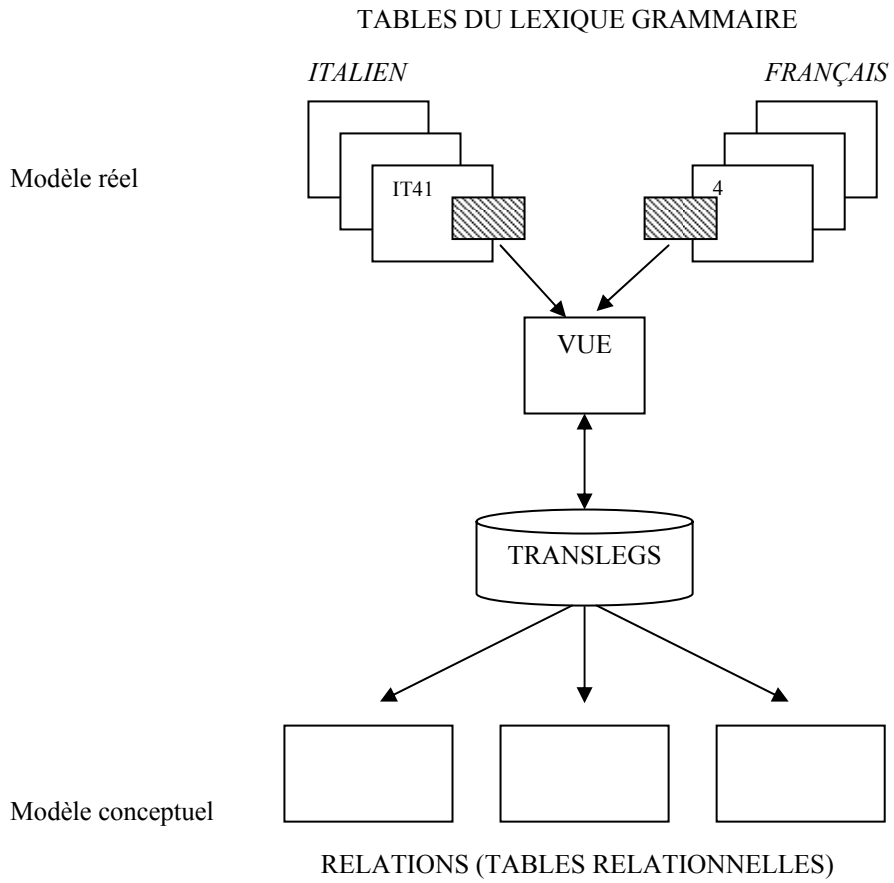


Figure 13. Construction d'une vue dynamique dans le système *TRANSLEGS*

L'utilisateur s'intéresse à un sous-ensemble du lexique grammaire italien et à un sous ensemble du lexique grammaire français. La juxtaposition de ces 2 sous-ensembles

constitue une vue partielle et temporaire. *TRANSLEGS* va reconstituer cette vue à partir de ses relations et des liens qui existent entre elles. Cette reconstitution est dynamique et n'existe que le temps de la consultation. Nous reprendrons cette notion de vue dynamique dans le système *FEELING*.

La base est construite autour d'un Système de Gestion de Bases de Données (SGBD) relationnel (*INFORMIX*), avec des modules d'interface écrites en langage procédural (langage C). Ce sont ces interfaces qui permettent de faire les déductions quand les constructions syntaxiques ne sont pas similaires dans les deux langues, et de communiquer avec l'utilisateur. Les tables du lexique-grammaire sont représentées sous forme de tables relationnelles, c'est-à-dire des données structurées avec un format fixe, dont on peut extraire l'information grâce à l'algèbre relationnel et à un langage de requêtes (SQL).

3.1.2. INTERFACILE

INTERFACILE est une interface en langage naturel entre un utilisateur et un ordinateur, qui permet la compréhension et la reformulation de questions mal comprise par la machine. *INTERFACILE* a été réalisé en 1986 en collaboration avec Paul Sabatier (LIF, CNRS).

Le système est constitué d'une petite grammaire du français utilisée pour comprendre les questions, résoudre certaines anaphores et générer des réponses correctes syntaxiquement ; il contient également des connaissances sémantiques du domaine concerné. L'application que nous avons développée est l'aide à l'utilisation et à la compréhension d'un système d'exploitation comme UNIX, pour des utilisateurs non informaticiens, à l'époque où il fallait taper directement des lignes de commande (car les icones et les souris n'existaient pas). Voici des exemples de dialogues (extraits de Mathieu et Sabatier 1986) ; le prompt > précède la question de l'utilisateur, le prompt – précède la réponse d'*INTERFACILE* en italiques. Les questions doivent commencer par « comment ».

Dialogue 1

> comment créer un fichier ?
>
> - *utilisez la commande : CREATE nomfichier*

> comment l'imprimer ?
>
> - *utilisez la commande : PRINT nomfichier*

> comment modifier son contenu ?
>
> - *utilisez l'éditeur avec la commande : EDIT nomfichier*

> comment copier un fichier dans un répertoire ?
>
> - *utilisez la commande :*

COPY nomfichier [nomrepertoire]

> comment renommer ce dernier ?
>
> - *utilisez la commande :*

RENAME/DIRECTORY [ancienrepertoire] [nouveaurepertoire]

Dans un but « pédagogique » pour les non spécialistes en informatique, *INTERFACILE* reformule les questions mal comprises, comme le montre le dialogue 2.

Dialogue 2

> comment envoyer un message à un utilisateur ?
- utilisez la commande : MAIL
> comment le supprimer ?
- le renvoie à : message ? (1)
utilisateur ? (2)
tapez le numéro de votre choix.
> 2
- INTERFACILE n'en sait rien, mais sait comment :
envoyer un message
écrire un message
supprimer un fichier
supprimer un répertoire
supprimer un message

Quand une question est déviante par rapport à ce qu'attend le système, soit parce que, telle qu'elle est formulée, la question elle même n'a pas de sens (dialogue 3), soit parce que le système n'a pas les connaissances pour y répondre (dialogue 4), *INTERFACILE* indique à l'utilisateur ce qu'il sait et le guide pour reposer sa question.

Dialogue 3

> comment ajouter à un fichier le contenu d'un fichier ?
- pour « ajouter », utilisez la construction :
ajouter [quelque chose]
à [quelque chose]
complétez ou tapez / pour annuler)
> comment ajouter ...

Dialogue 4

> comment envoyer un message dans un terminal ?
- pour « envoyer », utiliser les constructions :
envoyer [quelque chose]
à [quelqu'un]
envoyer [quelque chose]
sur [quelque chose]

complétez ou tapez / pour annuler)
> comment envoyer ...

L'aide n'est pas une reformulation linguistique : les informations sont déduites à partir des connaissances sémantiques du domaine de l'application et reformulées grâce à la grammaire.

Les connaissances dans *INTERFACILE* sont représentées dans un formalisme de grammaire « logique », sous forme de prédicats (clauses de Horn) avec un ensemble de règles de déductions qui permet de les exploiter. Le système a été implémenté dans un langage déclaratif avec un mécanisme d'unification (Prolog).

3.1.3. Comparaison de TRANSLEGS et INTERFACILE

Contrairement à *TRANSLEGS*, dont l'interface ne proposait que des écrans figés prédéfinis, *INTERFACILE* dialogue avec l'utilisateur, et ses réponses sont construites de façon dynamique à partir d'une grammaire et d'une base de connaissances sémantiques.

Le mode de représentation des données et la façon de les exploiter sont également différents dans les deux systèmes. La base de connaissances de *TRANSLEGS* est sous forme de tables relationnelles avec un format fixe, on y a accès grâce à un langage de requêtes, et des déductions sont faites avec des règles écrites dans un langage procédural, c'est-à-dire une succession d'instructions avec un ordre impératif. Les connaissances dans *INTERFACILE* sont décrites de façon déclarative sous forme de faits élémentaires et de règles associées. Leur exploitation est faite grâce à un mécanisme d'unification.

Cette évolution d'un système (*TRANSLEGS*) à l'autre (*INTERFACILE*) se traduit donc par un niveau conceptuel supplémentaire dans la représentation, le traitement et l'exploitation des données.

3.2. LES SENTIMENTS : *FEELING, FEELING-G* ET L'ONTOLOGIE DES SENTIMENTS

3.2.1. FEELING

En 1993, nous avons conçu et développé le système *FEELING* pour vérifier le bien-fondé de la description que nous avons proposée pour les verbes de sentiment. Il fournit une interprétation sémantique des verbes, génère des paraphrases syntaxiquement correctes, et (à un niveau simple) vérifie la bonne formation d'un énoncé. Nous avions également pour objectif la conception d'un système qui explique la raison des déductions qu'il fait et qui offre la possibilité d'enrichir à tout moment non seulement la connaissance du système, mais également ses moyens de traitements.

Nous donnons ci-dessous un exemple d'analyse de phrase réalisée par *FEELING*. Le système ne fait pas d'analyse syntaxique. Il faut donc lui fournir explicitement le sujet, le verbe et le complément. On doit lui indiquer également la forme du sujet, c'est-à-dire s'il est phrastique, un substantif humain (*Nhumain*), un infinitif, ou autre, mais pour une meilleure lisibilité nous avons écrit une phrase. Les réponses de *FEELING* sont divisées en trois parties :

(1). l'interprétation de la phrase,

(2). les phrases associées et, à la demande, les propriétés de l'item lexical (option non choisie dans cet exemple).

(3). les modifications d'intensité du sentiment.

Phrase à interpréter *Luc exaspère Marie*

(1). INTERPRÉTATION

Le sentiment est éprouvé par *Marie,* **La cause est** *Luc*
Le sentiment éprouvé est *plutôt désagréable*
Le sentiment éprouvé est *l'exaspération*

Phrases synonymes :
Luc énerve beaucoup Marie
Luc (courrouce / enrage / excède / hérisse / horripile / impatiente / ulcère) Marie

Phrases Antonymes
Luc apaise Marie

Paraphrases
Marie éprouve de l'exaspération (envers / contre) Luc
Marie est exaspérée par Luc

Nature du sujet
Luc peut être un sujet agentif : *Luc exaspère Marie volontairement*

(2). PHRASES ASSOCIÉES

Marie est (exaspérée / très énervée)
Marie est (exaspérée / très énervée) par Luc
Marie éprouve (de l'exaspération / beaucoup d'énervement)
Marie éprouve (de l'exaspération / beaucoup d'énervement) (envers / contre) Luc

(3). MODIFICATION DE L'INTENSITÉ DU SENTIMENT

(3.1) La **diminution** continue de l'intensité du verbe *exaspérer* donne les sens suivants :
Luc (agace / crispe /énerve / fâche / irrite) Marie
Luc (dérange / désoblige / emmerde / empoisonne) Marie
Luc (ennuie / enquiquine /gêne / importune / incommode / indispose) Marie

(3.2) L'**augmentation** continue de l'intensité du verbe *exaspérer* donne les sens suivants :
Luc (braque / choque / écœure / scandalise) Marie
Luc (/ indigne / rebelle / révolte) Marie

3.2.1.1. Représentation des connaissances dans *FEELING*

Le système *FEELING* exploite les classes sémantiques de verbes psychologiques, (du type *irriter*) que nous avons décrites, et les propriétés associées, classification élaborée à la fin des années 1990. Comme nous l'avons déjà indiqué, cette classification a été modifiée depuis pour prendre en compte les noms et les adjectifs. Certains passages sont extraits de Mathieu (2000) noté YM dans la suite.

Les classes

Nous avons adopté un formalisme qui utilise la notion de « prototype » pour représenter les classes sémantiques. Nous utilisons un prototype au sens de Eleanor Rosch (1975, 1978) qui considère que pour un être humain, les objets du monde réel sont structurés en catégories. Cependant, comme tous les exemplaires d'une même catégorie n'en sont pas également représentatifs, on choisit pour prototype l'« élément moyen », considéré comme le meilleur représentant de la catégorie. Les classes sémantiques de verbes ont été nommées par le verbe le plus typique de la classe, il y a donc analogie entre les noms des classes et leur prototype. Ainsi, le verbe *énerver* qui est le prototype des verbes de la classe sémantique regroupant les verbes exprimant l'énervement est également le nom de cette classe. Ce prototype possède l'ensemble des caractéristiques de la plupart des éléments de la catégorie et contient une connaissance par défaut. Le verbe *énerver* a pour propriétés : (*a un emploi non psychologique, a un sujet agentif, a un intensifieur Médian*). Chaque verbe de la même classe est une « spécialisation » du prototype *énerver,* et hérite donc de ses trois propriétés. Une spécialisation peut également avoir des propriétés qui lui sont propres. C'est le cas, par exemple, du verbe *exaspérer* qui, contrairement au prototype, n'a pas d'emploi non psychologique. Les valeurs des propriétés héritées sont alors masquées par

des valeurs propres à la spécialisation. Notre classification est ainsi représentée par une arborescence :

> « Nous avons défini une structure, appelée *RACINE*, qui contient les connaissances communes à tous les verbes psychologiques de notre étude. Ce « père » est la racine unique de la hiérarchie des connaissances, dont les prototypes définis à partir des classes sémantiques sont des spécialisations. [...] Chaque verbe est alors une spécialisation d'un prototype. » YM page 114.

Les propriétés

Parmi les propriétés que nous avons étudiées, certaines peuvent être représentées, pour chaque item lexical, par une paire (attribut-valeur), l'attribut est la propriété, la valeur est la valeur que prend cette propriété. Ainsi, pour le verbe *exaspérer*, la propriété [Expérienceur] a pour valeur « complément ».

Un exemple de ces propriétés « simples » pour la classe **Énerver,** le sous-ensemble des verbes psychologiques de la classe **Colère,** est donné dans le tableau 14. Chaque ligne correspond à un verbe (*agacer, énerver,* etc.), et chaque colonne à une propriété.

Verbe	**Sujet agentif**	**Intensifieur**	***N_1 est Vpp de N_0***	**Passif sans agent**	**Emploi non psychologique**
agacer	oui	Médian	oui	oui	non
énerver	oui	Médian	oui	oui	oui
exaspérer	oui	Fort	oui	oui	non
horripiler	non	Fort	oui	oui	non
irriter	oui	Médian	oui	oui	oui

Tableau 14. Exemple de propriétés simples des verbes de sentiment

Plusieurs représentations sont possibles pour décrire les propriétés plus complexes, comme des tableaux, des listes ou des règles.

Les bases de connaissances dans FEELING

Nous appelons **base de connaissances** l'ensemble des traitements et inférences qu'il est possible d'effectuer sur les propriétés et l'interprétation sémantique qui peut être déduite en fonction du triplet (verbe, sujet, complément). Nous avons représenté ces connaissances par des règles de production déclaratives de la forme « **si** condition(s) **alors** conclusion(s)*i* **sinon** conclusion(s)*j* ».

Une base de connaissances peut concerner et donc s'appliquer à tous les verbes, dans ce cas elle est contenue dans la **Racine**. Elle peut également ne s'appliquer uniquement qu'à

des prototypes particuliers ou à leurs spécialisations; elle est alors contenue dans ces prototypes ou ces spécialisations. La base de connaissances d'un prototype contient ainsi des informations de plusieurs types :

1-) les inférences qui déterminent la valeur de la préposition dans certaines constructions ;

2) les inférences qui concernent les verbes qui acceptent un complément de la forme *Dét N de Nhum*, et qui permettent :

- de déduire les possibilités de pronominalisation et de restructuration,

- de déduire le domaine sémantique du sujet quand le substantif approprié *N* du complément appartient à certains domaines de valeurs,

- de vérifier si une construction obtenue par restructuration ou pronominalisation d'une phrase de base, est acceptable ;

3-) les inférences qui permettent de générer des constructions passives et des nominalisations.

Pour représenter ces inférences nous avons utilisé des règles de déductions de la forme **si** condition(s) **alors** conclusion(s)*i* **sinon** conclusion(s)*j*. Par exemple, nous donnons ci-dessous un exemple simplifié d'une des règles qui expriment les contraintes liées à la nature du sujet dans la formation des constructions passives :

1. Si la phrase à analyser est de la forme *Sujet Verbe Complément (S V C)*

2. et si le sujet est une phrase à l'infinitif

3. et si il existe pour le verbe au moins une construction passive avec la préposition *de*

4. et si le complément est un nom humain

5. alors la phrase *Complément est (participe passé de V) de Sujet* est acceptable

Si on applique cette règle à la phrase *Voir Luc partir déçoit Marie*, la condition **1** est vérifiée avec *Voir Luc partir* comme sujet, le verbe est *déçoit* (*décevoir*) et le complément *C* est *Marie*. La condition **2** est vérifiée car le verbe *décevoir* accepte la formation du passif introduit par la préposition *de.* La condition **3** est vérifiée car le sujet *Voir Luc partir* est une phrase à l'infinitif et enfin la condition **4** est vérifiée car le complément *Marie* est un humain. La conclusion **5** permet donc de dire que *Marie est déçue de voir Luc partir* est une phrase acceptable.

Chacune des parties de *FEELING* (propriétés et base de connaissances) est optionnelle. Les propriétés sont les caractéristiques du prototype, ce sont des connaissances typiques, les plus fréquemment rencontrées chez les membres du prototype. La base de connaissances contient des règles qui décrivent le comportement du prototype.

3.2.1.2. Le mécanisme d'héritage

Un mécanisme d'héritage permet un partage des propriétés et des bases de connaissances.

> « Chaque spécialisation hérite de son prototype. Elle hérite à la fois des propriétés et du contenu de la base de connaissances. Mais une spécialisation peut être un élément exceptionnel, pour lequel il est nécessaire soit d'ajouter des propriétés et/ou des éléments dans la base de connaissances, soit de masquer les valeurs par défaut des propriétés et des connaissances héritées. Il y a donc un mécanisme de préemption de la spécialisation sur son prototype (Walter Daelemans, Koenraad De Smedt et Gerald Gadzar (1992)). » YM page118.

Un exemple de ce mécanisme d'héritage entre la **Racine**, la classe **Énerver** et le verbe *horripiler* est décrit dans la figure 14. Tous les verbes psychologiques acceptent pour complément d'objet direct un substantif strictement humain (propriété [objet humain]), cette propriété est donc indiquée au niveau de la Racine, et sera héritée par toutes les classes puis par tous les verbes. Le verbe *énerver*, comme la majorité des verbes de la classe **Énerver**, accepte un sujet agentif et un passif sans agent, ces propriétés seront donc indiquées au niveau de la classe. Tous les verbes de la classe **Énerver** héritent de ces trois propriétés, dont le verbe *horripiler,* mais ce dernier n'accepte pas de sujet agentif, ce qui « masque » l'héritage de cette propriété.

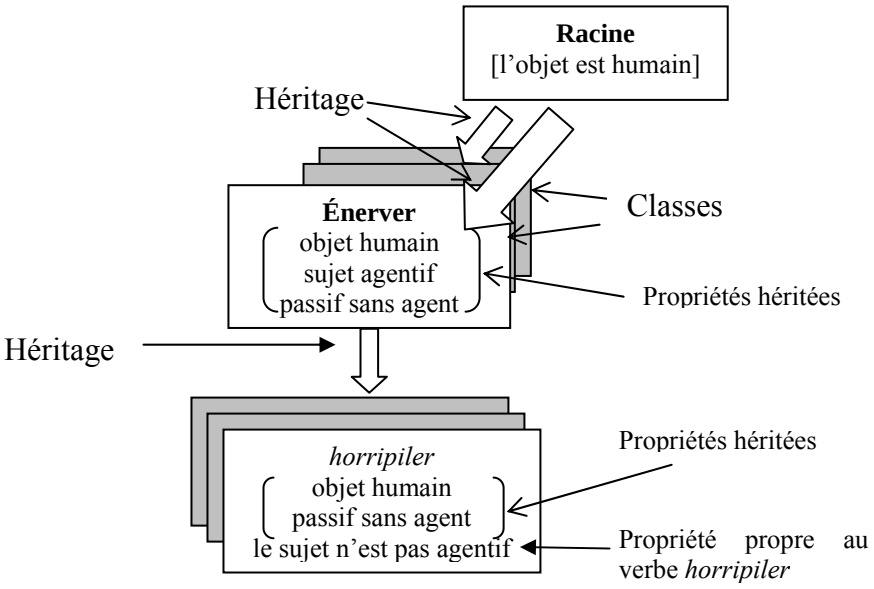


Figure 14. Mécanisme d'héritage dans le système *FEELING*

Nous avons utilisé un langage déclaratif (Prolog) pour implémenter le système *FEELING*. Pour permettre une indépendance des traitements vis à vis de la base sémantique, nous avons adopté une structuration du traitement des connaissances sous

forme d'opérateurs. Ils sont autonomes, chacun est spécialisé pour un type précis de recherche d'information ou de traitement des connaissances.

Cette modélisation sous forme d'opérateurs permet un enrichissement permanent de la base sémantique sans avoir à modifier le traitement de ses connaissances. On peut ajouter, modifier ou enlever indifféremment des verbes, des classes sémantiques, des propriétés, des liens d'intensité entre classes ou des règles déductives, sans avoir à modifier les opérateurs qui portent sur ces différentes informations.

La méthode de traitement est la même que dans *TRANSLEGS*, pour chaque phrase, le système crée un espace de travail temporaire dans lequel vont être mises toutes les informations (descriptions et déductions) générées par la (ou les) analyse(s) de la phrase. L'interface de communication pourra alors exploiter ces résultats. Cette interface comprend également un module destiné à l'enrichissement de la base sémantique et des opérateurs.

3.2.2. *FEELING-G*

Ce système reprend les caractéristiques du système *FEELING*, auxquelles nous avons ajouté une représentation sémantique des sentiments donnée sous forme de structure de traits. (Mathieu 2004). Rappelons qu'une structure de traits décrit des connaissances sous forme de liste de couples « attribut (nom du trait) : valeur ». Les attributs ont des valeurs atomiques, mais également des valeurs complexes représentées par d'autres structures de traits.

Nous avons défini trois principales catégories de traits : s***entim*** qui décrit le sens, l'intensité et les propriétés linguistiques du sentiment considéré, ***exper*** qui identife la personne qui éprouve le sentiment, et ***causob*** qui décrit la cause ou l'objet du sentiment.

Le trait s***entim*** contient 5 attributs : **Prédicat** dont la valeur est la lexicalisation du sentiment, **Polarité** et **Niveau d'intensité** du sentiment, **Classe** qui caractérise la classe sémantique à laquelle appartient le sentiment et **Synonymes** de ce dernier. Le nom de ce trait est suivi de « #i », ce qui permettra de l'unifier avec l'attribut Sentiment que nous allons décrire plus loin.

L'attribut Classe est lui-même une structure de 5 traits : **Sémantique** dont la valeur est le nom de la classe, **Héritage** qui est la classe dont cette classe hérite, la classe **Antonyme**, la classe où l'intensité du sentiment est plus faible, **W-classe,** et enfin celle où l'intensité du sentiment est plus élevée, **S-classe**, si elles existent.

Les traits ***exper*** et ***causob*** contiennent les caractéristiques de l'expérienceur et de la cause ou de l'objet du sentiment, dont l'attribut **Ident** qui les identifie dans la phase considérée. Le sentiment éprouvé par l'expérienceur est celui décrit dans le trait ***sentim***, il n'est pas dupliqué : on fait référence à lui dans ***expr*** par la même notation « #i », qui, grâce à un mécanisme d'unification, permettra de remplacer « #i » dans ***expr*** par la structure de traits ***sentim*** lors des traitements.

La représentation de l'exemple précédent *Luc exaspère Marie* est donné dans la figure 15.

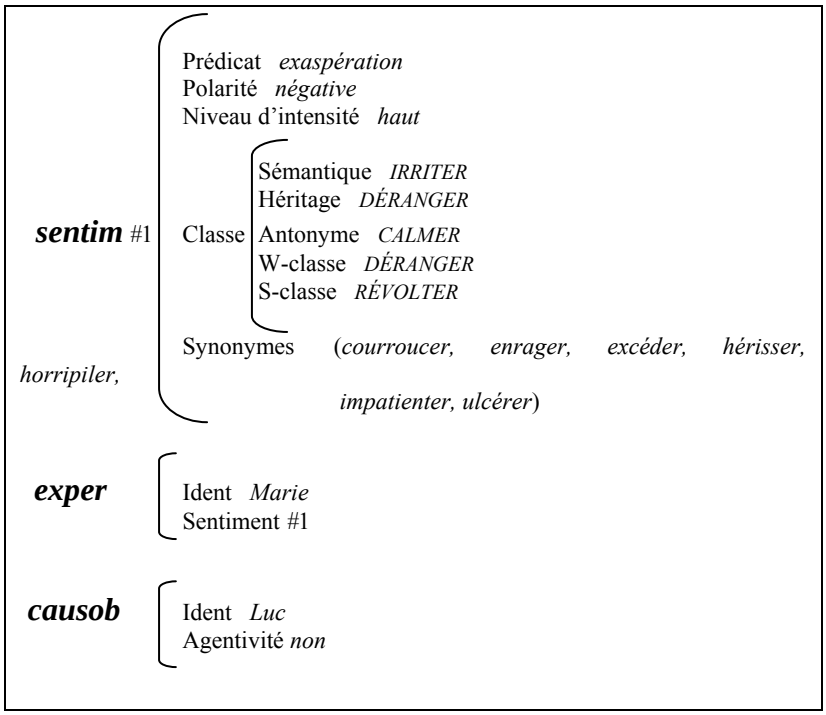


Figure 15. Représentation sémantique de *Luc exaspère Marie*

Les interfaces de *FEELING* et *FEELING-G* avec l'utilisateur sont interactives. On fournit une phase à analyser au système, et à tout moment on peut choisir d'avoir (option « verbeuse ») ou non les informations qu'il déduit de ses connaissances.

3.2.3. Du lexique sémantique à l'ontologie

Nous venons de le voir, la description du lexique sémantique et des propriétés a été faite avec un « format propriétaire », peu accessible. En effet, nous avons utilisé différentes représentations : des paires attribut-valeur, des listes, des tableaux, des graphes, des règles. Pour faciliter l'accès à ces connaissances et les rendre utilisables à la fois par des utilisateurs humains et des logiciels, nous les avons représentées sous forme d'ontologie, formalisme qui permet de représenter les connaissances avec un format standard et de disposer d'outils associés. Nous considérons une ontologie au sens où ce terme est utilisé dans le Web sémantique, c'est-à-dire un ensemble de concepts organisés hiérarchiquement auxquels sont associés des attributs qui caractérisent le concept (Uschold et Gruninger 1996), définition assez éloignée du sens philosophique d'Ontologie, qui est « Partie de la métaphysique qui s'applique à l'être en tant qu'être, indépendamment de ses déterminations particulières. » (Le Petit Robert 2001).

Les valeurs des attributs sont de type atomique, comme une chaine de caractère ou un nombre entier. Les concepts (aussi appelées « classes ») sont liés entre eux par des relations sémantiques.

La représentation sous forme d'ontologie permet d'avoir accès automatiquement aux connaissances sémantiques pour des traitements entièrement informatisés. Un accès *aux* connaissances, mais pas *à toutes* les connaissances, car il n'existe pas à l'heure actuelle d'outils pour les décrire dans toute leur complexité, ni de consensus dans la communauté pour savoir quels seraient ces outils.

Nous n'avons donc pu représenter qu'un sous-ensemble des connaissances que nous avons étudiées et formalisées, ce qui a entrainé une perte d'information et une dégradation de la description sémantique très fine que nous avons faite. Nous n'avons pas pu représenter, par exemple, les règles qui permettent de générer des constructions passives ou déterminent la valeur de la préposition dans des nominalisations, etc., ou les règles qui permettent de distinguer le sens « concret » de celui métaphorique d'un verbe dans une phase. Une solution partielle pourrait être apportée par le langage OWL[25] dont les spécificités ont été définies par le consortium W3C[26] (World Wide Web Consortium) pour des applications dans le domaine du Web sémantique. Ce langage permet de définir les éléments d'une ontologie et de les combiner avec un langage de règles pour produire des raisonnements, mais son emploi est très lourd et loin de tout résoudre.

Notre ontologie des sentiments contient 118 concepts (Mathieu 2005a]. Nous l'avons implémentée avec l'outil de gestion des connaissances *Protégé-2000*[27].

3.2.4. Annotation et navigation à travers les sentiments

3.2.4.1. Annotation

Comme nous l'avons indiqué au §2.1.4., nous avons exploité l'ontologie des sentiments comme ressources pour l'annotation et le filtrage des sentiments dans un texte, et la navigation à travers ce texte annoté. Nous avons adopté le cadre de travail de Javier Couto et Jean-Luc Minel que nous allons décrire brièvement. Les citations sont extraites de (Couto à paraître).

Le traitement informatique d'un texte par un système nécessite qu'il ait une représentation utilisable :

> « C'est le type de traitement à effectuer qui viendra déterminer habituellement la représentation la plus convenable selon des critères à choisir : rapport coût-performance, souplesse, exhaustivité, simplicité, entre autres. En conséquence, il existe des représentations d'un texte qui le conceptualisent comme une séquence de caractères, jusqu'à des représentations fortement imprégnées par des

[25] http://www.w3.org/TR/owl-features/

26 http://www.w3.org/

27 http://protege.stanford.edu/

approches linguistiques, où la représentation de texte permet de modéliser certains phénomènes discursifs. »

Le texte doit avoir les caractéristiques suivantes :

« ne pas restreindre à un ensemble prédéterminé le type d'unités textuelles qui composent un texte ; offrir à la fois une organisation hiérarchique des unités textuelles et une autre permettant d'exprimer des relations non hiérarchiques ; considérer les titres comme des unités textuelles ; admettre que toute unité textuelle, y compris les titres et les relations non hiérarchiques, est susceptible d'avoir un nombre non limité d'annotations de nature quelconque. »

Un texte est composé des trois éléments : son *titre*, la *tête*, et le *corps*, dans cet ordre. Le titre est composé d'unités textuelles dont l'organisation (une seule unité textuelle, une séquence d'unités, une hiérarchie d'unités) est définie par l'utilisateur. Cette représentation est indiquée dans la figure 16. Le corps est constitué d'une hiérarchie d'unités textuelles, la tête regroupe des relations non hiérarchiques construites à partir des d'unités textuelles existantes dans le *corps*. Le titre et la tête sont des éléments optionnels, le corps est un élément obligatoire.

Un texte est ainsi conçu comme une hiérarchie d'unités textuelles de base, permettant la définition d'unités plus complexes susceptibles de ne pas suivre la hiérarchie établie. »

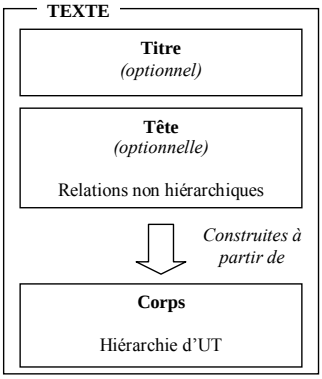


Figure 16.. Composition d'un texte

Une unité textuelle (UT) est typée, c'est-à-dire qu'au lieu d'avoir une unité textuelle *section* ou une unité textuelle *paragraphe* par exemple, les caractéristiques d'une unité textuelle sont données par la valeur de son type. Loin d'être une convention de représentation, cette décision conceptuelle entraine une grande généralité de la notion d'unité textuelle qui donne à l'utilisateur la possibilité de définir lui-même ses propres types. Ainsi, pour l'annotation des sentiments, nous avons défini un type « Marque_Sentiment » (voir exemple figure 17 ci-après).

L'unité textuelle de base contient des attributs fixes (*type*, *numéro*, *rang* et *chaîne*) et des attributs dynamiques (des *annotations* de nature quelconque). Ces attributs diffèrent selon la position, *feuille* ou *nœud,* de l'UT dans l'arborescence du texte. Une unité textuelle *feuille* porte l'information lexicale, indiquée par un attribut fixe nommé *chaîne,* alors qu'une unité textuelle *nœud* sert à représenter la hiérarchie textuelle qui permet l'emboîtement des différentes unités. Cette dernière est composée d'UT filles et ne contient pas d'information lexicale, mais peut s'inférer à partir des UT filles, en faisant un parcours dans l'arborescence.

Les unités textuelles peuvent être annotées. Ces annotations peuvent être héritées soit par les UT descendantes (*héritage descendant*), soit par les UT ascendantes (*héritage ascendant*), soit par les deux.

La tête du texte contient des unités textuelles construites par l'utilisateur pour représenter des constituants textuels qui ne respectent pas la hiérarchie établie dans le *corps* du texte, par exemple des phénomènes linguistiques comme la discontinuité ou le recouvrement. Ces unités sont construites à partir d'UT existantes dans le *corps* du texte, elles peuvent avoir des annotations (qui leur sont propres c'est-à-dire qui ne pourront pas être héritées par les UT de base du *corps*) et enfin elles peuvent se combiner afin de construire des éléments plus complexes. On peut ainsi construire quatre ensembles d'unités textuelles un *Ensemble*, une *Séquence*, une *Référence* ou un *Graphe* :

> « Un *Ensemble* déclare un ensemble non ordonné d'UT pour lesquelles existe, du point de vue de l'annotateur, une relation d'équivalence. [...]
>
> Une *Référence* décrit une relation orientée entre deux UT. Typiquement, ce constructeur permet de représenter le lien entre une anaphore et son référent discursif.
>
> Une *Séquence* est une suite ordonnée d'UT à laquelle l'annotateur attribue une cohésion.[...]
>
> Le dernier type d'objet, *Graphe,* est utilisé pour construire des relations multiples entre des UT. Cette structure permet de représenter des phénomènes linguistiques complexes comme par exemple l'imbrication de différents cadres de discours [...], les différentes pistes de cohérence qui structurent un texte [...], ou encore les expressions des sentiments d'un personnage dans un roman [...].»

Pour la description des expressions de sentiment dans un texte, les unités que nous avons annotées sont les verbes, les noms et les adjectifs exprimant un sentiment. Nous montrons un exemple d'annotation dans la figure 17.

Dans cet exemple, nous avons défini cinq attributs pour une UT :

- Catégorie : la catégorie grammaticale de l'expression du sentiment : verbe, nom, adjectif ;

- Classe sémantique (du sentiment) ;

- Intensité. : le niveau d'intensité (la force) du sentiment exprimé (Faible, Médian ou Fort) ;

- Expérienceur : la personne qui éprouve le sentiment ;

- Objet : l'objet ou la cause du sentiment ;

De façon à faciliter l'expression des opérations de navigation pour être utilisables par le langage Sextant, nous avons mis tous les attributs sur l'UT qui marque le sentiment. Par exemple, dans la phase « Elle était **amoureuse** de Léon, et elle recherchait la solitude, afin de pouvoir plus à l'aise se délecter en son image », nous avons agrégé l'ensemble des annotations sur l'adjectif **amoureuse** alors qu'un analyseur syntaxique, comme par exemple celui que nous avons utilisé dans l'exemple d'annotation du §2.1.4.1., aurait mis les annotations sur l'expérienceur **elle** (cf. figure 8). Nous avons également fait manuellement la résolution d'anaphore et remplacé **Elle** par **Emma**.

```
<UT Type="Paragraphe" Nro="85">
  <UT Type="segment" Nro="1">
      <Chaine> Elle était </Chaine>
  </UT>
  <UT Type="Marque_Sentiment" Nro="1">
      <Attribut Nom="Catégorie">Adjectif</Attribut>
      <Attribut Nom="Classe_Sémantique">Amour</Attribut>
      <Attribut Nom="Intensité">Médian</Attribut>
      <Attribut Nom="Experienceur">Emma</Attribut>
      <Attribut Nom="Objet">Léon</Attribut>
      <Chaine> amoureuse </Chaine>
  </UT>
  <UT Type="segment" Nro="2">
      <Chaine> de Léon, et elle recherchait la solitude, afin de
pouvoir plus à l'aise se délecter en son image.    </Chaine>
  </UT>
  ……
</UT>
```

Figure 17. Exemple d'annotation d'un texte

3.2.4.2. Navigation à travers les sentiments

Nous avons utilisé la plateforme de navigation textuelle NaviTexte pour « lire » le roman Madame Bovary en suivant différents parcours de lecture. Javier Couto et Jean-Luc Minel ont conçu un langage déclaratif, le langage Sextant, qui propose, dans NaviTexte, des opérations prédéfinies de visualisation et de navigation. Les opérations sont définies par une syntaxe formelle et leur sens est déterminé par une sémantique opérationnelle. Ainsi, les vues d'un texte peuvent être de type linéaire, arborescent ou graphe, et peuvent

contenir des chaînes lexicales ou des annotations. Un langage de conditions a été défini, composé de conditions simples, de conditions d'existence sur les éléments des UT et de conditions sur leur hiérarchie.

Les conditions simples portent sur les attributs et sur les annotations des UT. Il existe un opérateur *UT* qui comporte cinq opérandes correspondant aux propriétés suivantes d'une UT : le *type*, le *numéro*, le *rang*, les *annotations* et la *chaîne lexicale*. Le quatrième opérande *annotations* est utilisé pour indiquer l'existence ou non d'annotations, marquée par un nom d'annotation, une valeur ou un couple (nom d'annotation – valeur). Les autres opérandes sont des contraintes d'égalité, d'inégalité, d'ordre (inférieur et supérieur), de préfixe, de suffixe et de recherche de sous-chaîne.

Plusieurs contraintes peuvent s'exprimer pour les annotations, comme l'exemple de la figure 18 qui exprime la condition : *des UT de type « Marque_Sentiment » comportant une annotation de nom «Classe_Sémantique » dont la valeur est « Amour».*

```
UT(type = Marque_Sentiment, *, *, {(Classe_Sémantique, Amour)}, *)
```

Figure 18. Exemple de l'opérateur UT

La condition ci-dessus est donc vérifiée par les expressions de sentiment qui ont été annotées comme Amour (comme valeur de classe sémantique).

Pour vérifier les conditions d'existence des UT, un opérateur sans argument est défini pour chaque élément (cf. figure 19).

Des opérateurs unaires permettent de définir des conditions portant sur le rapport des UT entre elles dans leur hiérarchie. Pour une description détaillée, on peut se reporter à (Couto à paraître).

Enfin, les conditions peuvent se combiner en utilisant les opérateurs logiques classiques *OU*, *ET* et *NON*.

existeAnnotations : teste si l'ensemble d'annotations d'une UT n'est pas vide ;
existeChaîneLexicale : teste si la chaîne lexicale d'une UT est définie ;
existeTitre : teste si le titre d'une UT n'est pas vide ;
existeParent : teste si une UT a une UT parent ;
existeFils : teste si la suite d'UT filles d'une UT n'est pas vide.

Figure 19. Opérateurs d'existence sur les éléments des UT

Le langage Sextant propose des opérations de visualisation et de navigation sur les unités textuelles. Les opérations de visualisation sont des opérations de mise en relief et des opérations d'aide contextuelle. Plusieurs opérations de navigation peuvent être définies sur une UT, ce qui permet de proposer au lecteur plusieurs parcours de navigation à partir de n'importe quel endroit du texte dès lors qu'il contient une opération de navigation dans son annotation.

Une opération de navigation est définie comme une opération qui cherche l'UT cible à partir d'une UT source, en vérifiant les différentes conditions, exprimées dans le langage de conditions que nous venons de présenter, et en suivant l'orientation relative au type d'opération.

La source est définie en utilisant une condition sur les UT. Une opération de navigation est *disponible* pour une UT déterminée si celle-ci vérifie la condition exprimée par la source. La cible est déterminée à partir de deux paramètres : une condition à vérifier pour l'UT cible et le type d'opération de navigation. Une fois déterminée la source, plusieurs UT peuvent vérifier la condition de la cible, et c'est le type d'opération qui indique laquelle choisir.

Chaque opération est typée avec une valeur qui spécifie, d'une part l'orientation c'est-à-dire dans quel sens (avant ou après l'UT source) doit être effectué la recherche de l'UT cible, d'autre part le référentiel par rapport à la source. La valeur du type appartient à l'ensemble: {*premier, dernier, suivant[i], précédent[i]*}, *i* étant un nombre entier positif. *Premier* et *dernier* indiquent un référencement absolu : l'UT à afficher doit être la première (resp. la dernière) des UT cibles potentielles (= vérifiant les conditions spécifiées pour la cible), *suivant[i] et précédent[i]* indiquent un référencement relatif, l'index *i* permettant alors de spécifier le rang de la cible recherchée. Par exemple, le type « *précédent[1]* » s'interprète comme la recherche, dans les des UT cibles potentielles, de la première unité textuelle située avant l'unité textuelle source. La représentation d'une opération de navigation est montrée dans la figure 20.

$$OpNav = < Condition_{Source}, Condition_{Cible}, TypeOpération >$$

Figure 20. Définition d'une opération de navigation

Du point de vue de la navigation textuelle, une vue a toujours une *UT active*, c'est-à-dire une UT sur laquelle l'utilisateur est positionné. Lors de la création d'une vue, sa première UT active doit donc être définie.

Quand l'exécution d'une opération de navigation réussit (*i.e.* une cible est déterminée pour une source donnée), l'impact sur la vue se concrétise par le changement de l'UT active, si l'UT cible est différente de l'UT source. La représentation formelle de l'exécution d'une opération de navigation est donnée dans la figure 21.

ExécuterOpNav (ut_{Source}, opnav, vue, texte**)**

SI vérifie(*ut_{Source}*, *opnav*.$Condition_{Source}$)

ALORS

ut_{Cible} = sélectionner(*opnav*.$Condition_{Cible}$, *opnav*.TypeOpération, *texte*)

définirUTActive(*ut_{Cible}*, *vue*)

Figure 21. Exécution d'une opération de navigation

Si l'on veut offrir à un lecteur la possibilité de parcourir un texte « en suivant les sentiments », c'est-à-dire, si par exemple comme nous l'avons vu au §2.1.4.2., on s'intéresse à un personnage du roman et on veut suivre les sentiments qu'il ressent, pour qui les éprouve t-il ?, leur intensité, un autre personnage éprouve t-il un sentiment pour lui ?, si oui, lequel, de quelle intensité est-il, comment se modifie ce sentiment au fil du roman ?, etc., ces déplacements à travers le texte, si certaines conditions sont remplies, se traduisent par des opérations de navigation de Sextant : il y a une source, une cible et des conditions pour passer de l'une à l'autre.

Supposons que l'on souhaite, se déplacer d'une phrase du roman dans lequel le système a repéré une marque qui indique qu'Emma Bovary éprouve de l'amour pour quelqu'un, à une autre position dans la suite du roman (en allant vers la fin) où c'est elle qui est l'objet de l'amour de quelqu'un (sans condition sur la personne). Une représentation pseudo-formelle de cette opération de navigation est donnée dans la figure 22.

Source Expérienceur = « Emma »

Condition sur la source : Classe Sémantique = « Amour »

Cible Objet = « Emma »

Condition sur la cible : Classe Sémantique =« Amour »

Type de lecture : « narrative » (= suite du roman, vers la fin)

Figure 22. Exemple d'opération de navigation (1/2)

Par contre, si on veut aller à une phrase du roman où Emma éprouve aussi un sentiment d'amour pour quelqu'un (sans condition sur la personne qui est l'objet de cet amour, qui peut donc être différente de la précédente) mais avec force plus grande, ce qui correspond à la classe **Passion**, toujours dans la suite du roman (en allant vers la fin), la condition sur la cible ne sera pas la même comme le montre la figure 23.

Il s'agit bien d'une navigation qui, conceptuellement, est un chemin dans un graphe orienté dont les arcs correspondent, par exemple, à la force des sentiments des personnages du roman.

Source Expérienceur = « Emma »

Condition sur la source : Classe Sémantique = « Amour »

Cible Expérienceur = « Emma »

Condition sur la cible : Classe Sémantique = « Passion »

Type de lecture : « narrative » (= suite du roman, vers la fin)

Figure 23. Exemple d'opération de navigation (2/2)

La représentation de l'opération de la figure 22, que l'on peut définir comme « Aller où Emma est l'objet d'un sentiment identique », dans le formalisme imposé par le langage Sextant, est reproduite dans la figure 24.

La source et la cible sont typées : « TU « Marque_Sentiment » et les conditions s'appliquent sur les attributs de ces UT. Quand on déclenche cette opération, d'une phrase où Emma ressent de l'amour pour quelqu'un, le lecteur va directement à une phrase où c'est Emma qui est l'objet de l'amour de quelqu'un (cf. figure 9 du §2.1.4.2.).

```
<Op_Nav Titre=" Aller où Emma est l'objet d'un sentiment identique  "
Type="Forward">
  <Source>       <Condition Type="Simple">
  <UT Type=" Marque_Sentiment  ">
  <Attribut Nom="Experienceur">Emma</Attribut>
  <Attribut Nom="Classe_Sémantique">Amour</Attribut>
                 </Condition>
  </Source>
  <Cible>        <Condition Type="Simple">
  <UT Type=" Marque_Sentiment  ">
  <Attribut Nom="Object">Emma</Attribut>
  <Attribut Nom="Classe_Sémantique">Amour</Attribut>
                 </Condition>
  </Cible>
</Op_Nav>
```

Figure 24. Exemple d'opération de navigation en langage Sextant

3.3. REPRÉSENTATION DES EXPRESSIONS D'OPINION

Rappelons que nous avons décrit les expressions d'opinion à l'intérieur d'un segment de discours. Nous avons défini un segment de base comme une proposition contenant une expression d'opinion ou une séquence de propositions (reliées à un segment exprimant l'opinion par une relation de discours) exprimant une relation rhétorique avec un segment exprimant une opinion. Les segments complexes sont composés de segment de base reliés par des relations de discours.

Les objectifs que nous avons poursuivis dans l'étude des opinions sont différents de ceux de notre étude des sentiments. Nous voulons repérer les expressions d'opinion dans un ou plusieurs textes, les décrire et en faire une évaluation locale au niveau du segment contenant cette expression et globale au niveau du texte. Ce travail est actuellement fait manuellement, mais notre but final est de le faire automatiquement.

Nous avons fait une classification détaillée des opinions avec une fine granularité, en partie à partir de celles des sentiments, mais nous n'avons pas cherché à en faire une description linguistique détaillée. Ces différences expliquent que nous n'avons repris que partiellement les mêmes modes de représentation des connaissances que ceux définis pour le domaine des sentiments.

3.3.1. Représentation d'un segment d'opinion

La représentation sémantique que nous avons adoptée pour décrire un segment d'opinion est celle d'une structure de traits contenant cinq attributs : la catégorie sémantique à laquelle il appartient (Catégorie), la modalité associée (Modalité), celui qui émet l'opinion (Émetteur de l'opinion), le sujet ou thème de l'opinion (Topique), et le déclencheur lexical qui a permis d'identifier le segment d'opinion (Lexicalisation). Une modalité est définie comme un degré de préférence (Préf*i*) pour les expressions de la catégorie ***Conseil***, une combinaison de degré d'Engagement (C*i*) et de leur Force dans la catégorie ***Reportage***, ou une combinaison de Polarité (positive, négative, neutre) et de leur Force pour les expressions des catégories de ***Jugement*** et de ***Sentiment***. Des dimensions scalaires simples ont été utilisées pour représenter la force, les valeurs 2, 1 et 0 signifiant respectivement que l'expression a une forte, moyenne ou faible intensité. Les classes **Recommander**, **Suggérer** et **Espérer** sont associées respectivement aux modalités Pref1, Pref2, Pref3, telles que Pref1>Pref2>Pref3. Les classes **Informer** et **Soutenir** sont associées à la modalité C1, les classes **Dire** et **Remarquer** à C2 et les classes **Penser** et **Supposer** à C3, et on a la gradation C1>C2>C3. Par exemple, le segment [*Ce film est époustouflant*] a la structure de traits représentée dans la figure 25.

Quand les compléments du verbe contiennent une expression d'opinion, il y a un attribut supplémentaire (Contenu) dans la structure de traits qui décrit le contenu de l'expression d'opinion introduite par le verbe. Cet attribut est principalement utilisé pour les verbes de la catégorie ***Reportage***, sa valeur est également une structure de traits.

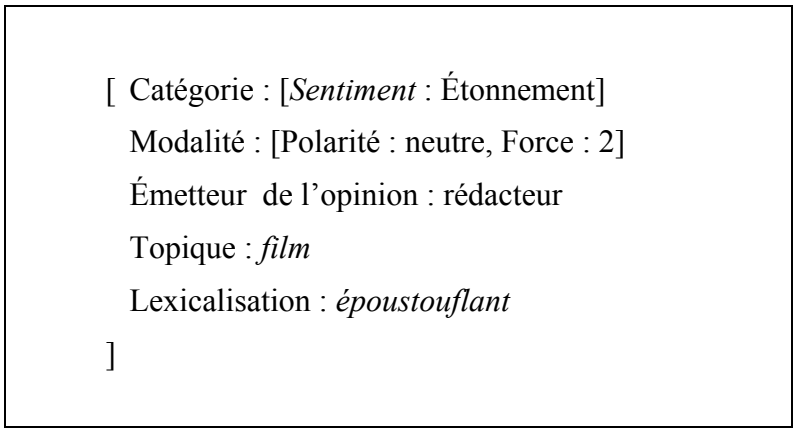


Figure 25. Structure de traits du segment [*Ce film est époustouflant*]

Par exemple, le segment [*Le Ministre a souligné que* [*trois personnes sur quatre étaient favorables à une politique des quotas*.]][28] est représenté par la structure de traits de la figure 26. Les valeurs des attributs ne sont pas dupliquées, elles sont représentées par des nombres qui seront remplacés par les valeurs grâce à un mécanisme d'unification. Ainsi, dans l'exemple de la figure 26, la valeur du Topique est obtenue en remplaçant le nombre (1) par la valeur du trait indicé par (1) c'est-à-dire la structure de traits du Contenu.

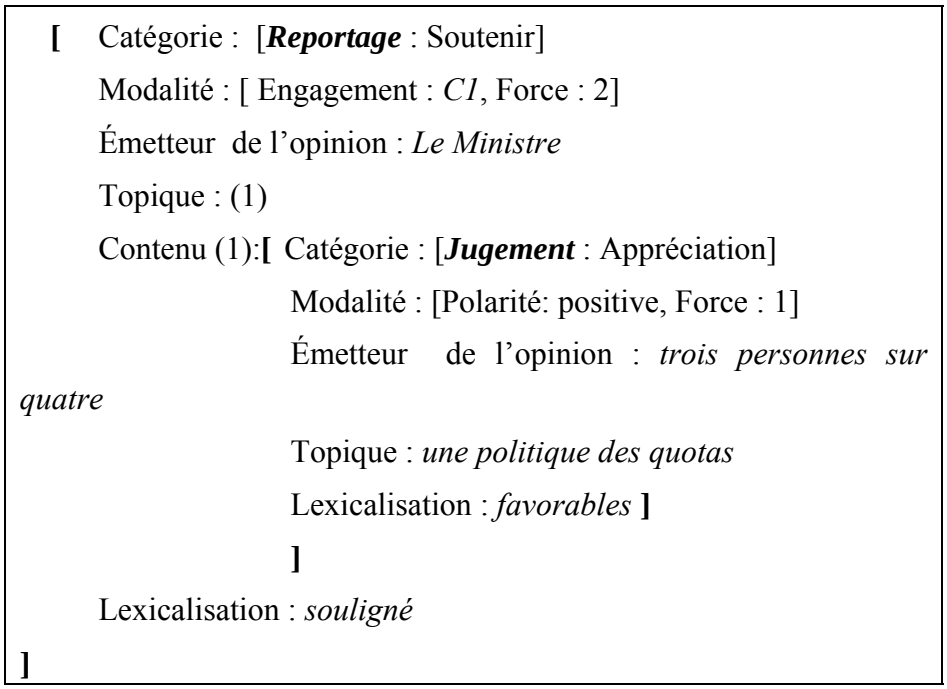


Figure 26. Structure de traits du segment *[Le Ministre a souligné que [trois personnes sur quatre étaient favorables à une politique des quotas.]]*

[28] Le Monde 21 janvier 2008

3.3.2. Représentation discursive d'un texte

La représentation discursive du texte à analyser est faite à l'aide de graphes. Par exemple, le commentaire **C** du film (contenu dans un DVD) suivant : *[ce film est époustouflant.]*a. *[[On peut ne pas être complètement convaincu]*b.1*, mais [on en ressort bouleversé]*b.2*].[[c'est poignant]*c.1*, [triste]*c.2*] et à la fois [superbe]* c.3*]*d.*[Achetez le]*e. est représenté par le graphe de la figure 27.

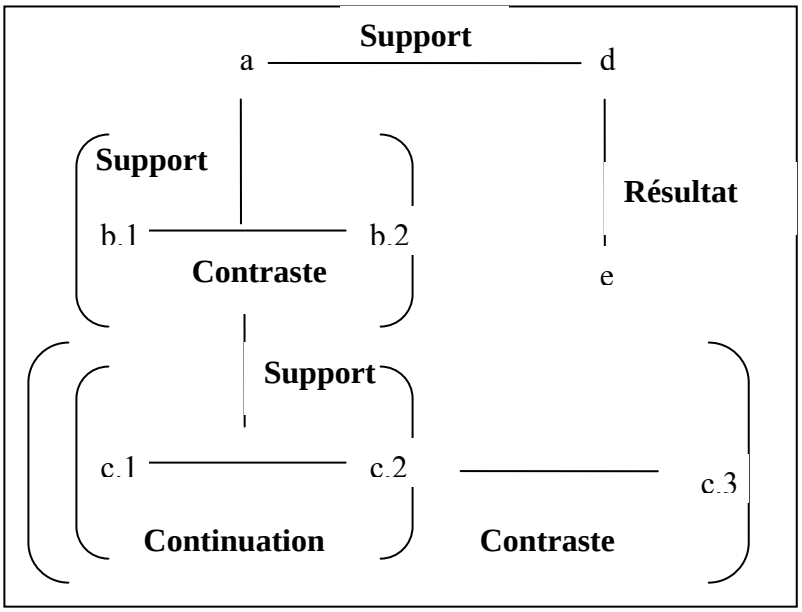


Figure 27. Exemple de représentation discursive d'un texte.

3.3.3. Représentation générale des opinions dans un texte

Une fois cette représentation discursive du texte construite, la combinaison des différentes structures de traits à l'aide des relations rhétoriques, permet d'obtenir la représentation générale du texte. Nous reprenons ici une partie décrite en détail dans (Asher *et al.* 2008b).

La combinaison de 2 structures de traits est effectuée en 2 étapes successives :

(1) on combine les structures de traits reliées par des relations de coordination (**Contraste** et **Continuation**). Ainsi dans l'exemple de la figure 27, on construit des nouvelles structures de traits à partir des segments b.1 et b.2, des segments c.1 et c.2 , puis c.3.

(2) on combine les structures de traits reliées par des relations de subordination (**Support** et **Résultat**), en opérant de façon ascendante. Dans l'exemple de la figure 27, la structure de traits du segment est combinée avec les structures déduites de l'étape 1. La formalisation de cette procédure est donnée figure 28.

Pendant ce procédé, un ensemble de règles dédiées est utilisé. La relation de **Continuation** peut, par exemple, renforcer la polarité de l'opinion en cours, comme le

montre l'extrait de règle donné figure 29 qui permet d'analyser la phase « *un film poignant et triste »*.

Soit a, b deux segments relies par la relation rhétorique R telle que : R(a, b).

Soit S_a, Sb les structures de traits associées respectivement à a et b

i.e S_a : [catégorie : [$group_a$: $subgroup_a$], modalité : [polarité : p_a, force : s_a] · · ·]

et S_b : [catégorie : [$group_b$: $subgroup_b$],modality : [polarity : p_b, strength : s_b] · · ·]

alors la structure de traits déduite de la combinaison de Sa et S_b est

S : [catégorie : [groupe],modalité : [polarité : p, force : f] · · · **]**

Figure 28. Combinaison de deux structures de traits

En effet, cette phrase contient deux segments S_a = « *un film poignant »* et S_b *« un film triste »*. Les caractéristiques de S_a sont [$groupe_a$ = Sentiment, sous-$groupe_a$ = émouvoir, $polarité_a$ = neutre, $force_a$=2], les caractéristiques de S_b sont [$groupe_b$ = Sentiment, sous-$groupe_b$ = tristesse, $polarité_b$= négative, $force_b$=1]. Par application de la règle (figure 29), on peut déduire que la phase entière exprime une opinion de type Sentiment de polarité négative de force 2.

si ($groupe_a$ = $groupe_b$) **et** (sous-$groupe_a$ ≠ sous-$groupe_b$))
alors si ((p_a = neutre) **et** (p_b ≠ neutre))
alors groupe = $groupe_a$ **et** p = p_b **et** f = max(f_a, f_b),

Figure 29. Extrait de règle pour la relation de **Continuation**

L'ensemble des opinions d'un texte est représenté par un graphe $\mathcal{G}$= ($\mathcal{N}$, $\mathcal{R}$), où $\mathcal{N}$ est un ensemble de nœuds tel que $\mathcal{N}$ = H ∪ T avec :

- H = {ho_i}, ho_i est l'émetteur de l'opinion
- et T = {to_i : *valeur*} avec to_i le topique et *valeur* une structure de traits telle que
  valeur = [Polarité : p, Force : s, Conseil : a],
  avec p = {positif, négatif, neutre} et s, a = {0, 1, 2}.

Il y a trois types de relations : entre deux émetteurs d'opinion, entre 2 topiques, et entre un émetteur et un topique. On a donc $\mathcal{R} = \mathcal{R}H \cup \mathcal{R}T \cup \mathcal{R}H{-}T$ où :

- $\mathcal{R}T = \{(t_i, t_j)/t_i, t_j \in T$ signifie que deux topiques sont reliés via une relation **Élaboration** (une des deux relations de **Support**). Cela se produit généralement entre un topique et un sous- topique, comme un film et son scénario
- $\mathcal{R}H = \{(h_i, h_j)$, type $\}$ avec $h_i, h_j \in H$ et type = **Support** ou **Contraste}** signifie que 2 émetteurs d'opinion sont reliés soit par une relation **Support** (les émetteurs partagent le même point de vue) soit par une relation **Contraste** (les émetteurs h_i et h_j ont une opinion contrastée sur le même topique);
- et $\mathcal{R}H{-}T = \{(h_i, t_j$, type)/ $h_i \in H$ et $t_j \in T$ et type = attribution ou engagement} signifie qu'une opinion à propos d'un topique t_j est attribuée à h_i ou est émise par h_i avec un certain degré d'engagement.

Par exemple, dans « *Jean a dis que le film était horrible* », l'opinion est attribuée seulement à *Jean* car les verbes de la classe **Dire** n'indiquent rien à propos du point de vue personnel de l'auteur. Ce n'est pas le cas dans « *Jean a informé la Commission que la situation était horrible »,* car ici le rédacteur prend l'information comme établie ; en effet, l'information introduite par les verbes de la classe **Informer** est estimée vraie et établie par l'auteur (cf §2.2.1).

La figure 30 montre la représentation générale de la critique de film **C** précédente : *« ce film est époustouflant. On peut ne pas être complètement convaincu, mais on en ressort bouleversé. c'est poignant, triste et à la fois superbe. Achetez le ».*

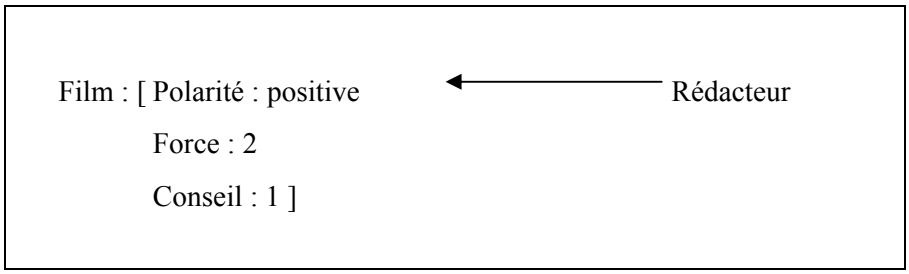


Figure 30. Exemple de représentation générale de la critique de film **C**

Une seconde représentation plus formelle (dite de « sémantique profonde ») est en cours d'élaboration. Elle permet de prendre en compte les Modalités (degré de préférence, degré d'Engagement, etc.), de les combiner aux structures rhétoriques et de calculer une opinion globale sur un thème particulier dans un texte en utilisant une logique modale propositionnelle non monotone telle celle de Asher et Lascarides (2003).

Cette analyse des opinions est actuellement manuelle, son implémentation est en projet.

3.4. ÉVOLUTION DES REPRÉSENTATIONS DES CONNAISSANCES UTILISÉES

Depuis le début de nos travaux, les représentations que nous avons utilisées ont évoluées. À chaque étape des systèmes que nous avons réalisés, la complexité des données linguistiques que nous avons traitées a été croissante. Cette complexité est à mettre en relation avec l'objet d'étude que nous avons considéré : des tables du lexique grammaire (des matrices) dans *TRANSLEGS* et de l'étude d'un item lexical, nous sommes passée à une petite grammaire du français, (qui permet la résolution d'anaphores et la génération de phrases correctes) et des connaissances sémantiques dans *INTERFACILE*, puis à des classes sémantiques, des tableaux, des graphes et des règles de déduction dans *FEELING* et *FEELING-G*. Dans ces trois derniers systèmes, l'étude a été faite au niveau de la proposition ou la phrase. Enfin, nous avons utilisé des annotations pour le filtrage sémantique et la navigation dans les textes, des relations rhétoriques discursives et des règles de combinaisons d'opinions et de relations rhétoriques pour l'étude des opinions, dans les textes également.

À cette complexité croissante des données à traiter a correspondu une complexité croissante de leurs modèles de représentation et des outils de traitement. Nous sommes ainsi passée des tables relationnelles à des représentations symboliques avec héritage et raisonnements.

Les données de *TRANSLEGS* sont sous forme de tables relationnelles, reliées par des jointures, exploitées par l'algèbre relationnel, un langage de requêtes (SQL) et des règles de déduction écrites en langage procédural (langage C). Dans *INTERFACILE,* les données sont sous forme de prédicats logiques exprimés en langage déclaratif (Prolog) avec un mécanisme d'unification qui permet de les exploiter.

Dans le domaine des sentiments, les données sont représentées sous forme de prototypes, de traits (attributs-valeurs), et de règles déclaratives dans *FEELING*, et par des structures de traits récursives dans *FEELING-G,* avec un mécanisme d'héritage, implémentées sous forme de clauses déclaratives (écrites en Prolog). Ces données sur les sentiments ont ensuite été décrites par des concepts et des bases de connaissances dans une ontologie, implémentée avec un outil de gestion des connaissances (Protégé 2000).

Dans le domaine des opinions, les données que nous traitons sont représentées sous forme de structure de traits, de graphes et de prédicats logiques.

Enfin, nous avons exprimé sous forme symbolique dans des opérations de navigation les connaissances contenues dans notre ontologie, ce qui nous a permis de faire des filtrages sémantiques et de parcourir des textes avec des objectifs précis en utilisant une plate-forme de navigation textuelle NaviTexte.

Cet accroissement de la complexité se reflète également dans les interfaces avec l'utilisateur qui ont évolué vers une interactivité et une dynamique de plus en plus grande, de *TRANSLEGS* à la navigation textuelle.

Ces différentes étapes sont synthétisées dans le tableau 15.

	DESCRIPTIONS	REPRÉSENTATIONS	OUTILS LOGICIELS	OBJET ÉTUDIÉ
AVANT LES SENTIMENTS				
TRANSLEGS	Tables du lexique-grammaire français et du lexique-grammaire italien	Relations, Algèbre relationnel, jointures, Règles de déduction	SGBD + langage procédural	Mot
INTERFACILE	Petite grammaire du français + lexique-grammaire du domaine + connaissances sémantiques du domaine	Prédicats logiques, règles de déductions, parsing , unification	Langage déclaratif	Phrase
LES SENTIMENTS				
FEELING	Lexique, classes sémantiques, tableaux, graphes et règles de déductions	Prototypes, traits attributs-valeurs, règles déclaratives, unification mécanisme d'héritage	Langage déclaratif	Phrase
FEELING -G	idem		Langage déclaratif	Phrase
ONTOLOGIE	idem	Concepts et relations	Gestionnaire d'ontologie	Phrase
NAVIGATION		Annotations, opérations de navigation	Plate forme de navigation textuelle	Texte
LES OPINIONS				
	Classes d'opinions, relations rhétoriques discursives, règles de combinaisons d'opinions et de relations rhétoriques discursives, lambda terms	Structure de traits, graphes, algèbre, logique		Texte

Tableau 15. Évolution des représentations des connaissances utilisées.

CONCLUSION

Au début des années 2000, différents bouleversements ont jalonné nos travaux. Quand nous avons commencé à étudier les sentiments, nous avions pour objectif des descriptions linguistiques et nous les avons établies avec des emplois construits, non attestés en corpus. Un premier basculement a été une modification de perspective dans nos recherches avec le souci de pouvoir utiliser ces descriptions comme ressources. Parallèlement avec la disposition de nombreux textes électroniques, nous avons pu vérifier les emplois en corpus et passer d'une linguistique « construite » à une linguistique « attestée », évolution qui a été également largement répandue dans la communauté linguistique et du TAL.

La possibilité d'utiliser nos descriptions a entraîné un autre tournant important, en nous amenant à réfléchir aux moyens à mettre en œuvre pour rendre accessible à des utilisateurs humains ou informatiques ces descriptions. En effet, une partie des représentations que nous avons adoptées dans notre étude sur les sentiments sont très proches de celles proposées dans les ontologies utilisées en Intelligence Artificielle dans le domaine du Web sémantique. Nous avons pu ainsi aisément en traduire certaines, comme par exemple l'organisation en classes qui sont devenues les concepts de l'ontologie, liés entre eux par des relations qui représentent certains de nos graphes (intensité et antonymie).

La récente disponibilité de formats standards (XML et sa galaxie RDF, RDFS, OWL, etc.) exploités par des outils de gestion et d'interrogation des ontologies, comme Protégé-2000, permet maintenant d'exploiter une partie des ressources linguistiques que nous avons construites. En contrepartie, nous sommes obligée d'accepter une part d'approximation et une dégradation dans nos représentations, liées aux formalismes utilisés.

Nous nous intéressons à présent à la détection et à l'annotation des sentiments et des opinions non plus au niveau du mot et de la phrase mais dans les textes et en corpus. Ce passage d'une linguistique construite à une linguistique textuelle de corpus a entraîné que les propriétés qui nous intéressent ne sont plus exactement les mêmes. Au début de notre étude sur les sentiments, nous avons établi la liste des verbes à partir de dictionnaires, et construit les exemples d'emploi pour la classification et l'étude des propriétés à partir de notre intuition personnelle que nous avons vérifiée auprès de locuteurs dont la langue maternelle est le français. Par la suite, la généralisation de corpus électroniques a entraîné une modification de cette classification et de l'acceptabilité de certaines propriétés. Ainsi, nous avons trouvé via un moteur de recherche du Web, des emplois attestés de « Nom hum *désemparé de* »[29] alors que notre propre intuition (et celle de Maurice Gross dans la table 4 du Lexique grammaire) pensait cet emploi inacceptable pour le verbe *désemparer*. Mais comme le remarque Celine Raynal (2005)

> « Quels statuts donner à des structures attestées uniquement sur le Web et en très petite quantité ? » page 140.

À l'inverse, certaines propriétés que nous avions estimées pertinentes dans une optique de TAL, pour générer des paraphrases par exemple, comme les constructions *N_1 se V de ce que P* ou la possibilité d'un sujet phrastique (*Que P V N_1*), se rencontrent très peu dans la langue des corpus (en 2008).

Nous continuons à penser, cependant, que l'introspection et l'intuition sont nécessaires pour mettre en évidence ces propriétés que l'on risquerait de ne pas remarquer en ayant recours uniquement aux corpus. Or, ces propriétés doivent être prises en compte car notre propre intuition dans notre langue maternelle est valide, et dès lors que la langue est un système on ne peut générer n'importe quelle phrase.

Ce recours, ou non, au corpus fait l'objet d'un large débat dans la communauté. La position d'un extrême est que « *La langue c'est les corpus* », et l'on décrit, par exemple, un phénomène syntaxique uniquement à partir d'un corpus censé représenter le fonctionnement d'une langue ; de nombreux travaux en linguistique computationnelle ont cette approche. Le choix du corpus est donc crucial, et l'on ne doit pas perdre de vue que ce n'est le reflet que d'une partie de la langue. À l'opposé, certains travaux, en linguistique notamment, reposent encore sur l'introspection « pure ».

Nous adoptons une position médiane, assez répandue, qui est d'utiliser à la fois l'intuition et l'introspection, et les corpus pour attester les emplois ou rechercher des informations.

[29] « Escobar **désemparé de** douleur », Libération 27 octobre 2005, « Œil-de-Faucon fut **désemparé de** la mort de sa femme » *http://hyperion2000.free.fr,*

Une autre problématique est venue depuis quelques années questionner nos recherches, il s'agit de l'exploitation possible des ressources que nous avons construites pour assister un lecteur.

De nombreux livres électroniques (« ebooks ») sont maintenant disponibles, scannés ou numérisés et présentés en mode texte ou dans des formats adaptés à la lecture sur écran (de périphérique électronique dédié ou d'ordinateur).

Le mode de lecture d'un document dépend de son support : « *on ne lit pas sur papier comme on lit sur écran* »[30], et le support numérique a des modalités de consultation propres. Lors des *Quinzièmes entretiens Jacques Cartier* (2002) sur le thème : « Les défis de la publication sur le Web », Christian Vandendorpe écrivait :

> « [...] La question est de déterminer si la nouvelle technologie va remplacer le livre ou si, au contraire, elle va exercer une telle pression sur nos modes de lecture que la forme incarnée traditionnellement par le livre et la littérature va tout simplement tomber en désuétude.
>
> [...] Plus précisément, le Web est devenu, grâce aux moteurs de recherche, une gigantesque base de données. Et celle-ci instaure un type de textualité qui lui est propre, organisé non plus en fonction d'une lecture continue ou d'une récitation à haute voix, mais selon les paramètres les plus susceptibles de correspondre à des besoins spécifiques de lecture et d'information. »

Si la lecture sur un support papier est le plus souvent une lecture continue et d'un seul document à la fois, on utilise des outils de navigation pour se positionner directement à des endroits particuliers d'un livre ou d'un document numérique en général, selon des objectifs que l'on a choisis.

Avec l'arrivée de ce nouveau support qui se répand rapidement, un nouveau champ d'investigation est la conception d'outils de navigation dans des textes. Des annotations dédiées à des applications particulières permettent, grâce à des plates formes de navigation « intelligentes », de se déplacer dans un document, ou dans plusieurs documents, selon des critères spécifiés (sur les annotations). Ce parcours peut être réalisé aussi bien vers le début que vers la fin du document.

Ces outils amènent de nouveaux besoins comme des modes de lecture « dynamiques », par exemple en suivant les sentiments des différents personnages d'un roman, comment se transforment ces sentiments au fil du texte, etc., ou encore en s'intéressant à l'évolution des opinions, de quelqu'un ou sur quelqu'un, un homme politique par exemple, à travers ses déclarations, ses discours, etc.

Cette transformation de l'écrit en une forme numérique offre de nouvelles possibilités de lecture, qui n'est plus linéaire et séquentielle. Ainsi, on peut « naviguer » dans des textes, grâce à des raisonnements sur des connaissances qu'on leur a associées.

---

[30] Le Monde 2, 16 février 2008, page 25

À l'inverse, si ces nouvelles technologies proposent de nouveaux modes de lecture, elles induisent de nouveaux modes d'écriture et peuvent également nous amener à exprimer nos idées et nos affects de façon différente. L'évolution vers une technologie toujours plus performante, loin d'exclure les agents humains de toutes applications, les y réintroduit. Cela nous conduira, peut être, à une meilleure compréhension des modes d'expression écrits de nos comportements et de nos émotions.

BIBLIOGRAPHIE

ASHER N., BENAMARA F., MATHIEU, Y. Y. 2008a. « Discourse Based Opinion Categorizing: A preliminary study », *Proceedings of ECAI2008*, Patras, Greece, p. 835-836.

ASHER N., BENAMARA F., MATHIEU, Y. Y. 2008b. « Categorizing Opinion in Discourse », *Proceedings of COLING2008* Companion Volume, Manchester, p. 7-10.

ASHER N., BENAMARA F., MATHIEU, Y. Y. 2008c. « Appraisal of Opinion Expressions in Discourse », *Actes du 27e Colloque International sur le Lexique et la Grammaire*, L'Aquila, Italie, p. 223-229.

ASHER N., LASCARIDES A. 2003. *Logics of Conversation*, Cambridge University Press.

BELLETI A., RIZZI L. 1988. « Psych-Verbs and θ–theory », *Natural Language and Linguistic Theory* 6, p. 291-352.

BETHARD S., YU H., THORNTON A., HATZIVASSILOGLOU V., JU RAFSKY D., 2004. « Automatic Extraction of Opinion Propositions and their Holders », *Proceedings of AAAI04.*

BOONS J.-P., GUILLET A., LECLÈRE C. 1976. *La structure des phrases simples en français. Constructions intransitives*, Genève, Droz.

BOURIGAULT D., 2002. « Upery : un outil d'analyse distributionnelle étendue pour la construction d'ontologies à partir de corpus », *Actes de TALN 2002*, Nancy, p. 75-84.

CHOI Y., CARDIE C., RILOFF E., PATWARDHAN S. 2005. « Identifying Sources of Opinions with Conditional Random Fields and Extraction Patterns », *Proceedings of HLT/EMNLP 2005.*

COUTO J. À paraître. « Exploiter par navigation textuelle les informations filtrées », In *Filtrage sémantique dans les textes. Approches symboliques*, J-L Minel (Ed.), Paris, Hermès Sciences, Lavoisier.

COUTO J., MINEL, J.-L. 2006. « Navigation textuelle : représentation des textes et des connaissances », *TAL*, Vol 47/2, p. 225−254.

COUTO J., MINEL, J.-L. 2008. « A linguistic and navigational knowledge approach to text navigation », *Third International Joint Conference on Natural Language Processing*, Hyderabad, India, p. 667-672.

DAELEMANS W., DE SMEDT K., GADZAR G. 1992. « Inheritance in Natural Language Processing », *Computational Linguistics* ; 18/2, p. 205-218.

DANLOS L., SAGOT B. 2007. « Comparaison du *Lexique-Grammaire* des verbes pleins et de DICOVALENCE : vers une intégration dans le Le*fff* », *Actes de TALN 2007*, Toulouse.

DAVITZ J. R., 1969. *The Langage of Emotion*, New York and London, Academic Press.

DESCARTES R. 1649/1988. *Les passions de l'âme*, Gallimard.

DUBOIS J., DUBOIS-CHARLIER F. 1997. *Les Verbes français*, Paris, Larousse-Bordas.

DUGAS A., COURTOIS B., MATHIEU Y. Y. 1992. « O stories », In *First Korean-French Colloquium : The Grammar and the Lexicon for NLP*, Séoul.

DUGAS A., MATHIEU, Y. Y., COURTOIS B. 1995. « Formes simples en *o-* », *Linguisticae Investigationes*, XIX/2, p. 429-441.

ELIA A., 1984. *Le verbe italien*, Biblioteca della Ricerca, Fasano di Puglia Italie, Paris, France, Schena-Nizet.

ELIA A., MATHIEU Y. Y. 1986. « Computational Comparatives Studies on Romance Language. A Linguistic Comparison of Lexicon-Grammars », Invited paper In *11th International Conference on Computational Linguistics*, *COLING'86*, Bonn, p. 146-150.

ESULI A., SEBASTIANI F. 2006. « SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining », *Proceedings of LREC 06*.

FALK I., FRANCOPOULO G., GARDENT C. 2007. « Evaluer SynLex. », *Actes de TALN'07*, Toulouse.

FELLBAUM C. 1999. « La représentation des verbes dans le réseau sémantique WordNet », *Langages* 33, p. 27-40.

FILLMORE C. J. 1988. « The mechanisms of 'Construction Grammar. », *BLS* 14, p. 35-55.

FLAUBERT, G. 1857/2001. *Madame Bovary*, Paris, Gallimard.

FRAMENET http://framenet.icsi.berkeley.edu

FRANÇOIS J, LE PESANT D., LEEMAN (eds.). 2007. *Classements syntactico-sémantiques des verbes français, Langue Française* 153.

FRANTEXT http://www.frantext.fr.

GARDENT C., GUILLAUME B., PERRIER G., FALK I. 2006. « Extraction d'information de sous-catégorisation à partir des tables du LADL », *Actes de Traitement Automatique des Langues Naturelles*, Louvain, Belgique.

GRACQ, J. 1972. *ANDRÉ BRETON*, Paris, José Corti.

GROSS M. 1975. *Méthodes en syntaxe*, Paris, Hermann.

GROSS M. 1981. « Les bases empiriques de la notion de prédicat sémantique », *Langages* 63, Paris, Larousse.

GUILLET A., LECLÈRE C. 1992. *La structure des phrases simples en français. Constructions locatives*, Genève, Droz.

HALLIDAY M.A. K. 1994/2004. *An Introduction to Functional Grammar,* London, Edward Arnold.

HARRIS Z. 1971. *Structures mathématiques du langage*, Paris, Dunod.

HARRIS Z. 1990. *La genèse de l'analyse des transformations et de la métalangue*, *Langages* 99, p. 9-20..

JACKENDOFF R. 1990. *Semantic Structures*, Cambridge, Massachussets, The MIT Press.

JOHNSON-LAIRD P.N., OATLEY K. 1989. « The Language of Emotions : An Analysis of a Semantic Field », *Cognition and Emotion* 3, p. 81-123.

KIM S. M., HOVY E. 2005. « Identifying Opinion Holders for Question Answering in Opinion Texts. », *Workshop on Question Answering in Restricted Domain, AAAI-05*.

KIM S. M., HOVY E. 2006. « Identifying and analyzing judgment opinions », *HLT NAACL*, New York.

KOEVA S., MAUREL D., SILBERZTEIN M. 2007. *Formaliser les langues avec l'ordinateur : de Intex à Nooj,* PU de Franche-Comté.

LE PETIT ROBERT. 2001. Version électronique.

LEEMAN D. 1991. « *Hurler de rage, rayonner de bonheur* : remarques sur une construction en *de* », *Langue française* 101, p. 80-101.

LEEMAN D. 1995. « Pourquoi peut-on dire Max est en colère mais non **Max est en peur* ? Hypothèse sur la construction *être en N* », *Langue française* 105, p. 55-69.

LEVIN B,. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*, Chicago, IL, University of Chicago Press.

MARTIN J. R., WHITE P.R.R. 2005. *Language of Evaluation: Appraisal in English*, Palgrave Macmillan.

MATHIEU, Y. Y., SABATIER P. 1986. « INTERFACILE : a Linguistic Coverage and Query Reformulation », In *11th International Conference on Computational Linguistics*, *COLING'86*, Bonn, p. 46-49.

MATHIEU, Y. Y. 1989. « TRANSLEGS : une station de travail linguistique », In *Colloque International sur la description des langues naturelles en vue d'applications informatiques*, Québec, Université Laval, p. 309-320.

MATHIEU, Y., 1994. *Interprétation par prédicats sémantiques de structures d'arguments. FEELING : une application aux verbes psychologiques*, Thèse en Informatique fondamentale, Paris, Université Paris 7.

MATHIEU Y. Y. 1998a. « Étude pour une extraction semi-automatique des néologismes », In *Cinquièmes Journées Scientifiques : « La mémoire des mots »,* septembre 1997*,* Tunis. p. 55-460

MATHIEU Y. Y., FOUQUERÉ C., GROSS G. 1998b. « Vers une extraction automatique des néologismes », *Cahiers de lexicologie* 72, p. 199-208.

MATHIEU, Y. Y. 1999. « Les prédicats de sentiment », *Langages* 136, p. 41-52.

MATHIEU, Y. Y. 2000. *Les verbes de sentiment. De l'analyse linguistique au traitement automatique*, Paris, CNRS Editions.

MATHIEU, Y. Y. 2004. « Linguistic Knowledge and Automatic Semantic Representation of Emotions and Feelings. », *International Conference on information Technology (ITCC2004)*, IEEE Computer Society, p. 314-318.

MATHIEU, Y. Y., 2005a. « Annotations of Emotions and Feelings in Texts », *Conference on Affective Computing and intelligent Interaction (ACII2005)*, Dordrecht, Springer Lecture Notes in Computer Science, p. 350-357.

MATHIEU, Y. Y. 2005b. « A Computational Semantic Lexicon of French Verbs of Emotion », In Shanahan, G., Qu, Y., Wiebe, J. (eds.): *Computing Attitude and Affect in Text*, Dordrecht, Springer, p. 109-123.

MEL'ČUK I., 1992. *Dictionnaire explicatif et combinatoire du français contemporain*, Recherches lexico-sémantiques III, Montréal, Les presses de l'Université de Montréal.

MILLER G., FELLBAUM C. 2007. « Wordnet Then and Now », *Language Resources and Evaluation* 41:2, Netherlands, Springer, p. 209-214.

MILLER, G. A. 1990. « Nouns in WordNet: a lexical inheritance system. », In *International Journal of Lexicography* 3 (4), p. 245-264.

MILLER, G. A., FELLBAUM C. 1991. « Semantic networks of English. », In *Cognition* 41, p. 197-229.

MILLER, G. A., FELLBAUM C., MILLER J. K. 1993. *Five papers on WordNet.* ftp://ftp.cogsci.princeton.edu/pub/wordnet/5papers.ps

MINEL J-L., À paraître. « Le filtrage sémantique dans les textes», In *Filtrage sémantique dans les textes. Approches symboliques*. J-L Minel (éd.). Paris, Hermès Sciences, Lavoisier.

ORTONY A, CLORE G.L, COLLINS A. 1988. *The Cognitive Structure of Emotions,* Cambridge University Press.

OSGOOD C.E., 1979. « From Yang And Yin To And Or But In Cross-Cultural Perspective », *International Journal of Psychology* 14**,** p. 1-35.

OSGOOD C.E., SUCI G.J., TANNENBAUM P.H. 1957. *The measurement of meaning*, Urbana, University of Illinois Press.

PANG ,B.; LEE L., VAITHYANATHAN S. 2002. « Thumbs up? Sentiment Classification using Machine Learning Techniques », *Proceedings of EMNLP 2002*.

PROTÉGÉ-2000. The Protégé Project, http://protege.stanford.edu

RUSSELL J. A., PRATT G. A. 1980. « A Description of the Affective Quality Attributed to Environments », *Journal of Personality and Social Psychology* 38, p. 311-322.

RUWET N. 1993. « Les verbes dits psychologiques : trois théories et quelques questions », *Recherches linguistiques de Vincennes* 22, p. 95-124.

RUWET N. 1994. « Être ou ne pas être un verbe de sentiment », *Langue française* 103, p. 45-55.

RUWET, N. 1995. « Les verbes de sentiment peuvent-ils être agentifs? », *Langue française* 105, p. 28-39.

SHANAHAN, G., QU, Y., WIEBE, J. eds., 2005. *Computing Attitude and Affect in Text*, Dordrecht, The Netherlands, Springer.

STRAPPARAVA C., VALITUTTI A. 2004. « WordNet Affect: an Affective Extension of WordNet », *Proceedings of LREC 2004*.

TRÉSOR DE LA LANGUE FRANÇAISE INFORMATISÉ, http://atilf.atilf.fr/tlf.htm.

TURNEY P., 2002. « Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews », *Proceedings of 2006 International Conference on Intelligent User Interfaces (IUI06)*.

USCHOLD, M., GRUNINGER, M. 1996. « Ontologies: Principles, Methods and Applications », *Knowledge Engineering Review* 11(2).

VAN DE VELDE, D. 1999. « La multiplication des sentiments », *Travaux de Linguistique* 38, p. 63-87.

VOORST J. V. 1992. « The Aspectual Semantics of Psychological Verbs », *Linguistics and Philosophy* 15, p. 65-92.

WHITELAW C., GARG N., ARGAMON S. 2005. « Using Appraisal Taxonomies for Sentiment Analysis », *Proceedings of Second Midwest Computational Linguistic Colloquium (MCLC 2005)*.

WIEBE J., WILSON T., CARDIE C. 2005. « Annotating Expressions of Opinions and Emotions in Language », *Language Resources and Evaluation* 1(2).

WIERZBICKA A. 1987. *Speech Act Verbs*, Sydney, Academic Press.

WIERZBICKA A. 1993. « La quête des primitifs sémantiques », *Langue française* 98, p. 9-23.

WILSON T., WIEBE J., HWA R. 2006. « Recognizing strong and weak opinion clauses », *Computational Intelligence* 22 (2), p. 73-99.

WORDNET, http://wordnet.princeton.edu/

YU H., HATZIVASSILOGLOU V. 2003. « Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences », *Proceedings of EMNLP 03*.

TABLE DES FIGURES

Figure 1. La relation d'hyperonymie pour « Doberman » (WordNet 2.1).....26
Figure 2. Exemples de relations d'intensité entre classes sémantiques 42
Figure 3. Exemple de relation de sens entre classes sémantiques (1/3).....42
Figure 4. Exemple de relation de sens entre classes sémantiques (2/3).....42
Figure 5. Exemples de relation de sens entre classes sémantiques (3/3) 43
Figure 6. Exemples de relation entre classes sémantiques.....43
Figure 7. Annotation d'un extrait de *Madame Bovary* 50
Figure 8. Processus d'annotation 51
Figure 9. Exemple d'écran *NaviTexte*.....52
Figure 10. Exemples d'antonymie dans WordNet 67
Figure 11. Exemple de comparaison des Lexiques grammaires italien et français 71
Figure 12. Propriétés de substitution du verbe *douter* 71
Figure 13. Construction d'une vue dynamique dans le système *TRANSLEGS* 72
Figure 14. Mécanisme d'héritage dans le système *FEELING* 80
Figure 15. Représentation sémantique de *Luc exaspère Marie* 82
Figure 16.. Composition d'un texte 84
Figure 17. Exemple d'annotation d'un texte.....86
Figure 18. Exemple de l'opérateur UT 87
Figure 19. Opérateurs d'existence sur les éléments des UT 87
Figure 20. Définition d'une opération de navigation.....88
Figure 21. Exécution d'une opération de navigation 89
Figure 22. Exemple d'opération de navigation (1/2) 89
Figure 23. Exemple d'opération de navigation (2/2) 90
Figure 24. Exemple d'opération de navigation en langage Sextant.....90
Figure 25. Structure de traits du segment [*Ce film est époustouflant*].....92
Figure 26. Structure de traits du segment *[Le Ministre a souligné que [trois personnes sur quatre étaient favorables à une politique des quotas.]]*.....92
Figure 27. Exemple de représentation discursive d'un texte. 93
Figure 28. Combinaison de deux structures de traits 94
Figure 29. Extrait de règle pour la relation de Continuation 94
Figure 30. Exemple de représentation générale de la critique de film C 95

TABLE DES TABLEAUX

Tableau 1. Organisation générale des tables du Lexique grammaire des verbes..... 10
Tableau 2. Tables des verbes à constructions complétives..... 11
Tableau 3. Extrait de la table 4 du Lexique grammaire (Gross 1975)..... 12
Tableau 4. Classes des verbes anglais par Beth Levin (1993)..... 18
Tableau 5. Contenu du réseau lexical WordNet (juillet 2008)..... 24
Tableau 6. Domaines sémantiques des verbes de WordNet..... 27
Tableau 7. Classes génériques et syntaxico-sémantiques des verbes français (Dubois et Dubois–Charlier 1997)..... 33
Tableau 8. Organisation des verbes psychologiques dans le dictionnaire des verbes français de Dubois et Dubois-Charlier (1997)..... 35
Tableau 9. Classes sémantiques des sentiments..... 41
Tableau 10. Classes des verbes de parole anglais par Anna Wierzbicka (1987)..... 57
Tableau 11. Classification des expressions d'opinion..... 58
Tableau 12. Correspondances entre les tables de verbes à complétives italiennes et françaises..... 70
Tableau 13. Extrait de la table italienne it42..... 70
Tableau 14. Exemple de propriétés simples des verbes de sentiment..... 78
Tableau 15. Évolution des représentations des connaissances utilisées. 97