

UNIVERSITÉ DE NANTES

ÉCOLE DOCTORALE

« SCIENCES ET TECHNOLOGIES DE L'INFORMATION ET DE MATHÉMATIQUES »

Année : 2011

T H È S E

DE

D O C T O R A T

DE L'UNIVERSITÉ DE NANTES

Spécialité : INFORMATIQUE

Présentée et soutenue publiquement par

Matthieu VERNIER

le 4 juillet 2011

à la faculté de Sciences et Techniques de Nantes

TITRE

Analyse à granularité fine de la subjectivité

JURY

<i>Président :</i>	Marc EL-BÈZE, Professeur	LIA, Université d'Avignon et des Pays de Vaucluse
<i>Rapporteurs :</i>	Nicholas ASHER, Directeur de Recherche	IRIT, CNRS
	Yves BESTGEN, Professeur	F.R.S-FNRS, Université Catholique de Louvain
<i>Examineurs :</i>	Marc EL-BÈZE, Professeur	LIA, Université d'Avignon et des Pays de Vaucluse

Directeur de thèse :	Béatrice DAILLE, Professeur
Laboratoire :	LINA - UMR CNRS 6241
Co-encadrant:	Laura MONCEAUX, Maître de Conférences
Laboratoire :	LINA - UMR CNRS 6241

Analyse à granularité fine de la subjectivité

Fine-grained Subjectivity Analysis

MATTHIEU VERNIER



Université de Nantes

Table des matières

Table des matières	iii
Table des figures	vii
Liste des tableaux	ix
1 Introduction	1
1.1 L'essor du « web social » : un contexte applicatif porteur	1
1.2 La fouille d'opinions : un domaine émergent	3
1.3 Développer des méthodes et des ressources pour une analyse à granularité fine	4
2 Vers une analyse à granularité fine	7
2.1 Analyse de l'opinion au grain du document	10
2.1.1 Un problème de catégorisation de documents	10
2.1.2 Limites des approches au grain document	17
2.2 Analyse à granularité fine de l'opinion	18
2.2.1 Une combinaison de problèmes hétérogènes	18
2.2.2 Synthèse des réflexions	23
2.3 Extraction des mots d'opinions d'une langue	24
2.3.1 Trois types d'approches	24
2.3.2 Des résultats difficilement évaluables	29
2.4 Conclusion	30
2.4.1 Synthèse du domaine de la fouille d'opinions	30
2.4.2 Positionnement de nos travaux	30
3 La subjectivité dans le langage	33
3.1 La subjectivité dans le langage : un phénomène énonciatif	35
3.1.1 La problématique de l'énonciation	35
3.1.2 Un 1 ^{er} niveau de subjectivité : les traces de l'énonciateur	39
3.1.3 Un 2 ^{ème} niveau de subjectivité : les traces d'une évaluation	41
3.1.4 Synthèse : deux configurations d'énonciations subjectives possibles	45
3.2 Différentes modalités énonciatives pour exprimer une évaluation	46
3.2.1 Définition de la modalité	46

3.2.2	Les modalités logiques	47
3.2.3	Les modalités axiologiques	48
3.2.4	Combinaison de modalités	49
3.3	Vers un modèle informatique du langage de l'évaluation	50
4	Développer des ressources pour le français	51
4.1	Un corpus annoté, multithématique et en français	52
4.1.1	Les blogs : un support riche en phénomènes évaluatifs	53
4.1.2	Annotation manuelle des objets textuels, des évaluations et de leur cible	55
4.1.3	Le corpus Blogoscopie : résultats et statistiques	58
4.1.4	Travaux liés et discussion	60
4.2	Un lexique de mots évaluatifs pour le français	61
4.2.1	Du corpus annoté aux entrées lexicales	61
4.2.2	Le lexique de l'évaluation : résultats et statistiques	63
4.2.3	Travaux liés et discussion	65
4.3	Conclusion	66
5	Délimiter et catégoriser les évaluations	67
5.1	Analyse à granularité fine des évaluations	68
5.1.1	Un double objectif : délimiter et catégoriser	68
5.1.2	Modéliser l'évaluation en articulant lexique, grammaire et sémantique	72
5.2	Acquisition de structures évaluatives	75
5.2.1	Extraction des passages évaluatifs à partir d'un corpus annoté	75
5.2.2	Transformation en structures évaluatives généralisées	76
5.2.3	Annotation des méta-données sur les structures évaluatives	79
5.2.4	Résultats et discussion	80
5.3	Délimitation et catégorisation des évaluations	82
5.3.1	Méthode par structures évaluatives	82
5.3.2	Méthode lexicale (<i>baseline</i>)	83
5.4	Expérimentation : intérêts et limites des structures évaluatives	85
5.4.1	Protocole d'évaluation	85
5.4.2	Une bonne délimitation des évaluations	87
5.4.3	Une couverture trop faible du vocabulaire évaluatif	88
5.4.4	Une bonne catégorisation des traits sémantiques et énonciatifs	89
5.4.5	Un intérêt pour résoudre les effets du contexte sur l'axiologie	90
5.4.6	Une bonne adaptation à des corpus multi-thématiques	91
5.5	Conclusion	93
5.5.1	Synthèse	93
5.5.2	Perspectives	94
6	Améliorer la couverture du vocabulaire évaluatif	97
6.1	Une couverture insuffisante des lexiques	98
6.1.1	Un silence rarement mesuré	98
6.1.2	À l'origine du silence	99
6.1.3	Romp le silence en limitant le bruit	101
6.2	Apprentissage de subjectivèmes par hypothèses sémantiques	103
6.2.1	Les tests sémantiques : des indicateurs pour discriminer les subjectivèmes	103
6.2.2	Extraire des candidats à partir d'un corpus et du web	106

6.2.3	Construire un modèle de classification	107
6.2.4	Expérimentation et résultats	110
6.2.5	Discussion	113
6.3	Apprentissage de subjectivèmes négatifs par hypothèses morphologiques	114
6.3.1	Les affixes : des indicateurs pour discriminer les subjectivèmes	114
6.3.2	Extraire des candidats à partir d'un dictionnaire	114
6.3.3	Construire un modèle de classification	116
6.3.4	Expérimentation et résultats	118
6.4	Conclusion	121
6.4.1	Synthèse	121
6.4.2	Perspectives	121
7	Identifier la cible d'une évaluation	123
7.1	Une problématique récente peu explorée	124
7.1.1	Quatre facteurs de complexité	124
7.1.2	Approches existantes pour l'anglais	127
7.1.3	Adapter et renforcer les méthodes pour le français	129
7.2	Un corpus multithématique de référence	129
7.2.1	Protocole d'utilisation du corpus Blogoscopie	129
7.2.2	Observations statistiques	130
7.3	Méthodes d'identification de la cible évaluée	132
7.3.1	Méthodes applicables à un niveau intraphastique	133
7.3.2	Méthodes applicables à un niveau intra- et interphrastique	135
7.4	Expérimentation et résultats	137
7.4.1	Résultats pour les cas intraphastiques	137
7.4.2	Résultats pour les cas interphastiques	138
7.5	Conclusion	139
8	Conclusion	141
8.1	Un bilan de l'analyse à granularité fine de la subjectivité	141
8.1.1	Le développement de ressources francophones	142
8.1.2	La délimitation et la catégorisation des évaluations	142
8.1.3	L'amélioration de la couverture des lexiques de mots évaluatifs	143
8.1.4	L'identification de la cible d'une évaluation	144
8.2	Perspectives	145
8.2.1	Développer des ressources	145
8.2.2	Structurer le discours évaluatif	147
A	Document non annoté du corpus	151
B	Document annoté du corpus	153
C	Extrait du lexique de l'évaluation	155

Table des figures

1.1	Nombre d'articles (presses, blogs) autour de <i>Nicolas Anelka</i> en 2010	1
2.1	Problématiques de la fouille d'opinions	8
2.2	Analyse de l'opinion au grain document	11
2.3	Processus de classification avec séparateur à vaste marge (SVM)	13
2.4	Analyse à granularité fine de l'opinion	19
3.1	Les deux niveaux de subjectivité dans le langage	34
3.2	Échelle de subjectivité intrinsèque aux unités lexicales – (Kerbrat-Orecchioni, 1980) . . .	41
4.1	Développement du lexique de l'évaluation à partir du corpus Blogoscopie	62
5.1	Structure de traits d'un constituant évaluatif	73
5.2	Constituants associés à l'évaluation « <i>pas plus utiles</i> »	74
5.3	Structure évaluative généralisée (<i>pas plus utiles, jamais plus mauvais</i> , etc.)	75
5.4	Structure évaluative non généralisée issue du corpus d'entraînement	76
5.5	Généralisation des mots de la même classe sémantique et grammaticale	77
5.6	Généralisation de l'axiologie d'un mot évaluatif	78
5.7	Nombre de structures en fonction de leur fréquence dans le corpus	80
5.8	Évolution du nombre de structures par rapport au nombre d'exemples d'évaluations . . .	80
6.1	Nombre d'occurrences des noms candidats en fonction de leur nombre d'occurrences précédées d'un marqueur d'intensité	109
6.2	Nuage de mots représentant les adjectifs connus du lexique qui s'opposent (polarité inverse) le plus fréquemment à l'adjectif <i>populiste</i>	109
6.3	Nuage de mots représentant les adjectifs connus du lexique qui s'associent (polarité identique) le plus fréquemment à l'adjectif <i>hilarant</i>	109
6.4	Chaîne de traitement pour collecter et filtrer les candidats évaluatifs du wiktionnaire . . .	117
7.1	Analyse des dépendances entre une évaluation et sa cible	135
8.1	Structuration des unités évaluatives du discours	148

Liste des tableaux

1.1	Analyse fournie en sortie d'un système de fouille d'opinions	3
2.1	Bigrammes grammaticaux d'opinions – (Turney, 2002)	14
2.2	Patrons subjectifs et objectifs – (Wiebe et Riloff, 2005)	20
4.1	Accord inter-annotateurs pour l'annotation des objets et des évaluations	58
4.2	Répartition des modalités d'évaluations dans le corpus Blogoscopie	59
4.3	Répartition des configurations d'énonciations dans le corpus Blogoscopie	59
4.4	Répartition des axiologies dans le corpus Blogoscopie	59
4.5	Caractéristiques des entrées du lexique de l'évaluation	64
4.6	Caractéristiques des lexiques francophones et anglophones existants	65
5.1	Structures évaluatives acquises selon leur modalité	81
5.2	Structures évaluatives acquises selon leur configuration d'énonciation	81
5.3	Structures évaluatives acquises selon leur action sur l'axiologie	81
5.4	Résultats de la détection et de la délimitation des évaluations	87
5.5	Résultats de la catégorisation de la modalité	89
5.6	Résultats de la catégorisation de la configuration d'énonciation	90
5.7	Résultats de la catégorisation axiologique	91
5.8	Résultats de la détection des passages subjectifs (DEFT'09)	93
6.1	Nombre de termes candidats par catégorie grammaticale	106
6.2	Résultats obtenus aux tests sémantiques	107
6.3	Termes classés par les juges humains : objectif (O), subjectif (S) ou ambigu (A)	108
6.4	Termes ajoutés au lexique de l'évaluation	110
6.5	Évaluation hors-contexte de l'enrichissement lexical par tests sémantiques	110
6.6	Tableau de contingence de l'accord observé	111
6.7	Précision obtenue avec le partie enrichie du lexique	112
6.8	Évaluation quantitative de l'enrichissement du lexique de l'évaluation.	112
6.9	Nombre d'entrées du Wiktionnaire selon leur affixe	116
6.10	Nombre de candidats évaluatifs du Wiktionnaire catégorisés manuellement	116
6.11	Évaluation hors-contexte de l'enrichissement lexical par hypothèses morphologiques	117
6.12	Nombre de mots négatifs et exemples appris à partir d'hypothèses morphologiques	118

6.13	Tableau de contingence de l'accord observé	119
6.14	Précision obtenue avec la seconde partie enrichie du lexique	120
6.15	Évaluation quantitative de l'enrichissement du lexique de l'évaluation	120
7.1	Exemple de structuration des évaluations et des cibles d'un corpus	127
7.2	Nombre d'objets principaux par billet	131
7.3	Nombre d'objets secondaires par billet	131
7.4	Étude de la position de la cible par rapport à son évaluation	131
7.5	Caractéristiques utilisées pour décrire chaque couple évaluation-cible	133
7.6	Chemins de dépendances syntaxiques les plus fréquents entre une évaluation et une cible	134
7.7	Résultats obtenus par les quatre méthodes d'identification de la cible d'une évaluation . .	138

Introduction

1.1 L'essor du « web social » : un contexte applicatif porteur

Avec l'avènement du web 2.0, le web est devenu « social ». Des outils de publication en ligne simples d'utilisation (forums, blogs et micro-blogs¹) ont permis à chacun de créer sa propre tribune sur internet. Ces lieux d'expressions sont ouverts aux commentaires des membres d'une communauté ou des lecteurs de passage. Ainsi, le modèle traditionnel du média, basé sur la diffusion descendante de masse, a été supplanté par un modèle décentralisé, dans lequel émetteur et récepteur se confondent, partagent leurs centres d'intérêts, argumentent leurs points de vue, affirment leur personnalité, voire cherchent volontairement à médiatiser leurs opinions pour influencer sur leurs communautés.

Un des artefacts les plus connus de cette sphère basée sur l'opinion est le « buzz », terme désignant ces moments courts où l'opinion concentre son attention sur un événement particulier. La facilité et la rapidité de communication permises par le web 2.0 donne alors à l'événement une exposition large que seuls les médias traditionnels pouvaient auparavant provoquer. La figure 1.1 illustre ainsi le bourdonnement autour du footballeur *Nicolas Anelka* en juin 2010 mesuré par l'outil *Wikio Trend*². Si ce type d'outils permet un certain aperçu quantitatif des opinions exprimées, en revanche, ils ne permettent pas de les caractériser qualitativement.

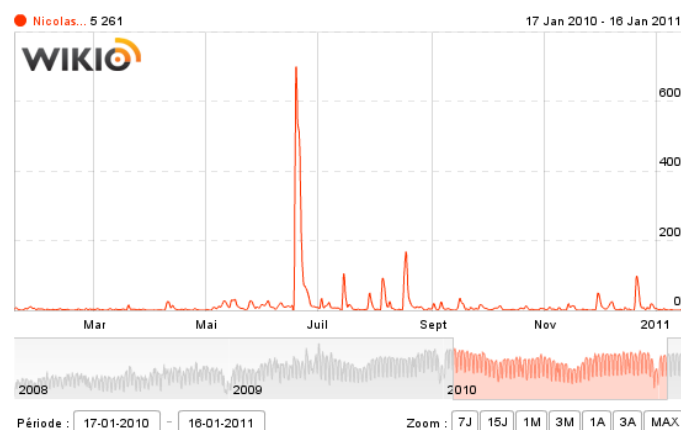


FIGURE 1.1 : Nombre d'articles (presses, blogs) autour de *Nicolas Anelka* en 2010

1. Exemples : <http://twitter.com/> et <http://www.facebook.com/>
2. <http://labs.wikio.net/fr/trends/trendscategory/footballeurs>

Devant l'ampleur du phénomène, pour certains demandeurs les opinions exprimées sur internet sont devenues des données stratégiques qu'il faut pouvoir mesurer en temps réel et analyser avec suffisamment d'avance et de précision pour pouvoir réagir. Les industriels ayant un fort besoin relationnel avec leurs clients (*SNCF, EDF, etc.*), les nouvelles technologies sensibles aux modes (*téléphone, ordinateur portable, lecteur MP3, etc.*), la politique, les médias et les sociologues en sont les principaux utilisateurs finaux. Profitant de cette vague, les agences de marketing proposent désormais des services de veille d'*e-reputation* basées sur l'observation manuelle des opinions émises sur internet autour d'une marque, d'un produit, d'une société ou d'une personne. La question centrale « *Quelle est l'opinion du public à propos d'un objet donné* » se spécialise en plusieurs questions :

- Quels sont les aspects de l'objet qui sont particulièrement évalués par le public ? (*le prix d'un produit, son utilité, son apparence, sa durée de vie, etc.*)
- Quels sont les termes employés ? Sont-ils positifs ou négatifs ? Quelles valeurs mettent-ils en jeu : *affectives, esthétiques, pragmatiques, intellectuelles, morales* ?
- Quelle est l'évolution de l'opinion au fil du temps ?

L'exemple (1) illustre le texte d'un blog à propos d'une publicité de *McDonald*, jugée mensongère. Les passages en rouge révèlent un discours négatif, ceux en bleu, un discours positif ou ambigu (ironique). Une lecture de ce texte permet surtout de faire ressortir des opinions négatives, liées au champ de la morale, au sujet de la publicité de *McDonald* en pointant une manipulation et des mensonges (*matraquage, mensongère, etc.*). Les quelques marques positives ou ironiques portent sur des détails liés à *McDonald* (*logo, frites*). Néanmoins, dans le reste du billet, le locuteur montre qu'il existe un réel attachement affectif à la marque, représentée dans l'extrait par quelques indices (*Très cher Ronald, je suis déçue, etc.*). Bien qu'elle possède un lien affectif fort avec ses clients, *McDonald* a ainsi tout intérêt à identifier que ses publicités et ses produits peuvent être mal perçus si elle joue la carte de l'éthique et de la morale.

(1) **Quand Mc Donald's passe son Happy Meal au Bio**

Texte extrait d'un blog - 4 novembre 2010

La première fois que j'ai vu la pub je me suis dit qu'il fallait s'y attendre, Mc Donald's n'a pas supporté que Quick fasse son burger bio et a répliqué rapidement en ciblant les enfants (ceux qui veulent au final le plus aller chez Ronald). Sauf, que **matraquage oblige**, la pub est passée deux fois en une série ... et là je me rends compte qu'en fait **j'ai mal compris** l'happy meal n'est pas bio, il n'y a que deux produits à l'intérieur qui sont issus de l'agriculture biologique ... **publicité mensongère**... On dira plutôt que le service marketing a bien insisté sur uniquement deux produits (les laitages) sur dirons nous 4 ... alors que le burger lui bon ben il **n'est pas bio du tout ! !** [...] A ce stade là, sincèrement **il vaut mieux** prendre les frites [...].

Très cher Ronald, **je suis vraiment très déçue**, tu as **un joli** logo vert mais **le chemin est encore long** pour que tu sois complètement green.

Pour faire face à la quantité de données mises en ligne chaque jour sur les médias « officiels » (la presse en ligne), sur les blogs ou sur les forums, les systèmes d'analyses automatiques représentent une alternative ambitieuse pour répondre aux attentes applicatives. Néanmoins, ces systèmes se heurtent à des problématiques relevant du traitement automatique des langues. Pour le TAL, ces nouveaux défis sont une opportunité à l'origine du domaine de la « fouille d'opinions »³.

3. Terme anglais : *Opinion Mining*.

1.2 La fouille d'opinions : un domaine émergent

Le début des années 2000 marque le début de l'éclosion de la fouille d'opinions et des opportunités scientifiques et applicatives que ce nouveau domaine du TAL soulève. Plusieurs facteurs semblent expliquer cette ruée vers l'opinion :

- l'essor d'internet donnant des quantités de données accessibles sur lesquelles entraîner et tester les systèmes. Le développement de sites commerciaux de critiques en ligne, des forums, puis des blogs et enfin des micro-blogs sont autant de vagues perceptibles ayant eu des répercussions sur la dynamique du domaine ;
- les enjeux applicatifs commerciaux, politiques et sociologiques que l'on peut imaginer considérables, voire dont on peut questionner les limites aussi bien théoriques, techniques qu'éthiques ;
- et le développement des méthodes d'apprentissage automatique en traitement automatique des langues et en recherche d'information.

Génèse et choix terminologiques Les expressions **opinion mining** ou encore *sentiment analysis* (traduit de façon erronée en français par *analyse des sentiments*) pour nommer ce domaine sont apparues parallèlement au début des années 2000. Dave, Lawrence et Pennock (2003) décrivent ainsi l'idée d'un outil de *fouille d'opinions* capable de synthétiser, pour un objet cible donné, une liste de caractéristiques évaluatives exprimées dans un ensemble de documents. L'idée était alors essentiellement de pouvoir guider le processus décisionnel d'un utilisateur souhaitant acheter un produit, en lui proposant un résumé des avis d'autres consommateurs sur un produit et un ensemble de ses caractéristiques (Tableau 1.1 - (Kobayashi, Inui et Matsumoto, 2007)). Cependant, plus récemment, ce terme a inspiré de nombreux dérivés pour tenter de différencier les différents supports textuels ou les différentes granularités d'analyses automatiques possibles liées à l'opinion dans les textes. On parlera alors selon les cas de *review mining* (Zhuang, Jing et Zhu, 2006), de *blog mining* (Attardi et Simi, 2006), d'*appraisal extraction* (Bloom et Argamon, 2010), etc.

objet évalué	objet associé	évaluation	polarité
Restaurant A	curry	<i>délicieux</i>	positif
Restaurant A	atmosphère	<i>relaxante</i>	positif
Restaurant B	serveur	<i>impoli</i>	négatif

TABLE 1.1 : Analyse fournie en sortie d'un système de fouille d'opinions

Turney (2002) et Pang, Lee et Vaithyanathan (2002) emploient pour la première fois le terme *sentiment analysis* dans le cadre de leurs travaux en catégorisation de critiques de films. De nombreux articles mentionnant le concept de *sentiment analysis* se sont intéressés à la catégorisation de critiques selon leur polarité positive ou négative. Cela vaut à cette expression d'être souvent associée à la tâche de catégorisation globale de textes d'opinions, au point d'avoir été francisée en *analyse des sentiments* (Maurel, Curtoni et Dini, 2008) bien que cette tâche ne soit pas spécifique aux sentiments.

Nous considérons que ces différentes variantes de termes en concurrence dénotent tous le même champ d'étude : celui de l'**analyse automatique de la subjectivité**. Nous justifions l'emploi du terme de *subjectivité* et d'*évaluation* plutôt que ceux d'*opinion*, de *jugement*, d'*émotion*, etc., souvent employés en TAL, car la subjectivité et l'évaluation sont les phénomènes linguistiques qui englobent chacun de ces termes alternatifs (voir chapitre 3). Ce foisonnement terminologique témoigne de la diversité des problématiques liées aux traitements de données d'opinions au détriment parfois d'une bonne lisibilité terminologique.

Diminuer la granularité d'analyse Dans les premières années d'existence de la fouille d'opinions, les travaux se sont intéressés essentiellement à catégoriser un document textuel selon :

- son aspect globalement subjectif ou objectif ;
- sa polarité axiologique globalement positive, négative ou intermédiaire.

Les méthodes proposées paraissent trop réductrices par rapport aux enjeux applicatifs en termes de structuration d'informations. Surtout, elles ne permettent pas de rendre compte des informations véhiculées par les textes dont le contenu est hétérogène. Par exemple, dans le texte (1) sur *MacDonald*, une catégorisation au niveau du document réduit l'information à une seule caractéristique, alors que :

- plusieurs objets sont évalués : *la publicité, les frites, le logo, MacDonald* ;
- plusieurs modalités d'évaluations sont employées, activant plusieurs domaines de valeurs : l'affect (*Très cher, je suis vraiment très déçue*), l'esthétique (*joli logo*), la morale (*matraquage, mensongère*), la conviction (*sincèrement, vraiment, !!*) ;
- plusieurs polarités axiologiques sont utilisées (*positif, négatif*), impliquant même un phénomène d'*ironie*.

Des considérations plus récentes en fouille d'opinions (Liu, 2010) alimentent l'idée que le document est un grain d'analyse trop large pour pouvoir répondre à ces attentes. Dès lors, les analyses à granularité fine font l'objet d'un nouvel axe de recherche très récent en fouille d'opinions. Dans cet axe, les travaux s'intéressent à caractériser les passages subjectifs d'un texte et à relier ces passages à leur auteur (*la source*) et au sujet qu'ils évaluent (*la cible*). Néanmoins, par rapport à la catégorisation au niveau du document, cet axe est beaucoup moins doté en ressources et en méthodes informatiques. En particulier, pour le traitement du français, il n'existe à notre connaissance aucun travail de référence dans cet axe. De plus, l'analyse à granularité fine de la subjectivité nécessite d'améliorer la qualité et la couverture des lexiques existants. Il faut donc pouvoir ré-explore des méthodes pour développer des lexiques et évaluer précisément leur qualité et leur couverture. Nous détaillons ces différents axes de problématiques, nous présentons les méthodes existantes et synthétisons notre positionnement dans le **chapitre 2**.

1.3 Développer des méthodes et des ressources pour une analyse à granularité fine

Nous nous positionnons dans l'axe de l'analyse à granularité fine des opinions dans un cadre multithématique et pour le traitement du français en particulier. Notre démarche nous amène à recadrer le terme d'opinion dans les notions linguistiques plus larges de subjectivité et d'évaluation. Nous adoptons un modèle pour décrire la subjectivité et le langage de l'évaluation en se basant sur les théories linguistiques de Charaudeau (1992) et Galatanu (2000) (**chapitre 3**). Dans l'axe choisi, l'apport de notre travail joue sur deux types d'aspects : les ressources et les méthodes.

Ressources pour l'analyse à granularité fine de la subjectivité Les corpus annotés et les lexiques sont deux types de ressources critiques dans le domaine de la fouille d'opinions. Ils sont à la base des protocoles expérimentaux des problématiques scientifiques liées à l'opinion. Néanmoins, ces ressources sont particulièrement rares ou absentes dans le domaine, en particulier pour le français. Nous avons développé un corpus annoté, le corpus *Blogoscopie*, et un lexique de mots évaluatifs, le lexique de l'évaluation. Ces ressources sont structurées en tenant compte du modèle linguistique de l'évaluation cité précédemment. Elles constituent le premier type de contribution de cette thèse. Nous les présentons dans le **chapitre 4**.

- Le *corpus Blogoscopie* est un corpus de blogs, dont les billets et commentaires ont été annotés manuellement. L'annotation porte sur deux catégories d'éléments : les objets signifiés dans les

billets et les commentaires, et les évaluations émises par l'auteur du billet ou du commentaire à propos de certains de ces objets. Certaines relations sémantiques entre objets ou entre un objet et une évaluation sont annotées. Il s'agit du premier corpus francophone où ces éléments sont annotés.

- *Le lexique de l'évaluation* structure des informations lexicales et sémantiques autour des valeurs prises par des termes évaluatifs lors de leurs apparitions en contexte. Nous avons évalué ce lexique qualitativement et quantitativement en contexte sur la tâche de détection de passages d'évaluations (**chapitre 5**). Initialement construit manuellement, ce lexique est constitué, après enrichissement automatique, de 3 974 entrées. Par sa taille, il est à l'heure actuelle le lexique de mots évaluatifs le plus important pour le français (**chapitre 6**).

Ces deux ressources sont distribuées sous licence *LGPLLR* afin de faciliter les recherches futures. Elles sont disponibles à l'adresse suivante : <http://www.lina.univ-nantes.fr/Ressources.html>

Méthodes pour l'analyse à granularité fine de la subjectivité Pour répondre à différentes problématiques liées à l'analyse de la subjectivité à granularité fine, nous proposons plusieurs méthodes. Celles-ci constituent le second type de contribution de cette thèse.

- notre première méthode a pour objectif de délimiter les passages évaluatifs de documents et de catégoriser leur modalité sémantique, leur configuration d'énonciation et leur polarité axiologique. Cette méthode s'appuie sur une extraction préalable de structures linguistiques utilisées pour exprimer une évaluation. Ces structures combinent des informations lexicales, grammaticales et sémantiques. L'évaluation de cette méthode par structures sur un corpus de test montre notamment son intérêt pour redéfinir la polarité axiologique d'une entrée lexicale en tenant compte de son contexte (tournure de négation, etc.). Cette méthode a également été testée lors de la campagne d'évaluation DEFT'09 où elle a obtenu les meilleurs résultats. Nous présentons nos travaux sur ces aspects dans le **chapitre 5** ;
- dans le **chapitre 6**, nous proposons ensuite deux méthodes pour enrichir automatiquement le lexique de l'évaluation. Ces méthodes s'appuient des hypothèses sémantiques et des hypothèses morphologiques qui contraignent le sens de mots ou d'expressions inconnus. À partir des usages linguistiques sur le web, d'un dictionnaire et d'une méthode d'apprentissage supervisé, nous inférons le degré de subjectivité ou la polarité de ces mots ou expressions pour compléter le lexique de l'évaluation. Notre protocole d'évaluation montre que ces deux méthodes permettent d'augmenter la taille du lexique de l'évaluation et détecter un nombre de nouveaux passages évaluatifs significatif en préservant une bonne précision. Les types de mots évaluatifs appris par ces méthodes sont rarement présents dans les ressources du domaine et sont de notre point de vue des indices intéressants (stéréotypes culturels, néologismes, mots rares, expressions) ;
- enfin, dans le **chapitre 7**, nous proposons une étude comparative de quatre méthodes (dont deux baselines) pour la problématique d'identification de la cible d'un passage d'opinion. Nous comparons des méthodes heuristiques, syntaxiques ou combinant différents types d'indices pour résoudre des cas non traités par les méthodes existantes. Notre méthode n'est pas limitée au cadre de la phrase et traite tous les types d'évaluations dans un contexte multithématique. Elle a été entraînée à partir du corpus Blogoscopie et évaluée sur une partie réservée aux tests.

Le bilan et les perspectives de cette thèse font l'objet de la conclusion de ce manuscrit.

Vers une analyse à granularité fine

2.1	Analyse de l'opinion au grain du document	10
2.1.1	Un problème de catégorisation de documents	10
2.1.1.1	<i>Identifier la polarité d'une critique de produit</i>	10
2.1.1.2	<i>Identifier l'opinion politique d'un texte</i>	15
2.1.1.3	<i>Identifier la subjectivité d'articles de presse</i>	16
2.1.2	Limites des approches au grain document	17
2.2	Analyse à granularité fine de l'opinion	18
2.2.1	Une combinaison de problèmes hétérogènes	18
2.2.1.1	<i>Catégoriser les passages d'opinions</i>	20
2.2.1.2	<i>Identifier la cible des passages d'opinions</i>	22
2.2.1.3	<i>Identifier la source des passages d'opinions</i>	23
2.2.2	Synthèse des réflexions	23
2.3	Extraction des mots d'opinions d'une langue	24
2.3.1	Trois types d'approches	24
2.3.1.1	<i>Approche manuelle</i>	25
2.3.1.2	<i>Approche par dictionnaire</i>	26
2.3.1.3	<i>Approche par corpus</i>	27
2.3.2	Des résultats difficilement évaluable	29
2.4	Conclusion	30
2.4.1	Synthèse du domaine de la fouille d'opinions	30
2.4.2	Positionnement de nos travaux	31

Ce premier chapitre a pour objectif de présenter notre position par rapport à l'évolution des problématiques en fouille d'opinions. Pour adopter les usages de la communauté du TAL jusqu'ici majoritaires et faciliter la compréhension, nous utilisons dans ce chapitre le terme d'opinion. Nous définissons ultérieurement la notion linguistique de subjectivité (voir chapitre 3) et précisons l'emploi du terme évaluation qui englobe l'opinion en tant que phénomène complexe à la fois sémantique, pragmatique et énonciatif. La figure 2.1 représente les différentes problématiques traitées dans la littérature du domaine. Historiquement, les travaux en fouille d'opinions se scindent en deux axes :

- l'**analyse au grain document** – axe représentatif des travaux pionniers du domaine et majoritaires dans la première moitié des années 2000 (Pang et collab., 2002 ; Dave et collab., 2003 ; Mishne et Glance, 2006) ;
- l'**analyse à granularité fine** – axe représentatif de problématiques plus récentes qui se sont développées parallèlement à nos travaux (Wilson, 2008 ; Ruppenhofer, Somasundaran et Wiebe, 2008 ; Kessler et Nicolov, 2009).

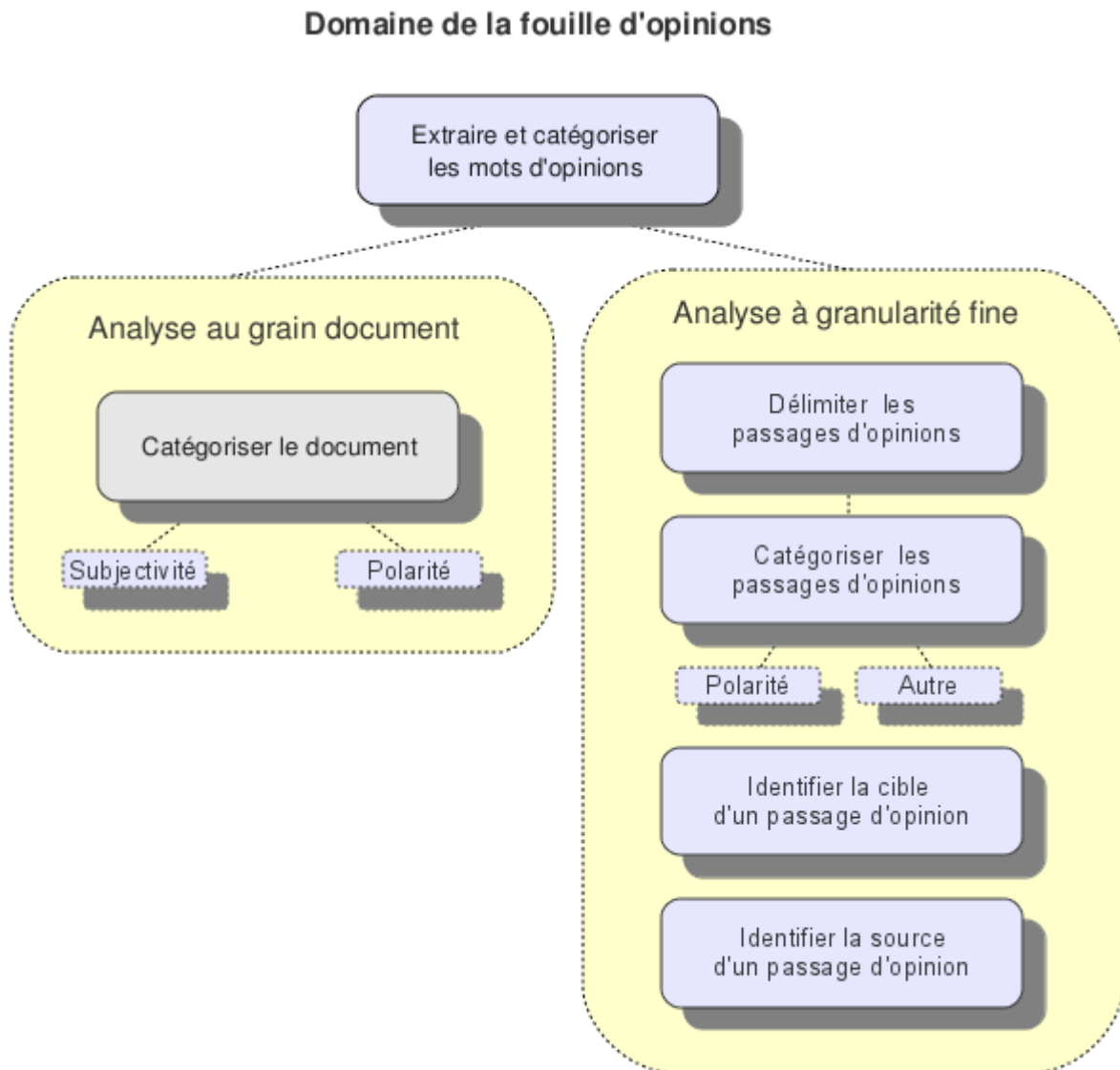


FIGURE 2.1 : Problématiques de la fouille d'opinions

Analyse de l'opinion au grain document Les premiers travaux se sont intéressés à des méthodes de catégorisation au grain du document. Ils permettent de répondre aux questions suivantes :

- *le document est-il globalement subjectif ou objectif ?*
- *la polarité du document est-elle globalement positive, négative ou intermédiaire ?*

Ces travaux s'inspirent en grande partie des méthodes existantes en catégorisation de documents. D'un point de vue applicatif, ces méthodes ne sont pertinentes que si le document parle d'un seul objet principal et s'il n'exprime une opinion que sur cet objet. Pour le traitement du français, cet axe a fait l'objet du DÉfi Fouille de Textes 2007 où un panel de méthodes ont été évaluées et comparées (Grouin, Berthelin, Ayari, Heitz, Hurault-Plantet, Jardino, Khalis et Lastes, 2007). Pour l'anglais, des campagnes d'évaluations ont proposé des tâches analogues (Ounis, de Rijke, Macdonald, Mishne et Soboroff, 2006 ; Macdonald, Ounis et Soboroff, 2007), les méthodes utilisées sont relativement similaires quelque soit la langue support.

Analyse à granularité fine de l'opinion Plus récemment, pour mieux tenir compte des spécificités des besoins applicatifs liés aux données d'opinions et analyser tous types de documents, un deuxième axe s'est développé pour analyser l'opinion à un grain plus fin. Cet axe regroupe plusieurs problématiques visant à répondre aux questions suivantes :

- *Quels sont les passages qui expriment une opinion dans un document ?*
- *Quelles sont les caractéristiques sémantiques de ces passages ?*
- *Quels sont les objets concernés par ces passages d'opinions (cible) ?*
- *Quels sont les auteurs de ces passages d'opinions (source) ?*

Par rapport à l'approche au grain document, ces quatre questions permettent de répondre aux besoins applicatifs en synthétisant des données d'opinions hétérogènes issues d'un même document. En revanche, ces problématiques sont beaucoup moins abordées dans la littérature. Elles font l'objet de quelques travaux récents pour l'anglais, mais à notre connaissance cet axe n'a pas été exploré pour le français. Une tâche du DÉfi Fouilles de Textes 2009 était proposée dans cet axe, mais seules deux équipes, dont la nôtre, ont proposé des méthodes (Grouin, Arnulphy, Berthelin, Ayari, García-Fernandez, Grappy, Hurault-Plantet, Paroubek, Robba et Zweigenbaum, 2009).

Extraire et catégoriser les mots d'opinions d'une langue Les deux axes évoqués précédemment dépendent d'une première problématique qui consiste à extraire les *mots d'opinions*¹ d'une langue. Dans les deux cas, avoir comme ressource une liste de mots d'opinions est un pré-requis pour pouvoir analyser des données d'opinions au grain document ou à un grain fin. En revanche, nous observons que les exigences pour cette problématique ne sont pas les mêmes selon le grain d'analyse. Les analyses au grain document tolèrent beaucoup mieux que les listes de mots d'opinions extraites soit partiellement erronées (les erreurs étant globalement compensées par les mots d'opinions corrects au niveau du document). Pour des analyses à granularité fine, les erreurs ne sont pas « cachées », et les ressources existantes provoquent trop de bruit et de silence. Il est donc nécessaire de redéfinir des méthodes pour améliorer l'extraction et la catégorisation des mots d'opinions d'une langue pour obtenir des listes plus précises et avec une meilleure couverture.

Dans ce chapitre, nous articulons notre réflexion autour de ces trois axes. Nous présentons :

- les objectifs des méthodes d'analyses au grain document (§. 2.1) et montrons en quoi ceux-ci ne sont pas adaptés pour répondre aux défis actuels ;
- les méthodes existantes dans l'axe des analyses à granularité fine (§. 2.2). Celles-ci ont besoin d'être développées et adaptées au français pour lequel il n'existe aucune méthode ;

1. Le terme *opinion words* est utilisé dans les travaux anglophones.

- la problématique d'extraction et de catégorisation des mots d'opinions d'une langue (§. 2.3). Bien que cette problématique soit riche en méthodes, celle-ci a besoin d'être ré-explorée pour proposer de nouvelles méthodes permettant d'extraire des listes de mots avec une meilleure précision et un meilleur rappel.

2.1 Analyse de l'opinion au grain du document

La catégorisation de l'opinion au niveau document² est l'axe le plus souvent traité dans la bibliométrie du domaine entre 2002 et 2007. Cet axe répond à certains types de besoins applicatifs en utilisant des méthodes issues des travaux en catégorisation de documents (§. 2.1.1). Néanmoins cet axe ne semble pas adapté pour traiter tous les aspects complexes de l'opinion dans langue et répondre aux attentes de nouveaux défis applicatifs (§. 2.1.2).

2.1.1 Un problème de catégorisation de documents

Les travaux existants en catégorisation de l'opinion au niveau document peuvent se concevoir comme un problème de classification de documents. Étant donné un corpus de documents dont on ignore la classe, l'objectif est d'affecter correctement chaque document à sa classe correspondante (Fig. 2.2). En fouille d'opinions, les travaux ont décliné ce problème selon différents axes de classification. Il s'agit de catégoriser soit :

- la polarité *positive* ou *négative*, voire *intermédiaire*, du document : 2 ou 3 classes ;
- l'*opinion politique* globalement véhiculée par un document : de 2 à 4 classes ;
- la *subjectivité* ou l'*objectivité* globale du document : 2 classes.

Ces travaux sont généralement indissociables de leur cadre applicatif. Dans cet axe, ces applications se limitent principalement à trois cas :

- catégoriser des critiques de produits selon la polarité des opinions ;
- catégoriser des articles politiques ou des textes de débats parlementaires selon l'étiquette politique de leur auteur ;
- catégoriser des articles de presse selon leur subjectivité ou leur objectivité globale.

Les méthodes proposées sont en partie liées aux caractéristiques de ces applications. Nous présentons ci-dessous les méthodes existantes réparties selon leur application.

2.1.1.1 Identifier la polarité d'une critique de produit

L'éclosion des nombreux sites web orientés sur le partage d'*avis de consommateurs* ou de *critiques de spectateurs* (Amazon, Fnac, Allocine, etc.) a nourri l'imagination d'un système de fouille d'opinions capable d'aggréger l'avis de multiples utilisateurs (Dave et collab., 2003) pour les présenter de manière synthétique. Une majorité d'utilisateurs de ces sites reconnaissent en effet que les avis d'autres consommateurs influencent significativement leurs décisions d'achats (comScore/the Kelsey group, 2007). Par conséquent, connaître le taux d'affection global d'un produit est un enjeu économique pour les entreprises et les distributeurs. C'est dans cette problématique applicative que se sont situés la très grande majorité des travaux en fouille d'opinions depuis 2002. L'objectif consiste à catégoriser les textes de critiques issus de ces sites selon leur polarité. Ces textes ont un sujet principal unique : *un film, une voiture, un livre, un album musical, un jeu vidéo, un spectacle, un appareil photo, un lecteur MP3*, etc.

2. Terme anglais : *Document-level sentiment classification*.

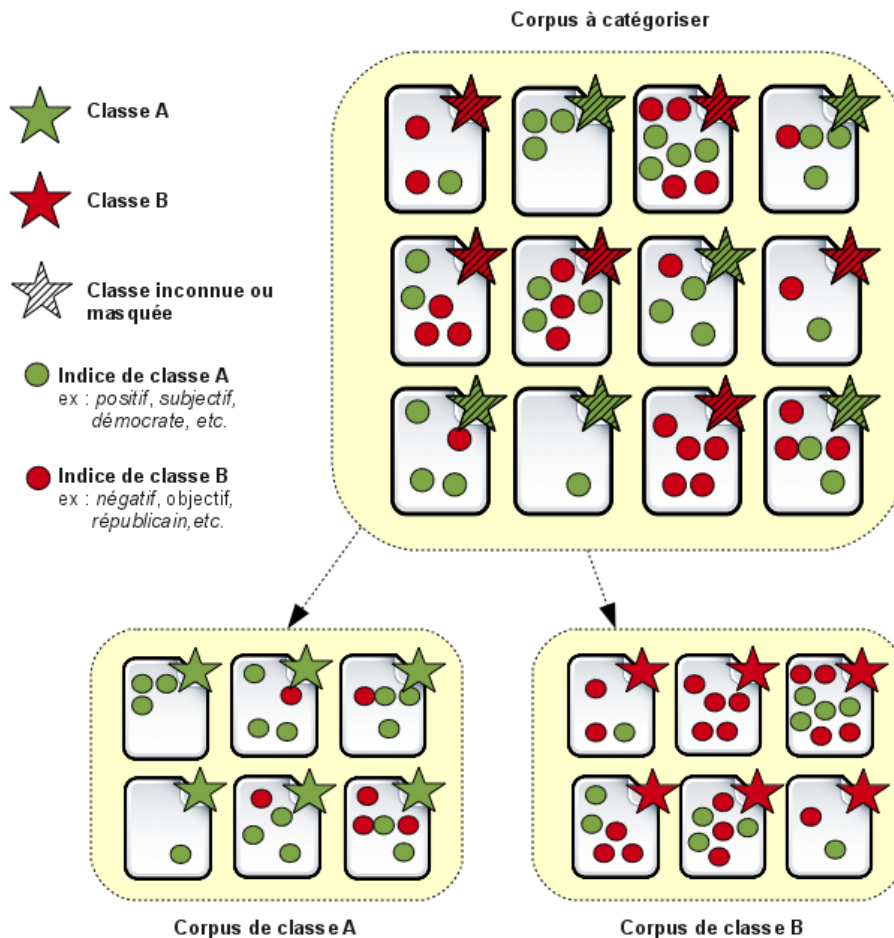


FIGURE 2.2 : Analyse de l'opinion au grain document

Le DÉfi Fouille de Textes 2007 (DEFT'07)³ a permis de confronter des systèmes informatiques sur des corpus francophones de critiques de produits de ce type. Les données utilisées sont issues de sites de critiques en ligne⁴ et la classe des documents est connue à l'avance grâce à la démarche réalisée par chaque auteur de critique pour attribuer une note globale à leur texte (par exemple, une à cinq étoiles). La seule adaptation effectuée consiste à réduire l'échelle de notes des sites web (par exemple, une note d'une à deux étoiles devient une critique négative alors qu'une note de 4 à 5 étoiles correspond à une critique positive). Dans le cadre de DEFT'07, les méthodes proposées par Torres-Moreno, El-Bèze, Béchet et Camelin (2007) et Vernier, Mathet, Rioult, Charnois, Ferrari et Legallois (2007) obtiennent les meilleurs résultats sur ces corpus. La mesure de F-Score varie entre 0.60 et 0.78 selon les corpus.

D'autres campagnes d'évaluations anglophones (TREC) et d'autres travaux ont adopté ce protocole, ou un protocole voisin, pour proposer des méthodes (Pang et collab., 2002 ; Aue et Gamon, 2005 ; Zhuang et collab., 2006 ; Das et Chen, 2007). Les méthodes sont en générales analogues aux méthodes de classification supervisée de documents et ce sont surtout les descripteurs linguistiques utilisés qui font varier la qualité des résultats.

3. <http://deft07.limsi.fr/>

4. <http://www.avoir-alire.com> et <http://www.jeuxvideos.com>

Classification supervisée Pang et collab. (2002) sont les premiers à avoir expérimenté des algorithmes d'apprentissages supervisés (naïve bayes, maximum entropie, séparateur à vaste marge (SVM)) pour classer des documents selon leur polarité. De nombreux travaux dérivés ont ensuite suivi cette idée en adaptant et en comparant des méthodes de classification sur différents types de corpus. Pour le français, Plantié, Dray et Roche (2007) ont expérimenté quatre modèles de classification : naïve bayes multinomial, réseaux de neurones RBF, SVM et arbre de décision. Lors de DEFT'07, Torres-Moreno et collab. (2007) montrent qu'une combinaison de plusieurs systèmes de classification départagés par une méthode de vote permet d'obtenir des résultats encore plus performants. Prises indépendamment, les méthodes utilisant les SVM (Boser, Guyon et Vapnik) sont les plus souvent citées dans les meilleurs résultats obtenus. L'idée des méthodes SVM (voir Fig. 2.3) est de :

- représenter un document par un vecteur de descripteurs. La nature de ces descripteurs est variable (cf. paragraphe suivant) ;
- utiliser un corpus d'apprentissage où les classes des documents sont connues afin de calculer un hyperplan qui maximise la marge séparant les vecteurs de descripteurs des deux classes de documents ;
- appliquer cette fonction pour classer les documents inconnus d'un corpus de test.

Les conclusions des travaux du domaine tendent à montrer que ce sont surtout les choix des descripteurs⁵ qui ont un rôle majeur dans les résultats obtenus. Le choix des descripteurs les plus efficaces est variable selon le corpus considéré. Les descripteurs les plus fréquemment utilisés sont :

- les **mots** sous forme lemmatisée. Torres-Moreno et collab. (2007) montrent néanmoins que la morphologie peut être discriminante sur certains corpus : *vous auriez dû, serait* ;
- les mots limités à certaines **catégories grammaticales**. Les documents sont souvent réduits à leur liste d'adjectifs (Mullen et Collier, 2004 ; Whitelaw, Garg et Argamon, 2005 ; Harb, Dray, Plantié, Poncelet, Roche et Trouset, 2008) de part la corrélation importante entre les adjectifs et les opinions ;
- les mots limités à certaines **catégories sémantiques**. La plupart des méthodes utilisent un vocabulaire de mots d'opinions pour réduire le document à une liste de mots d'opinions polarisés (Mullen et Collier, 2004 ; Torres-Moreno et collab., 2007 ; Maurel, Curtoni et Dini, 2007 ; Génereux et Santini, 2007 ; Vernier et collab., 2007 ; Choi et Cardie, 2009 ; Denecke, 2009 ; Dang, Zhang et Chen, 2010) ;
- l'utilisation de **n-grammes** à la place des mots est une alternative utilisée notamment par Pang et collab. (2002), Dave et collab. (2003), Matsumoto, Takamura et Okumura (2005) et Vernier et collab. (2007). Les n-grammes permettent d'extraire certaines expressions spécifiques à un corpus qui peuvent être discriminantes sans néanmoins contenir de mots d'opinion : *courez l'acheter* dans un corpus de critiques de livre, *on s'endort* dans un corpus de critiques de spectacle, etc. Ces n-grammes sont également utilisés pour repérer les marqueurs de négation qui inverse la polarité : *ai pas aimé, pas un bon roman* ;
- enfin, certaines **tournures de phrases**, interrogatives ou exclamatives peuvent également être des descripteurs pour discriminer certaines classes de document d'opinion selon les corpus (Torres-Moreno et collab., 2007).

Ces descripteurs peuvent être pondérés par plusieurs aspects :

- leur simple **présence** dans le document ;
- une mesure de leur **importance** dans le document ou dans la classe (fréquence, TF.IDF, test de χ^2 , information mutuelle, etc.). Pang et collab. (2002) montrent toutefois que considérer

5. Terme anglais : *features*.

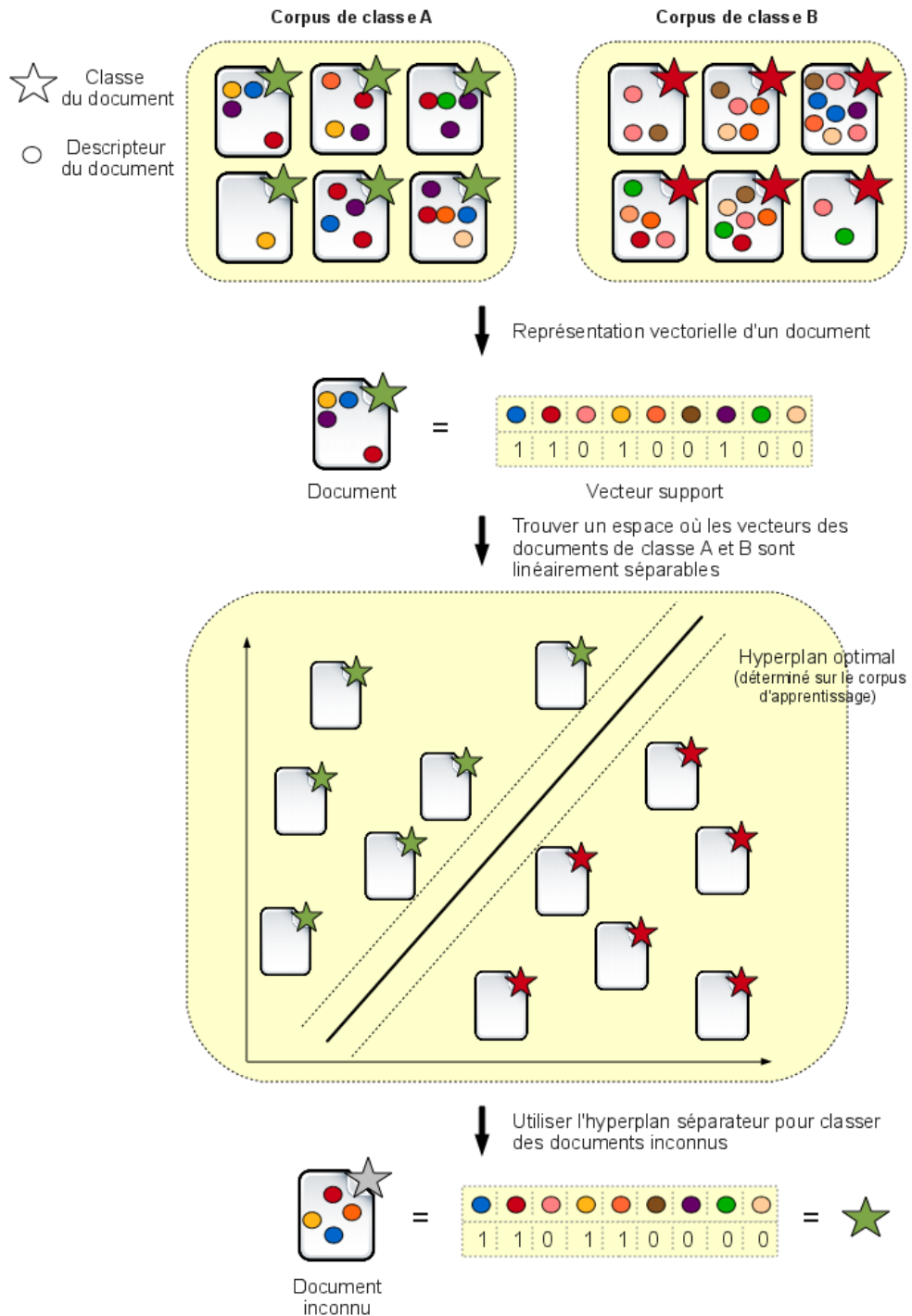


FIGURE 2.3 : Processus de classification avec séparateur à vaste marge (SVM)

la simple présence ou absence d'un mot plutôt que sa fréquence est plus efficace sur certains corpus, en particulier des corpus de critiques de films ;

- une prise en compte de leur **position** dans le document. La position d'un terme dans le document (au début, au milieu ou à la fin d'un texte) peut potentiellement avoir un effet sur la façon dont ce terme affecte la polarité globale du texte (Kim et Hovy, 2006a ; Vernier et collab., 2007). Ce dernier critère est néanmoins rarement considéré dans les approches du domaine.

Les méthodes dites supervisées nécessitent toujours l'existence d'un corpus d'apprentissage pré-étiqueté, ce qui limite souvent la portabilité de ces méthodes aux corpus de critiques. Pour contourner le besoin d'un corpus d'apprentissage, Turney (2002) propose une méthode non supervisée de catégorisation de la polarité d'un document.

Classification non supervisée La méthode proposée par Turney (2002) détermine de façon non supervisée la polarité des mots d'un document pour pouvoir le classer. Il s'appuie pour cela sur les adverbes et les adjectifs présents dans le document. Il considère ces deux catégories grammaticales comme étant davantage susceptibles de contenir une opinion. Néanmoins, pour pallier le problème d'ambiguïté des mots simples en contexte, il n'extrait du texte qu'une liste de bigrammes, contenant au moins un adjectif ou un adverbe, respectant un des patrons grammaticaux fixés manuellement (Tab. 2.1). L'étape 1 de son algorithme consiste à extraire tous les bigrammes du document analysé (1).

- (1) Son comportement est particulièrement irascible.

2. Adverbe Adjectif

	mot 1	mot 2	mot 3 (non extrait)
1.	Adjectif	Nom	Tout type
2.	Adverbe	Adjectif	Tout type sauf Nom
3.	Adjectif	Adjectif	Tout type sauf Nom
4.	Nom	Adjectif	Tout type sauf Nom
5.	Adverbe	Verbe	Tout type

TABLE 2.1 : Bigrammes grammaticaux d'opinions – (Turney, 2002)

L'étape 2 permet d'estimer la polarité d'un bigramme extrait en utilisant la mesure du point d'information mutuelle⁶ (PMI) établie dans la théorie de l'information et en statistique (équation 2.1).

$$\text{PMI}(T_1, T_2) = \log_2 \frac{\text{Pr}(T_1 \wedge T_2)}{\text{Pr}(T_1)\text{Pr}(T_2)} \quad (2.1)$$

Dans cette équation $\text{Pr}(T_1 \wedge T_2)$ est la probabilité de trouver les termes T_1 et T_2 simultanément, alors que $\text{Pr}(T_1)\text{Pr}(T_2)$ est la probabilité de trouver ces deux termes ensemble de façon fortuite. Le ratio entre ces deux expressions mesure ainsi le degré de dépendance statistique entre les deux termes. La polarité d'un bigramme extrait du document est calculée en se basant sur la différence entre la mesure du PMI avec un mot grain positif (*excellent* pour Turney) et la mesure du PMI et un mot grain négatif (*poor* pour Turney) (équation 2.2). Les différentes probabilités de ces formules sont estimées en interrogeant un moteur de recherche et en collectant le nombre de réponses de chaque requête. Turney (2002) a utilisé le moteur de recherche AltaVista pour pouvoir inclure l'opérateur *NEAR* dans ces requêtes. Cet opérateur permettait, à l'époque, d'ajouter une contrainte de proximité

6. Terme anglais : *pointwise mutual information*.

(une co-occurrence dans une fenêtre de 10 mots) entre les deux mots recherchés dans un document. Dans l'équation 2.2, la fonction *hits* représente ainsi le nombre de réponses fournies par le moteur de recherche pour la requête passée en argument.

$$\text{Polarité}(\text{candidat}) = \log_2 \frac{\text{hits}(\text{candidat NEAR excellent}) \text{ hits}(\text{poor})}{\text{hits}(\text{candidat NEAR poor}) \text{ hits}(\text{excellent})} \quad (2.2)$$

Enfin, l'étape 3 permet de classer globalement un texte selon la polarité en calculant la moyenne des polarités de chaque candidat catégorisé à l'étape 2. Si le modèle d'algorithme de Turney a été précurseur dans le domaine de la fouille d'opinions, quelques autres méthodes non supervisées ont également été proposées (Dave et collab., 2003).

Turney s'est évalué sur des critiques d'automobiles, de banques, de films et de voyages. Il obtient un F-Score moyen de 74,39 %. Au regard des bons résultats obtenus, Turney formule néanmoins la réserve suivante : les films qui contiennent des scènes violentes décrites dans la critique ont tendance à être classés négatifs quand bien même la critique est bonne. Inversement, des mauvaises critiques peuvent contenir du vocabulaire positif pour décrire une scène, le jeu d'un acteur, etc. Par cette remarque, Turney soulève implicitement dès le début des travaux en fouille d'opinions le problème de l'identification de la cible évaluée par une opinion (*le film, une scène particulière, un acteur*, etc.). Même les textes de critiques contiennent en réalité plusieurs objets et toutes les opinions ne concernent pas nécessairement l'objet principal.

2.1.1.2 Identifier l'opinion politique d'un texte

L'opinion est un aspect central en politique et savoir déterminer l'opinion politique de l'auteur d'un texte est un enjeu applicatif au coeur de plusieurs travaux du domaine (Laver, Benoit et Garry, 2003 ; Mullen et Malouf, 2006 ; Martin et Vanberg, 2008). Pour classer un texte politique, les travaux existants utilisent généralement un calcul de similarité (*k* plus proches voisins, cosinus, etc.) entre une représentation vectorielle du document et un ensemble de documents dont la classe est connue. L'idée est de considérer qu'un document inconnu est de la même catégorie que son voisin connu le plus proche. Pour représenter les documents, ces travaux ont tous privilégié des indices indirectement liés à l'opinion. Ils s'appuient plutôt sur des caractéristiques textuelles propres au cadre applicatif :

- le nombre de citations d'une entité nommée (une personnalité de droite a tendance à citer une personnalité de gauche pour la critiquer et inversement) ;
- les thématiques abordées et le champ lexical.

Selon cette idée, citons les travaux de Adamic et Glance (2005). Au cours de la campagne des élections américaines 2005, Adamic et Glance (2005) ont constitué un corpus séparé en deux catégories : 20 blogs *libéraux* et 20 blogs *conservateurs* les plus influents. Un de leurs objectifs est de relier les blogs par un coefficient de similarité et les catégoriser en deux classes : blog *libéral* et blog *conservateur*. Ces 40 blogs ont publié 22 800 billets pendant la période observée et sont utilisés comme corpus d'entraînement. Les auteurs s'appuient sur trois caractéristiques textuelles pour représenter un document :

- les **liens vers d'autres blogs** politiques. Les auteurs font remarquer que les blogs conservateurs ont tendance à citer (sous forme de lien) un autre blog du même camp 40% plus souvent que les libéraux ;
- une pondération (mesure Tf.Idf) des **bigrammes de mots** les plus importants ;
- les **entités nommées** (noms de personnalités politiques) citées sur chacun des blogs. Les personnalités politiques citées par la droite sont essentiellement des adversaires politiques : *Dan Rather, Michael Moore, Yassar Arafat* et *Terry McAuliffe*. Inversement, les blogs de gauche citent davantage des personnalités de droite : *Donald Rumsefeld, Colin Powell* et *Tim Russert*.

Au cours du DÉfi Fouille de Textes 2009 (DEFT'09)⁷, une des tâches proposées consistait à traiter un corpus de débats parlementaires européens datant de 1999 à 2004 et *in fine* retrouver l'étiquette du parti politique auteur d'un texte parmi quatre catégories : *ELDR*, *PPE-DE*, *PSE*, *Verts*. Forest, Bélange, Létourneau et van Hoeydonc (2009) sont les seuls à avoir proposé une méthode. Celle-ci s'appuie sur un test de χ^2 pour identifier les mots les plus discriminants d'une classe donnée. Chaque document est représenté par un vecteur de fréquence de chacun de ces mots dans le document. Par exemple, pour les *Verts*, dans le corpus d'entraînement les mots les plus discriminants sont les suivants : *écosse*, *galle*, *basque*, *nucléaire*, *flandre*, *catalogne*, *environnement*, *régionale*, *durable*. Avec la méthode des k plus proches voisins, un texte du corpus de test qui contient fréquemment ces mots a donc tendance à être classé « opinion Verte ».

Bien que ces travaux se revendiquent du domaine de la fouille d'opinions, ils s'intéressent finalement assez peu aux aspects linguistiques de l'opinion. Ils utilisent des caractéristiques propres aux corpus analysés et au domaine de la politique. Nous retrouvons la particularité d'utiliser des descripteurs indirectement ou non liés à l'opinion dans le cadre applicatif suivant.

2.1.1.3 Identifier la subjectivité d'articles de presse

Dans l'édition 2009, du DÉfi Fouilles de Textes (DEFT'09), une tâche consistait à catégoriser dans leur globalité des articles de presse issus de journaux (*Le Monde* pour le français, *The Financial Times* pour l'anglais) en deux catégories : *objectif* et *subjectif*. Cette même tâche a également fait l'objet de plusieurs autres travaux anglophones, toujours appliquée aux corpus d'article de presses (Riloff, Wiebe et Phillips, 2005 ; Lin, Wilson, Wiebe et Hauptmann, 2006 ; Kim, Zhang et Lee, 2010). À l'occasion de DEFT'09, Bestgen et Lories (2009) ont parfaitement illustré l'idée qu'au niveau du document, il est souvent plus efficace d'utiliser des caractéristiques propres aux corpus plutôt que de réellement s'intéresser aux caractéristiques de la subjectivité et de l'objectivité. Ils ont ainsi obtenu les meilleurs résultats pour le traitement du corpus français et du corpus anglais en n'ayant volontairement aucun intérêt spécifique pour la notion de subjectivité. Ils ont montré l'intérêt de prendre en compte la longueur des documents en plus de descripteurs plus classiques :

- les bigrammes et trigrammes de mots lemmatisés ;
- la suppression des mots fonctionnels.

Les descripteurs sont sélectionnés en testant plusieurs systèmes de pondération : la présence, la fréquence et l'analyse sémantique latente (LSA) (Deerwester, Dumais, Furnas, Landauer et Harshman, 1990). La pondération LSA obtient les meilleurs résultats d'optimisation lors de la phase d'entraînement (éq. 2.3).

$$f'_{ij} = \frac{\log(f_{ij} + 1)}{-\frac{\sum_j p_{ij} \log(p_{ij})}{\log(\text{nb de doc})}} \quad (2.3)$$

où :

- f'_{ij} est la pondération du mot i , dans le document j avec l'analyse sémantique latente (LSA) ;
- $\log(f_{ij} + 1)$ est la pondération locale du mot dans le document j , basée sur la fréquence ;
- $-\frac{\sum_j p_{ij} \log(p_{ij})}{\log(\text{nb de doc})}$ est l'inverse de l'entropie, pondérant le mot au niveau global.

À partir de ces descripteurs et d'un algorithme de classification SVM, ils obtiennent un score de f-mesure de 0,925%, soit 17 à 26% de plus que les équipes concurrentes⁸.

7. <http://deft09.limsi.fr/>

8. Actes de DEFT'09 : <http://deft09.limsi.fr/index.php?id=9&lang=fr>

Finalement, les conclusions de DEFT'09 montrent que la catégorisation de la subjectivité s'avère plus pertinente au grain phrase, voire à un grain intraphrastique, plutôt qu'au grain document. La plupart des documents sont en fait un mélange de passages objectifs et de passages subjectifs. Dans ce cadre applicatif, plusieurs travaux ont d'ailleurs cherché à catégoriser des phrases plutôt que des documents (Wiebe et Riloff, 2005 ; Kim et Hovy, 2005a ; Ku, Liu, Lee, Chen et Chen, 2008 ; Wang, Spencer, Ling et Zhang, 2008). Ces travaux représentent une première étape intermédiaire vers l'idée d'une analyse à granularité fine de l'opinion. Nous développons cette idée dans la section 2.2.

2.1.2 Limites des approches au grain document

L'axe des analyses d'opinions au grain document a longtemps marqué l'essor de la fouille d'opinions. L'attrait scientifique pour cette problématique s'explique en partie par :

- l'existence de corpus de références. Ces corpus sont rendus disponibles via les sites de critiques sur le web, les textes politiques où l'étiquette politique est connue et les textes journalistiques dont certaines rubriques sont considérées en amont comme étant subjectives et d'autres comme étant objectives ;
- l'attrait du TAL pour les problématiques de catégorisation de documents à partir de méthodes d'apprentissage artificiel.

Ces corpus de références permettent de définir des protocoles expérimentaux couramment employés dans les travaux de catégorisation de documents (thématique, genre, etc.) et donc de bénéficier d'un cadre scientifique relativement stabilisé. Les mesures de *précision* (éq. 2.4) et de *rappel* (eq. 2.5), pour chaque classe i , et une pondération de ces scores, le *F-Score* (éq. 2.6), permettent d'évaluer les méthodes proposées en les comparant à la classe initiale du document. Ces protocoles mettent relativement peu de moyen humain en jeu et sont donc facilement mis en place.

$$\text{Précision}_i = \frac{\text{NB de documents classés } i \text{ corrects}}{\text{NB de documents classés automatiquement } i} \quad (2.4)$$

$$\text{Rappel}_i = \frac{\text{NB de documents classés } i \text{ corrects}}{\text{NB de documents } i} \quad (2.5)$$

$$\text{F-Score}_\beta = \frac{(\beta + 1) \times \text{Précision} \times \text{Rappel}}{\beta^2 \times \text{Précision} + \text{Rappel}} \quad (2.6)$$

Les corpus utilisés dans cet axe obéissent à des règles discursives particulières qui rendent pertinente l'idée d'une catégorisation de l'opinion au niveau document :

- l'auteur d'une critique a l'*intention* de classer le produit qu'il évalue au moment où il écrit sa critique, donc de lui associer une étiquette particulière ;
- le journaliste a également l'*intention* d'écrire un article objectif pour certaines rubriques ;
- les parlementaires et les sites de partis politiques sont associée à une seule étiquette politique. Il y a également une *intention* de représenter cette étiquette politique dans les discours ;

Dès lors, cet axe nous apparaît comme étant de moins en moins représentatif des besoins applicatifs liés à l'opinion et de moins en moins adapté à l'analyse des discours d'opinions sur d'autres corpus. Par nature, sur tout autre type de document, le discours d'opinion s'articule autour d'une argumentation parfois complexe, mettant en jeu des opinions hétérogènes. On n'y retrouve pas nécessairement l'intention d'une opinion catégorique, il n'est donc pas pertinent de vouloir réduire ces documents à une seule information.

Par ailleurs, nous avons observé que l'analyse de l'opinion au niveau document a tendance à ne pas s'intéresser à caractériser les opinions et la subjectivité dans la langue, mais plutôt à rechercher des éléments qui caractérisent des corpus. Cet aspect entraîne selon nous un biais : les classificateurs sont entraînés sur un type de corpus et ne sont pas portables vers d'autres corpus. Les descripteurs

utilisés sont spécifiques aux corpus, ils ne décrivent pas l'opinion et la subjectivité dans la langue. Nous mettons ces différentes limites en parallèle avec l'intérêt très récent pour l'axe de l'analyse à granularité fine présenté ci-dessous.

2.2 Analyse à granularité fine de l'opinion

2.2.1 Une combinaison de problèmes hétérogènes

Pour pallier les limites de l'analyse au niveau document, l'analyse à granularité fine de l'opinion⁹ est un axe qui a émergé à partir de 2006 (Wilson, Wiebe et Hwa, 2006 ; Wilson, 2008). L'intérêt grandissant pour cet axe est motivé par plusieurs aspects :

- la grande majorité des textes ne véhicule pas une information d'opinion homogène (*un seul sujet, une seule opinion, une seule polarité*, etc.). Le discours d'opinion ne peut pas se résumer à un problème de classification de document où l'opinion est soit totalement positive, soit totalement négative. Il faut donc pouvoir délimiter des unités d'opinions minimales sémantiquement homogènes ;
- la polarité axiologique positive ou négative d'une unité textuelle n'est pas la seule information intéressante. Le champ sémantique d'une opinion (hédonique, esthétique, pragmatique, éthique, croyance, etc.), son intensité, l'engagement du locuteur, etc. sont d'autres aspects qu'il est intéressant de catégoriser au niveau de chaque grain d'opinion ;
- puisqu'un document parle de plusieurs objets et articule plusieurs passages d'opinions, il faut pouvoir identifier le sujet (cible) sur lequel porte chaque passage d'opinion ;
- les différents passages d'opinions d'un document ont potentiellement des énonciateurs différents (source). Identifier chaque source est intéressant dans les documents qui présentent beaucoup de discours rapportés, notamment les articles journalistiques.

L'exemple (2) illustre la complexité des opinions au fil d'un texte (*changement de cible évaluée, rupture du champ lexical et de la polarité axiologique, opinion affective contre opinion morale*, etc.) et rend compte des limites de la catégorisation au niveau document sur ce type de textes.

- (2) Ce livre, *j'avais hâte* de lire car outre le fait de dévoiler la réalité sur Fadela Amara, il m'a permis de découvrir une partie du combat de *ma chouchou !!* [...] *une des personnalités préférées des Français*, [...] et qui *n'a pas froid aux yeux*. Le parcours de Fadela Amara *tient* à la fois *du conte de fées* et de la construction médiatique et politique. [...]
Je me disais : "Tiens, c'est *vraiment un pas en avant* de mettre dans le gouvernement une femme [...], issue des banlieues donc qui *sait de quoi elle parle*. [...] qui n'a pas été dans le monde politique, qui *aura donc une autre vision*". *Quelle déception !!* [...] *Je n'avais pas vraiment une bonne image* des hommes politiques en général [...] qui me donnaient plus l'impression d'être là pour leur propre image personnelle et leurs ambitions *sur-développées*. *Une belle cause, un beau combat détournés pour les ambitions* de certains. Que de *mensonges*, que de *manipulation*... Comment, au nom du pouvoir, peut-on ainsi *briser l'élan* de personnes qui *se battent avec leurs tripes* pour *soutenir des gens qui souffrent* [...]. C'est *à vomir !*

Texte extrait de la plateforme Over-blog - 09/09/2009

9. Terme anglais : *Fine-grained opinion analysis*.

La figure 2.4 présente les objectifs de l'axe de recherche à granularité fine par opposition à ceux de l'analyse au niveau document (voir Fig. 2.2, p. 13). Dans cet axe, on considère que le document n'est pas centré autour d'un seul objet et qu'il présente des passages d'opinions hétérogènes. Par exemple, d'un point de vue applicatif, si l'on recherche les opinions exprimées autour de l'objet C, d'après la figure 2.4 trois documents contiennent cet objet (*doc 1*, *doc 2* et *doc 3*) à des degrés différents. Une analyse de l'opinion à granularité fine cherche notamment à rendre compte que :

- ces documents contiennent des passages d'opinions de catégories sémantiques différentes (par exemple, *positif* ou *négatif*) ;
- ces passages d'opinions concernent un objet particulier parmi les objets du document.

Par exemple, pour l'objet C, le document 1 est neutre, le document 2 contient trois passages d'opinions positifs. Dans le document 3, l'objet C est relié à une opinion négative. Le document 3 est globalement positif, mais les opinions positives portent sur d'autres objets.

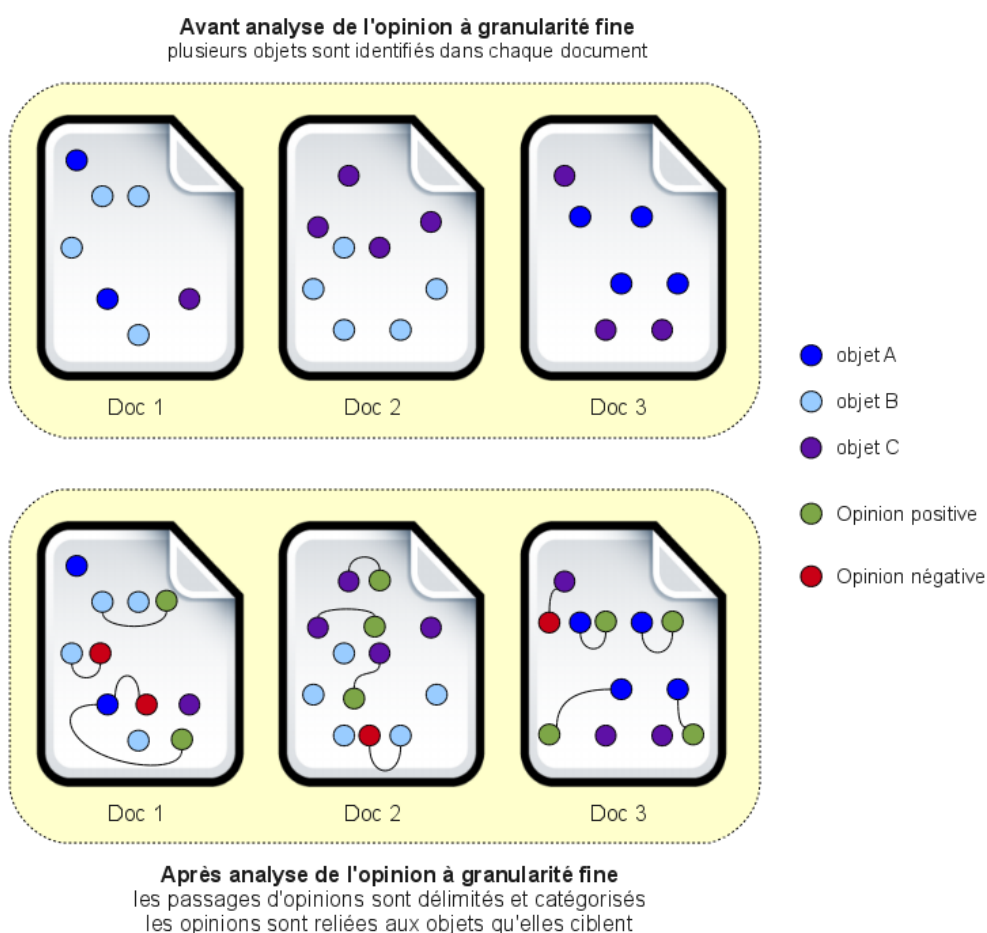


FIGURE 2.4 : Analyse à granularité fine de l'opinion

Nous distinguons trois problématiques dans les travaux appartenant à l'axe de l'analyse de l'opinion à granularité :

- catégoriser sémantiquement les passages d'opinions ;
- identifier la cible des passages d'opinions ;
- identifier la source des passages d'opinions.

Dans cet axe, les travaux du domaine s'intéressent à une problématique particulière à la fois. Elles font l'objet de méthodes et de protocoles évaluatifs distincts.

2.2.1.1 Catégoriser les passages d'opinions

La problématique consistant à délimiter et catégoriser des passages d'opinions à granularité fine se scinde en deux branches de solutions. Dans la première, plusieurs travaux proposent de s'intéresser à catégoriser des **phrases** (Hatzivassiloglou et Wiebe, 2000 ; Yu et Hatzivassiloglou, 2003 ; Wiebe et Riloff, 2005 ; Kim et Hovy, 2005a ; Lin et collab., 2006 ; Ku et collab., 2008 ; Wang et collab., 2008). De façon analogue au grain document, les méthodes proposées cherchent à répondre aux questions suivantes :

- *une phrase est-elle globalement subjective ou objective ?*
- *une phrase est-elle globalement positive, négative, intermédiaire ou neutre ?*

Si cette étape est un premier pas vers une analyse fine, elle n'est toutefois pas encore suffisante pour traiter des longues phrases contenant des passages d'opinions hétérogènes. Dans une seconde branche, les travaux de Wilson (2008) sont une rare contribution s'intéressant aux opinions à un **grain intraphrastique**. Leur méthode cherche à déterminer si :

- *un mot issu d'un lexique de mots d'opinions conserve sa polarité dans son contexte ?*

Nous développons les caractéristiques des travaux existants dans ces deux catégories ci-dessous et nous nous positionnons par rapport à eux.

Grain phrastique Les travaux majeurs de Wiebe et Riloff (2005), englobant une série de travaux sur la subjectivité au grain phrase pour l'anglais (Hatzivassiloglou et Wiebe, 2000 ; Wiebe, Wilson, Bruce, Bell et Martin, 2004 ; Wiebe, Wilson et Cardie, 2006 ; Wiebe et Mihalcea, 2006), sont la principale référence pour ce grain d'analyse. Ce travail a depuis été développé et adapté pour le chinois (Ku et collab., 2008) ou pour le français (Généreux et Poibeau, 2009). Par opposition avec les méthodes d'apprentissage supervisé utilisées au grain document, il est ici nécessaire de se confronter au manque de corpus de référence. La méthode de Wiebe et Riloff (2005) consiste à construire un classifieur automatique de phrases objectives et subjectives à très haute précision pour amorcer automatiquement l'annotation d'un corpus. L'identification à très haute précision de phrases subjectives est effectuée en utilisant un lexique d'indices de subjectivité considérés comme étant certains¹⁰ et un classifieur à base de règles.

patrons subjectifs	patrons objectifs
<subj> believes	<subj> increased production
<subj> was convinced	<subj> took effect
aggression against <np>	delegation from <np>
to express <dobj>	occurred on <np>
support for <np>	plans to produce <dobj>

TABLE 2.2 : Patrons subjectifs et objectifs – (Wiebe et Riloff, 2005)

La démarche de Wiebe et Riloff (2005) consiste à :

- identifier des phrases subjectives et des phrases objectives en se basant sur des indices « forts » et sur quelques règles de catégorisations manuelles. Par exemple, si la phrase ne contient aucun indice fort, elle est objective, si elle contient deux indices forts, elle est subjective ;

10. Terme employé par les auteurs : *Strong subjectivity clues*.

- utiliser les phrases annotés automatiquement pour extraire de nouveaux mots ou des groupes de mots qui discriminent statistiquement un catégorie de phrase (voir Tab. 2.2) ;
- ajouter les mots ou les groupes de mots les plus discriminants à la liste d'indices forts ;
- répéter l'algorithme à partir de l'étape 2.

Les auteurs arrivent ainsi à catégoriser par amorçage un corpus non annoté. Ils utilisent ensuite une méthode d'apprentissage supervisé Naïve Bayes sur ce corpus en utilisant les indices de subjectivité comme descripteurs de phrases. Ils obtiennent une précision de 86.3% et un rappel de 78.1% sur la catégorisation de phrases subjectives. Cette méthode est la méthode état de l'art du domaine. Elle a permis d'initier la création d'un corpus annoté d'opinion, le corpus *MPQA*, et du lexique des mots d'opinions anglais, le *Subjectivity Lexicon* (voir chapitre 4).

Une fois que les phrases subjectives ont été catégorisées, identifier leur polarité est le second objectif. Pour cette problématique, citons les travaux de Ku et collab. (2008) pour le chinois. Ils utilisent une méthode attribuant un score à chaque phrase. Ce score tient compte de la polarité associée à chaque mot de la phrase, de la présence d'un marqueur de négation dans la phrase (celui inverse le score) et la position du mot dans la phrase (la polarité d'un mot en fin de phrase à davantage de poids). Les résultats obtenus (la précision est d'environ 49 %) semblent souffrir des phrases contenant à la fois des opinions positives et négatives, des tournures de négations et des phrases contenant des mots d'opinions bruités. La simple présence d'un marqueur de négation dans la phrase pour inverser la polarité de la phrase est une hypothèse également utilisée par Hu et Liu (2004). Elle nous paraît trop naïve, nous revenons particulièrement sur cette hypothèse dans le chapitre 5 pour proposer une solution alternative.

Grain intraphrastique Les travaux de Wilson (2008) sont ceux qui sont allés le plus loin dans l'idée d'une analyse à granularité fine. Ces travaux se rapprochent le plus de nos propres objectifs. Dans sa thèse, Wilson (2008) considère qu'un mot d'opinion possède une polarité prioritaire que l'on peut stocker dans une ressource. Néanmoins, dans son contexte phrastique, cette polarité prioritaire peut être modifiée par plusieurs facteurs :

- des tournures de négations locales (ex : *not good*) ;
- des tournures de négations dont les dépendances sont éloignées (ex : *does not seem very good*) ;
- certaines modalités affirmant la réalité ou la non réalité d'une proposition (ex : *there is no reason to believe that it is good*) ;
- le sens du mot dans son contexte (ex : *Environmental Trust/ he has won the people's trust*)

Son travail consiste à reconnaître l'effet du contexte sur la polarité d'un mot d'opinion. Dans cet objectif, ils ont annotés manuellement le corpus MPQA en indiquant la polarité de mots ou d'expressions d'opinions dans leur contexte phrastique. Nous présentons plus précisément ce corpus dans le chapitre 4 (§. 4.1.4). Elle utilise ensuite une méthode d'apprentissage visant à entraîner deux classifieurs binaires. Le premier classifieur a pour objectif de catégoriser les mots d'opinions qui expriment une polarité neutre ou non neutre dans leur contexte. Le second classifieur détermine si un mot d'opinion non neutre est positif ou négatif. Wilson (2008) utilisent quatre types de classifieurs : *BoosTexter*, *TiMBL*, *Ripper* et un *SVM* à noyau linéaire. Pour le premier classifieur, chaque mot d'opinion est décrit par une liste de descripteurs :

- la thématique du document ;
- la partie du discours du mot d'opinion, du mot précédent et du mot suivant ;
- un booléen indiquant si le mot d'opinion est précédé d'un adverbe ou d'un adjectif ;
- un booléen indiquant si le mot d'opinion est précédé d'un marqueur d'intensité ;
- un booléen indiquant si le mot d'opinion est précédé d'un marqueur de négation ;
- le nombre d'autres mots d'opinions dans la phrase courante ;

- le nombre d'autres mots d'opinions dans la phrase précédente et suivante ;

Les résultats des différentes méthodes d'apprentissages sont relativement similaires et ne permettent pas de distinguer un classifieur meilleur qu'un autre :

- les mots d'opinions dont la polarité est neutre dans le contexte sont reconnus avec une précision allant de 74 % à 78 % et un rappel de 84 % à 88 % ;
- les mots d'opinions qui ont une polarité dans le contexte sont reconnus avec une précision allant de 64 % à 73 % et un rappel de 49 % à 59 % ;

Nous notons donc que l'apprentissage réalisé à tendance à classer, à tort, des mots d'opinions polarisés comme n'ayant pas de polarité en contexte. Pour le second classifieur, Wilson (2008) utilise les mêmes méthodes d'apprentissage mais avec des descripteurs différents :

- la présence d'un marqueur de négation dans le contexte du mot d'opinion (ce marqueur doit être dans le même syntagme) ;
- la présence dans la phrase d'un marqueur de modalité inversant la polarité du mot d'opinion.

Les résultats des différents classifieurs sont également assez similaires :

- les mots d'opinions positifs dans le contexte sont reconnus avec une précision allant de 71 % à 74 % et un rappel allant de 74 à 77 % ;
- les mots d'opinions négatifs dans le contexte sont reconnus avec une précision allant de 87 % à 89 % et un rappel de 88 % à 89 %.

Parmi les méthodes existantes à granularité fine, l'approche de Wilson (2008) est selon nous la plus ambitieuse et celle qui prend le plus en compte les caractéristiques linguistiques de l'opinion. Elle permet de re-catégoriser des mots ou des expressions d'opinions en tenant compte de leur contexte phrastique. Néanmoins, cette méthode ou une méthode approchante nécessite d'être adaptée pour le français pour lequel les tournures de négation peuvent être plus complexes ou plus éloignées. L'ambiguïté des marqueurs de négation et des marqueurs de modalité est un également élément à mesurer pour le français.

2.2.1.2 Identifier la cible des passages d'opinions

L'identification de la cible ¹¹ d'un passage opinion est une problématique qui a souvent été ignorée dans le domaine. En catégorisation de l'opinion au niveau document, on considère que la cible est un élément connu en amont des traitements et qu'il n'y a qu'une seule cible dans le document. Pourtant, afin d'analyser des textes dont le contenu est de plus en plus hétérogène, cette problématique correspond à un besoin particulièrement criant. Si l'on cherche les opinions exprimées sur une cible (ex : *Raymond Domenech*), il faut pouvoir filtrer toutes les opinions qui co-occurrent dans le cotexte et qui ne porte pas directement sur elle (3).

(3) Domenech [...] la *commission d'enquête* risque d'**accoucher d'une souris**.

Cette problématique nécessite de combiner des techniques connues en traitement automatique des langues (analyse syntaxique, résolution d'anaphore pronominale et nominale, segmentation thématique, etc.) pour proposer une solution souple et efficace. Ce n'est que récemment que cette problématique fait l'objet de quelques travaux pour l'anglais directement centrés sur elle (Nicolov, Salvetti et Ivanova, 2008 ; Kessler et Nicolov, 2009 ; Jakob et Gurevych, 2010). Néanmoins, les méthodes existantes ne proposent que des solutions très partielles pour résoudre des cas simples où restreints à certaines catégories grammaticales de mots d'opinions. Nous présentons plus précisément cette problématique et les méthodes existantes dans le chapitre 7 de cette thèse.

11. Terme anglais : *target*

2.2.1.3 Identifier la source des passages d'opinions

À l'image de la problématique précédente, la question de la source¹² d'une opinion est soulevée par des besoins applicatifs récents. Si cette problématique se rapproche des travaux sur le *discours rapporté* (Jackiewicz, 2006), son inscription et ses particularités dans le domaine de l'opinion en fait l'objet d'études dédiées (Bethard, Yu, Thornton, Hatzivassiloglou et Jurafsky, 2004 ; Choi, Cardie, Riloff et Patwardhan, 2005 ; Kim et Hovy, 2005b ; Ruppenhofer et collab., 2008). Pour Bethard et collab. (2004), l'intérêt pour ce problème dérive des travaux sur les systèmes de questions-réponses où il s'agit de répondre à la question :

- *Que pense X à propos de Y ?* (où X représente une personnalité publique) ;

Ils considèrent ainsi que tout passage textuel d'opinion provient d'un énonciateur *source* dont il s'agit de retrouver l'identité. Dans sa forme la plus simple, l'expression d'une opinion dans un énoncé est portée par l'auteur du texte lui-même, mais elle peut également être rapportée. Il existe deux types de discours rapporté selon la typologie de Jackiewicz (2006) :

- les **citations authentifiées** dans un style direct (4) ou indirect ;
- les **citations interprétées**, qui sont plus complexes à analyser. Jackiewicz (2006) précise ainsi que « dans la plupart des cas, les paroles citées ne font qu'illustrer un discours plus large dont le but communicatif est restitué grâce à un travail interprétatif de l'ensemble de la situation d'énonciation d'origine. L'énonciateur prend alors en charge ce qu'il a identifié et reconstruit comme essentiel dans l'intention communicative manifestée par le locuteur. C'est le rapport subjectif de ce dernier à sa propre parole et à l'effet qu'elle est censée produire qui se trouve ainsi mis en évidence ». Dans (5), la subjectivité d'un énonciateur et d'un locuteur distinct sont ainsi entremêlées. Dans cet exemple, *Jean-Pierre Papin* est l'auteur d'une opinion sur *Karim Benzema* que l'énonciateur résume subjectivement par *est une opinion élogieuse*.

(4) « C'est scandaleux de présenter les choses comme ça », a déclaré *Domenech*.

(5) Invité à s'exprimer sur le potentiel offensif de l'équipe de France, *Jean-Pierre Papin* s'est montré élogieux au sujet de *Karim Benzema*.

Les méthodes existantes se sont essentiellement intéressées aux citations authentifiées. La méthode de Choi et collab. (2005) est basée sur des CRF (Conditionnal Random Fields) et des patrons lexicaux pour extraire les entités nommées d'un texte. Ils utilisent le corpus anglophone NRRC (Wiebe et collab., 2006), où les opinions et les sources d'opinions ont été annotées, pour apprendre statistiquement les chemins lexicaux qui relient le plus fréquemment une source et une opinion. Leur méthode obtient une précision de 81% et un rappel de 60% sur la tâche de détection de la source d'opinion lorsque l'opinion est connue.

Bien que les méthodes du domaine semblent encore perfectibles et que cette tâche n'ait pas été traitée sur le français, nous ne nous positionnons pas sur cette tâche dans le cadre de cette thèse.

2.2.2 Synthèse des réflexions

L'analyse à granularité fine est un axe très récent en fouille d'opinions. Nous en distinguons trois problématiques :

- délimiter et catégoriser sémantiquement les passages d'opinions ;
- identifier les cibles des passages d'opinions ;
- identifier la source des passages d'opinions.

12. Terme anglais : holder ou source

Pour **délimiter** les passages d'opinions à grain fin, la plupart des travaux se sont intéressés au grain de la phrase. Néanmoins ce grain ne paraît pas totalement adapté. Les solutions mises en places ne permettent pas de catégoriser avec précision :

- les phrases qui contiennent des opinions de polarités différentes ;
- les phrases qui contiennent des marqueurs de négation.

De plus, si la phrase contient plusieurs objets, la question de savoir identifier quelle est la cible évaluée n'est pas résolue. De ce point de vue, le travail de Wilson (2008) au grain du mot d'opinion nous paraît le plus prometteur. Celui-ci cherche à désambiguïser un mot ou une expression d'opinion dans son contexte en tenant compte de la thématique du document, de la présence de marqueur de négation dans le syntagme et de celle de marqueur de modalité inversant la polarité. Dans une autre partie, le travail de Wilson (2008) s'intéresse également à **catégoriser** d'autres aspects sémantiques des opinions comme l'intensité forte ou faible. L'idée de dépasser la simple opposition sémantique positif/négatif est un point important. Des travaux vont récemment dans ce sens en introduisant l'idée de prendre en compte différentes catégories discursives ou sémantiques des opinions (Asher, Benamara et Mathieu, 2008 ; Ferrari, Mathet, Charnois et Legallois, 2008).

Enfin, les tâches d'**identification de la cible et de la source** d'un passage d'opinion permettent, une fois qu'un passage d'opinion est correctement détecté, de connaître sur quel objet porte l'opinion et le locuteur qui s'exprime. Pour les besoins applicatifs, ces deux tâches permettent de filtrer les opinions en fonction d'une cible ou d'une source. Les méthodes existantes semblent perfectibles dans le sens où elles permettent de traiter les cas où la source, la cible et le passage d'opinion sont dans la même phrase, mais elles s'avèrent limitées pour traiter des cas plus complexes.

Pour l'analyse à granularité fine, la détection des opinions repose essentiellement sur la qualité des lexiques. Pour constituer ces ressources, trois types d'approches ont été proposées, nous présentons ces approches dans la section suivante. Dans le cadre de l'analyse à granularité fine, celles-ci nous semblent à ré-explore pour améliorer l'aspect qualitatif et quantitatif des méthodes.

2.3 Extraction des mots d'opinions d'une langue

L'extraction et la catégorisation des mots d'opinions d'une langue est une problématique dont les résultats alimentent les systèmes d'analyses au grain document et à grain fin. Si plusieurs types méthodes ont été proposées (§. 2.3.1), il reste assez difficile de mesurer la réelle qualité des ressources qu'elles permettent de développer. En particulier, pour l'analyse à granularité fine, la précision et le rappel des listes de mots d'opinions obtenues est un aspect critique que nous discutons (§. 2.3.2).

2.3.1 Trois types d'approches

Historiquement, la problématique d'extraction et de catégorisation des mots d'opinions fait l'objet des premiers travaux liés à l'opinion (Hatzivassiloglou et McKeown, 1997 ; Wiebe, 2000 ; Turney, 2002). Les travaux anglophones introduisent l'expression *opinion words*, traduite ici par *mots d'opinions*¹³, pour évoquer les mots utilisés fréquemment pour exprimer une opinion. Dans ces travaux, il s'agit, pour une langue donnée :

- d'identifier quels sont les mots d'opinions ;
- de catégoriser leur polarité axiologique (*positive, négative, neutre* ou *ambigü*) ;
- de désambiguïser les mots dont le sens varie selon le contexte.

13. Nous rappelons que nous utilisons dans ce chapitre la terminologie du domaine et que les expressions *mots subjectifs*, *mots évaluatifs* ou *subjectivèmes* nous paraissent plus adéquates (voir chapitre 3).

L'objectif de cette problématique est de construire des ressources lexicales permettant d'alimenter les systèmes d'analyses d'opinions. Dans le domaine, ces ressources ont plus souvent été utilisées pour des analyses à grain large.

En linguistique, la notion de *mot* est mal identifiée. En fouille d'opinions, nous précisons que cet emploi est généralement entendu au sens de lemme¹⁴ (*espoir, gratitude, se lamenter, trembler, etc.*). La question de l'importance des expressions, des néologismes ou des stéréotypes culturels pour exprimer une opinion n'est jamais posée directement dans les travaux du domaine. Dans les ressources, la présence de ces types de mots d'opinions est très réduite. Cela constitue une des limites actuelles (§. 2.3.2). Pour constituer les ressources de mots d'opinions, trois approches ont été proposées :

- une constitution manuelle ;
- une constitution automatique en utilisant des dictionnaires existants ;
- une constitution automatique en utilisant des corpus.

Nous présentons les caractéristiques de ces types d'approches dans les sous-sections suivantes et nous montrons en quoi elles sont perfectibles pour améliorer la qualité des analyses à granularité fine (§. 2.3.2).

2.3.1.1 Approche manuelle

Au delà de l'effort humain à fournir, l'idée d'une approche entièrement manuelle n'est pas une mauvaise idée puisqu'elle permet d'extraire et de catégoriser avec précision le sens de mots d'opinions. Pour l'anglais, plusieurs travaux utilisent des listes de mots d'opinions polarisées et développées manuellement (Morinaga, Yamanishi, Tateishi et Fukushima, 2002 ; Tong, 2001 ; Yi, Nasukawa, Bunescu et Niblack, 2003 ; Das et Chen, 2007). De façon voisine, Strapparava et Valitutti (2004) ont développé manuellement une extension de WordNet (Fellbaum, 1998), WordNet-Affect, pour ajouter des labels affectifs à certains synsets¹⁵ de WordNet (*émotion, état psychologique, signal hédonique, etc.*). La précision et la couverture des ressources ainsi créées sont rarement évaluées directement. La qualité de ces ressources est en général estimée en observant l'amélioration obtenue par les systèmes de catégorisation d'opinions au grain document qui utilisent ces ressources.

Pour la langue française, citons le *lexique des sentiments* de Mathieu (2005), développé manuellement. À notre connaissance, ce lexique était la seule ressource de mots d'opinions existante pour le traitement du français. Il comporte 950 mots lemmatisés (unigrammes) exprimant des sentiments, des émotions et des états psychologiques. Ces mots sont répartis en 38 classes sémantiquement homogènes :

- 22 classes négatives (*Peur, Tristesse, Irritation, etc.*),
- 14 classes positives (*Amour, Intérêt, Passion, etc.*),
- 2 classes sans polarité (*Étonnement et Indifférence*).

Chaque classe est nommée par le sentiment ou l'état psychologique décrit, comme la classe *Peur* qui contient les mots relatifs à un sentiment de peur (*peur, crainte, frayeur, effrayer, effrayant, etc.*). Ces classes sémantiques sont liées entre elles par des relations de sens, d'intensité et d'antonymie. Bien qu'étant une ressource riche de part les traits sémantiques décrits, ce lexique des sentiments est limité par sa taille et la nature de ses entrées (majoritairement des lemmes simples). L'auteur relève aussi l'importance du niveau pragmatique et énonciatif du texte pour mesurer la subjectivité d'un mot et précise que « *la reconnaissance et l'interprétation du simple vocabulaire du domaine ne sont pas suffisantes le plus souvent, il faut prendre en compte d'autres éléments comme les expressions idiomatiques ou figées telle que être la prune de ses yeux* » - p.320.

14. Un lemme est une unité sémantique formée d'une suite de caractères et pouvant constituer une entrée de dictionnaire.

15. Dans WordNet, un synset est une signification particulière d'un mot.

Les listes construites manuellement sont souvent combinées avec des méthodes automatiques (§2.3.1.2, §2.3.1.3). Parmi les protocoles d'évaluations des méthodes automatiques du domaine, les listes de mots d'opinions développées manuellement jouent parfois un rôle de référence pour évaluer la qualité des résultats des approches automatiques. Kamps, Marx, Mokken et de Rijke (2004) évaluent ainsi la liste de mots d'opinions qu'ils extraient automatiquement en se comparant à liste du *General Inquirer* (Stone, 1966) développée manuellement pour l'anglais.

2.3.1.2 Approche par dictionnaire

Les approches par dictionnaire, appliquées essentiellement à l'anglais, tirent principalement profit de la ressource lexicale WordNet (Fellbaum, 1998). Typiquement, il s'agit de choisir manuellement quelques mots grains dont l'axiologie positive ou négative est connue, puis d'augmenter l'ensemble des mots d'opinions en explorant le graphe de synonymes et d'antonymes de WordNet. Les nouveaux mots sont alors ajoutés à la liste de mots grains pour répéter l'expérience et avancer d'un pas dans le graphe de synonymie. Cette méthode est utilisée par Hu et Liu (2004) et Kim et Hovy (2004) pour catégoriser des mots dans le cadre d'analyses au grain document ou au grain phrase. Selon cette idée, Kamps et collab. (2004) s'intéressent uniquement à la polarité axiologique des adjectifs en explorant les graphes sous-jacents de WordNet. À partir d'un adjectif inconnu à catégoriser, il s'agit de parcourir les arcs du graphe pour trouver les chemins les plus courts vers deux mots grains : *good* et *bad*. La polarité de l'adjectif ω est donnée par l'équation 2.7.

$$\text{Polarité}(\omega) = \frac{d(\omega, \text{bad}) - d(\omega, \text{good})}{d(\text{bad}, \text{good})} \quad (2.7)$$

Kamps et collab. (2004) évaluent la qualité de leur méthode de catégorisation d'adjectifs, hors-contexte, en comparaison avec le *General Inquirer* (Stone, 1966). Un défaut des méthodes basées sur WordNet est qu'elles vont ainsi considérer tous les adjectifs ou tous les mots de WordNet comme étant subjectifs dès qu'il existe un chemin de synonymie ou d'antonymie reliant le mot trouvé et un mot grain. Ce biais n'entraîne pas de conséquence trop négative si l'application consiste à catégoriser l'opinion au niveau du document (§. 2.1) (les « faux » mots d'opinions étant globalement compensés par les vrais), mais il entraîne beaucoup de bruits pour les systèmes à granularité fine (§. 2.2).

Dans un autre ordre d'idée, Banea, Mihalcea et Wiebe (2008) présentent une rare approche par dictionnaire n'utilisant pas WordNet, pour l'apprentissage des mots d'opinions du roumain. La méthode développée s'appuie sur 60 mots grains d'opinions roumains choisis manuellement comme étant des mots d'opinions sûrs. Elle utilise un dictionnaire électronique du roumain¹⁶. Les auteurs ne considèrent pas la polarité des mots d'opinions, ils cherchent juste à déterminer si un mot est subjectif ou non. La méthode utilisée se scinde en plusieurs étapes :

- les mots, de plus de trois caractères, qui apparaissent dans la définition des 60 mots grains et qui constituent eux-même une entrée du dictionnaire sont extraits et sont considérés comme des candidats ;
- in filtrage statistique est effectué pour conserver uniquement les mots les plus similaires des 60 mots grains initiaux et répéter le processus d'extraction ;
- la calcul de similarité est basé sur le score de l'analyse sémantique latente (LSA) (Dumais, Furnas, Landauer, Deerwester et Harshman, 1988) des co-occurrences de deux mots dans le corpus SemCor (Miller, Leacock, Tengi et Bunker, 1993).

Après chaque itération, seuls les candidats dont le score LSA est supérieur à un certain seuil sont conservés pour continuer l'expansion du lexique roumain. Les auteurs évaluent en contexte leur

16. <http://www.dexonline.ro>

lexique sur la tâche de catégorisation de phrases subjectives. 504 phrases du corpus SemCor ont été traduites manuellement en roumain. Chacune de ces phrases a une étiquette *subjective* ou *objective*. Banea et collab. (2008) utilisent le classifieur à base de règles de Riloff et Wiebe (2003) et obtiennent un F-Score de 66,2 %.

Nous voyons trois principales lacunes pour ces approches à base de dictionnaire :

- elles ne prennent pas en compte la dimension contextuelle des mots pour décider s'ils sont subjectifs/objectifs et positif/négatif. Or, la subjectivité d'un mot et sa polarité axiologique est souvent dépendante d'un domaine thématique ou d'une cible particulière, par exemple le mot *navet* est négatif s'il évalue un *film* (6) (7) (8) ;
- les approches par dictionnaires sont difficilement adaptables pour le français (et pour les autres langues que l'anglais), puisqu'il n'existe pas de dictionnaire équivalent et aussi exhaustif que WordNet pour le français ;
- les listes extraites semblent efficaces pour catégoriser des documents ou des phrases, elles provoquent beaucoup de bruits pour des analyses à granularité fine.

(6) Je ne suis pas d'accord pour ghost of mars, qui est loin d'être un **navet**.
- **Thème** : Cinéma

(7) Cinéma 3D : «Même en relief, un **navet** reste un **navet**». - **Thème** : Cinéma

(8) Le **navet** est un légume disponible en toute saison. Riche en vitamine C, il est léger en calories et assure une protection contre le cancer. - **Thème** : Cuisine

Un dernier type d'approche, à l'aide de corpus, permet d'extraire et de catégoriser des mots d'opinions. Celle-ci a l'intérêt de ne pas nécessiter de ressources ou d'efforts humains et semble pouvoir permettre de résoudre quelques cas d'ambiguïtés.

2.3.1.3 Approche par corpus

Les approches par corpus sont les plus fréquemment utilisées. Elles sont moins coûteuses que les approches manuelles en plus d'être potentiellement plus exhaustives. Elles ont l'avantage d'être accessibles pour des langues peu dotées en ressources lexicales et semblent également être les seules à proposer une solution pour désambigüiser un mot d'opinion selon le contexte.

Catégoriser la polarité des mots Hatzivassiloglou et McKeown (1997) sont les premiers à s'intéresser à déterminer automatiquement la polarité axiologique des *adjectifs* à partir d'un grand corpus : le *Wall Street Journal* composé de 21 millions de mots. Ils utilisent pour cela deux types d'indices pour l'anglais :

- les adjectifs ayant une même racine et un préfixe ou suffixe différent ont presque toujours une orientation sémantique différente : *adequate/inadequate, thoughtful/thoughtless*,
- le rôle des conjonctions de coordinations entre deux adjectifs contraint leur polarité axiologique. Deux adjectifs liés par une conjonction ont tendance à avoir la même polarité sauf si la conjonction est **but** : *The tax proposal was (9) (10) (11) by the public.*

(9) simple **and** well-received

(10) simplistic **but** well-received

(11) * simplistic **and** well-received

9 256 paires distinctes d'adjectifs ont été extraites. Ces adjectifs sont liés par 5 types de conjonctions : *and, or, but, either-or, neither-nor*. Parallèlement, 1 336 adjectifs et leur polarité ont été annotés manuellement pour servir de référence lors de la phase d'entraînement : 657 positifs, 679 négatifs. Le taux d'accord (89,15%) pour cette tâche d'annotation manuelle a été calculé avec la participation de 4 annotateurs sur 500 adjectifs de cet ensemble. L'algorithme de catégorisation des adjectifs consiste ensuite à se servir de ces indices pour mesurer une dissimilarité entre chaque pair d'adjectifs et regrouper en deux partitions l'ensemble des adjectifs pour retrouver leur polarité. La méthode est évaluée sur des sous-ensembles des adjectifs catégorisées manuellement et obtient un score de précision variant de 78% à 92%.

Kanayama et Nasukawa (2006) se sont appuyés sur une idée voisine selon laquelle l'usage cotextuel de deux mots d'opinions dans un corpus contraint leur sens. Ils considèrent que deux mots utilisés plus fréquemment dans une même phrase ou dans un voisinage proche auront ainsi tendance à avoir la même polarité axiologique selon une hypothèse de cohérence discursive. Kaji et Kitsuregawa (2007), Turney (2002), Bestgen (2002) et Wiebe et Wilson (2002) ont également suivi cette hypothèse dans leur méthode. Nous revenons sur ces hypothèses dans le chapitre 6. Kanayama et Nasukawa (2006) travaillent spécifiquement sur les forums et considèrent quatre domaines thématiques : les *appareils photos*, les *films*, les *téléphones portables* et les *voitures*. Les auteurs émettent la remarque que certains mots ou expressions d'opinions appris sont spécifiques à un corpus thématique (*posséder un mode caméra* est une expression d'opinion seulement dans le corpus des *appareils photos* par exemple). L'utilisation de corpus monothématiques est ainsi la première idée pour identifier certains mots d'opinions dont la polarité est spécifique à une thématique.

Désambigüiser les mots d'opinions selon leur contexte thématique Harb et collab. (2008) considèrent que l'utilisation des adjectifs est statistiquement dépendante du domaine thématique. Par exemple, sur un certain volume de données, on n'emploiera statistiquement pas les mêmes adjectifs pour exprimer une opinion sur un film que pour parler d'un homme politique. De plus, un adjectif n'a pas toujours le même sens et la même polarité selon le contexte d'utilisation (12) (13) :

(12) La qualité d'image de cet appareil est **haute/basse**. - **positif/négatif**

(13) Le plafond de l'appartement est **haut/bas**. - **neutre**

Pour retrouver la polarité des mots d'opinions selon une thématique, la méthode développée par Harb et collab. (2008) consiste à :

- extraire automatiquement un corpus monothématique à partir du web (ici : *le cinéma*). Chaque document exprime un avis positif ou négatif. Ils sont extraits à partir de requêtes envoyées à Google (par exemple : (14)) et à une liste de 14 mots grains dont la polarité est connue ;
- extraire les adjectifs qui co-occurrent fréquemment avec les mots grains ;
- catégoriser les nouveaux adjectifs à partir de leur similarité avec les deux classes de mots grains. La mesure de similarité *Acrodef* (Roche et Prince, 2007) est utilisée pour établir la dépendance fréquente de deux mots (eq. 2.8).

(14) +opinion +review +cinema +good -bad -nasty -poor -negative -unfortunate
-wrong -inferior - **requête Google**

$$\text{Acrodef}(X, Y) = \log_2 \frac{nb(X, Y)^3}{nb(X).nb(Y)} \quad (2.8)$$

- avec NB(X,Y) : nombre de co-occurrences du mot X et Y dans un même document.

Le lexique développé pour la thématique *cinéma* n'est pas évalué directement mais les auteurs l'utilisent pour catégoriser la polarité de l'opinion au niveau document dans un corpus de critiques de film. Les résultats ainsi obtenus sont meilleurs avec le lexique appris qu'avec une méthode d'apprentissage statistique SVM utilisant uniquement les mots comme descripteurs ou un lexique non désambiguïé. Néanmoins, Ding, Liu et Yu (2008) ont montré qu'un même mot pouvait avoir différentes polarités dans un même domaine. Par exemple, dans le domaine des appareils photos numériques, le mot *long* exprime différentes polarités dans les exemples (15) et (16) :

(15) La durée de vie de la batterie est **longue**.
cible

(16) Le temps d'ajustement du focus est **long**.
cible

Ainsi, résoudre l'ambiguïté des mots d'opinions selon un domaine thématique est encore insuffisant. L'idée clé de Ding et collab. (2008) consiste alors à désambiguïser le mot d'opinion à partir de l'objet cible qu'il évalue. Dans (15) et (16), c'est la connaissance sur la *durée de vie des batteries* et le *temps d'ajustement du focus* qui permet de déterminer la polarité de l'adjectif *long* dans ces deux cas. Pour valider leur hypothèse, Ding et collab. (2008) ont travaillé sur un corpus monothématique de textes d'opinions issus d'un forum sur les appareils photos. À partir d'une liste de mots grains dont la polarité est connue (ex : *belle, joli*, etc.) et grâce à une hypothèse de cohérence cotextuelle, leur méthode permet dans un premier temps de déduire la polarité des mots du cotexte. Ainsi, dans l'exemple (17) et grâce à la présence de l'adjectif *belle* dans le cotexte, l'adjectif *long* est considéré comme étant positif. Le fait de travailler sur un corpus monothématique dont le concept central (*appareil photos*) est stable permet aux auteurs de connaître à l'avance les objets gravitant autour de l'objet principal (*qualité d'image, photos, durée de vie de batterie*, etc.). À partir d'une petite liste de patrons syntaxiques et du corpus d'étude, leur méthode détermine ainsi que le couple *<long, durée de vie de batterie>* est positif, alors que *<long, focus>* est déterminé comme étant négatif à partir d'un autre exemple.

(17) Cet appareil prend de **belles** photos et sa durée de batterie est **longue**.
cible

Bien que l'idée d'établir des couples *<opinion, objet cible>* pour désambiguïser certains mots d'opinions soit séduisante, elle soulève des questions sur la faisabilité de cette méthode pour traiter d'autres domaines que celui des appareils photos. En effet, les textes de forums sur les *appareils photos* ont une structure assez stable, impliquant des objets bien identifiés et une richesse de phénomènes linguistiques relativement limitée. Typiquement, les mots d'opinions sont des adjectifs simples, les sujets cibles sont situés à proximité. La désambiguïstation des mots d'opinions restent donc une question majeure.

2.3.2 Des résultats difficilement évaluables

D'après l'étude des travaux existants, la qualité des méthodes d'extraction et de catégorisation de mots d'opinions et des ressources lexicales qu'elles permettent d'obtenir nous semblent en fait difficilement évaluables :

- sur l'aspect **qualitatif**, elles sont soit évaluées par rapport à une ressource lexicale développée manuellement (elle-même non évaluée), soit elles sont évaluées indirectement sur les résultats obtenus par une analyse d'opinion au grain document ou au grain phrase. Pour ces grains d'analyses, les erreurs d'extraction et de catégorisation de mots d'opinions ont tendance à être dissimulées car elles peuvent être compensées par d'autres mots d'opinions présents dans le document ou la phrase ;

- sur l'aspect **quantitatif**, elles ne sont pas évaluées. Seul un indicateur du nombre d'entrées lexicales est connu pour chaque ressource. Nous savons également que certaines ressources sont restreintes aux adjectifs ou aux lemmes simples.

Dès lors, s'il est assez difficile de se faire une idée précise sur ces méthodes et ressources, nous observons néanmoins que les ressources lexicales de référence pour l'anglais contiennent de nombreuses entrées susceptibles de provoquer du bruit. Par exemple, le Subjectivity Lexicon considère comme fortement subjectif des mots comme *baby* (bébé) ou *inundated* (inondé). Le SentiWordNet (Esuli et Sebastiani, 2006) considère négatif les mots *little* (petit) ou positif le mot *to proof* (prouver).

Par ailleurs, de notre point de vue, les méthodes et les ressources actuelles ne permettent pas de saisir une grande partie du vocabulaire d'opinions (expressions, néologismes, stéréotypes culturels). Il nous semble nécessaire d'évaluer la qualité des méthodes et des ressources, en contexte, à partir d'une analyse à granularité fine pour voir quels sont réellement les mots d'opinions qui échappent aux méthodes actuelles et ceux qui provoquent des erreurs ou des ambiguïtés. Nous rediscutons plus précisément de ces questions dans le chapitre 6.

2.4 Conclusion

2.4.1 Synthèse du domaine de la fouille d'opinions

L'évolution des réflexions en fouille d'opinions depuis une dizaine d'années peut se résumer à travers trois axes :

- l'analyse de l'opinion au grain document ;
- l'analyse de l'opinion à granularité fine ;
- l'extraction et la catégorisation des mots d'opinions d'une langue ;

Les travaux sur l'opinion au **grain document** sont à l'origine de l'émergence du domaine. Ils tirent principalement profit de corpus de référence disponibles et des méthodes de classification de documents pour traiter trois types d'applications : la catégorisation de critiques de produits selon la polarité des opinions, la catégorisation de textes politiques selon l'étiquette politique et la catégorisation d'articles de presse selon la subjectivité. Néanmoins, la portabilité des méthodes à d'autres types de corpus et à d'autres applications reste limitée. Ces limites justifient l'idée d'une analyse à **granularité fine** qui puisse être davantage générique à tous types de thématiques et de corpus. L'axe d'analyse à granularité fine est beaucoup moins abordé dans la littérature. Il s'intéresse récemment à catégoriser des passages d'opinions phastiques ou intraphrastiques dans un document, à en catégoriser certains aspects sémantiques (polarité, intensité, etc.) et à en identifier la cible et la source. L'essor récent de cet axe et la complexité de ses problématiques rend encore très perfectibles les méthodes proposées jusqu'à présent. De plus, il existe peu de travaux francophones dans cet axe. Enfin, la problématique d'**extraction et de catégorisation des mots d'opinions** d'une langue est un point central de ces deux axes d'analyses. Néanmoins, à granularité fine, la qualité et la couverture des lexiques est davantage un facteur clé. De la même façon, comparativement à l'anglais, le français reste assez pauvre en ressources lexicales de mots d'opinions.

2.4.2 Positionnement de nos travaux

Notre travail se positionne dans l'axe de l'analyse de l'opinion à granularité fine. Cet axe a commencé à se développer parallèlement à cette thèse et la jeunesse de ce domaine explique qu'il n'existe pas encore à l'heure actuelle de réel consensus autour :

- du grain d'analyse à adopter¹⁷ ;

17. Pour pouvoir évaluer les méthodes des participants, les organisateurs de la campagne d'évaluation DEFT'09 ont ainsi

- des protocoles expérimentaux à mettre en place : *quels corpus utiliser ? quelles mesures d'évaluations des méthodes et des ressources utiliser ? , comment comparer les méthodes entre elles ? ;*
- des caractéristiques sémantiques de l'opinion (mis à part l'aspect positif/négatif).

Dans ce cadre, notre travail se veut, d'une part, exploratoire en visant à structurer les différents problèmes de TAL liés à la notion d'opinion. Dans cet optique, notre démarche nous amène à recadrer l'opinion pour l'englober dans les notions linguistiques plus larges de subjectivité et d'évaluation. Nous adoptons un modèle pour décrire le langage de l'évaluation basé sur les théories linguistiques de Charaudeau (1992) et Galatanu (2000). Nous approfondissons cet aspect dans le chapitre 3. D'autre part, notre travail vise à développer des méthodes et des ressources pour s'attaquer aux problématiques de cet axe. Notre effort se concentre sur la langue française pour laquelle il existe beaucoup moins de lexiques et de corpus que pour l'anglais. Nous nous intéressons particulièrement aux problématiques :

- de délimitation et de catégorisation de passages d'opinions. Nous souhaitons dépasser la simple opposition positif/négatif en caractérisant davantage d'aspects sémantiques de l'opinion ;
- d'amélioration de la précision et du rappel des méthodes d'extraction et de catégorisation des mots d'opinions ;
- d'identification de la cible évaluée par un passage d'opinion ;

Nous proposons des méthodes pour chacune de ces problématiques et nous les évaluons en contexte sur des corpus tests. Nous n'aborderons pas la problématique de l'identification de la source d'un passage d'opinion. Dans le cadre de cette thèse, nous considérerons pour l'instant que cette source est toujours l'auteur du document. D'un point de vue plus général, nos objectifs s'articulent autour de l'idée de travailler dans un cadre multithématique pour permettre aux méthodes et ressources créées d'être utilisables quelque soit la thématique.

Le chapitre suivant nous amène au besoin initial d'identifier les caractéristiques sémantiques des opinions. Pour cela, nous souhaitons mieux prendre en compte les notions de subjectivité et d'évaluation définies dans des travaux majeurs en linguistique. Ceux-ci définissent l'opinion, ou plutôt l'évaluation, comme un phénomène sémantique et énonciatif complexe et proposent plusieurs modèles pour le représenter. Nous nous inspirons de ces modèles pour en proposer une version simplifiée et adaptée à nos travaux informatiques.

proposés un compromis entre le grain phrase et le grain mot/expression en considérant le passage entre ponctuations comme unité minimale d'opinion.

La subjectivité dans le langage

3.1	La subjectivité dans le langage : un phénomène énonciatif	35
3.1.1	La problématique de l'énonciation	35
3.1.1.1	<i>Définition de la subjectivité</i>	36
3.1.1.2	<i>Deux attitudes énonciatives opposées : le discours et le récit</i>	36
3.1.2	Un 1 ^{er} niveau de subjectivité : les traces de l'énonciateur	39
3.1.2.1	<i>Les embrayeurs de personnes</i>	39
3.1.2.2	<i>Les embrayeurs spatio-temporels</i>	40
3.1.3	Un 2 ^{ème} niveau de subjectivité : les traces d'une évaluation	41
3.1.3.1	<i>Définition de l'axiologie</i>	42
3.1.3.2	<i>Une structuration axiologique complexe au niveau lexical</i>	42
3.1.3.3	<i>Quelques caractéristiques morphologiques de l'axiologie</i>	43
3.1.3.4	<i>Effets des structures grammaticales de négation sur l'axiologie</i>	44
3.1.4	Synthèse : deux configurations d'énonciations subjectives possibles	45
3.2	Différentes modalités énonciatives pour exprimer une évaluation	46
3.2.1	Définition de la modalité	46
3.2.2	Les modalités logiques	47
3.2.2.1	<i>Opinion</i>	47
3.2.2.2	<i>Accord et désaccord</i>	47
3.2.3	Les modalités axiologiques	48
3.2.3.1	<i>Appréciation</i>	48
3.2.3.2	<i>Jugement</i>	49
3.2.4	Combinaison de modalités	49
3.3	Vers un modèle informatif du langage de l'évaluation	50

Ce chapitre a pour premier objectif de définir les concepts linguistiques liés à notre travail et d'éclaircir la terminologie utilisée en fouille d'opinions : *opinion*, *subjectivité*, *sentiment*, *évaluation*, *polarité*, etc. En linguistique, la notion de **subjectivité dans le langage** est celle qui regroupe ces différents termes. Nous la définissons en nous appuyant sur le travail fondateur de Benveniste (1966), puis principalement sur les travaux de Kerbrat-Orecchioni (1980). Leurs travaux replacent la problématique de l'*énonciation* au coeur de la définition de subjectivité (§. 3.1). Ils distinguent deux niveaux de subjectivité dans le langage (Figure 3.1) :

- le premier niveau témoigne du degré de présence de l'énonciateur dans son énoncé. Cette présence est rendue *visible* ou *dissimulée* par la présence ou l'absence de certains marqueurs (§. 3.1.2). Ce niveau de subjectivité n'implique pas l'expression d'une évaluation ;
- le second niveau rend compte des évaluations exprimées par l'énonciateur. Elles se manifestent par un phénomène de modalisation (Bally, 1932) et la sélection d'unités linguistiques évaluatives spécifiques qui peuvent être axiologiques (positives ou négatives) (§. 3.1.3).

Cette subjectivité à double niveau est souvent à l'origine de confusions sur la définition de cette notion, notamment dans les travaux en informatique. En synthèse de cette première partie, nous rediscutons de l'importance de la configuration d'énonciation d'une évaluation (§. 3.1.4).

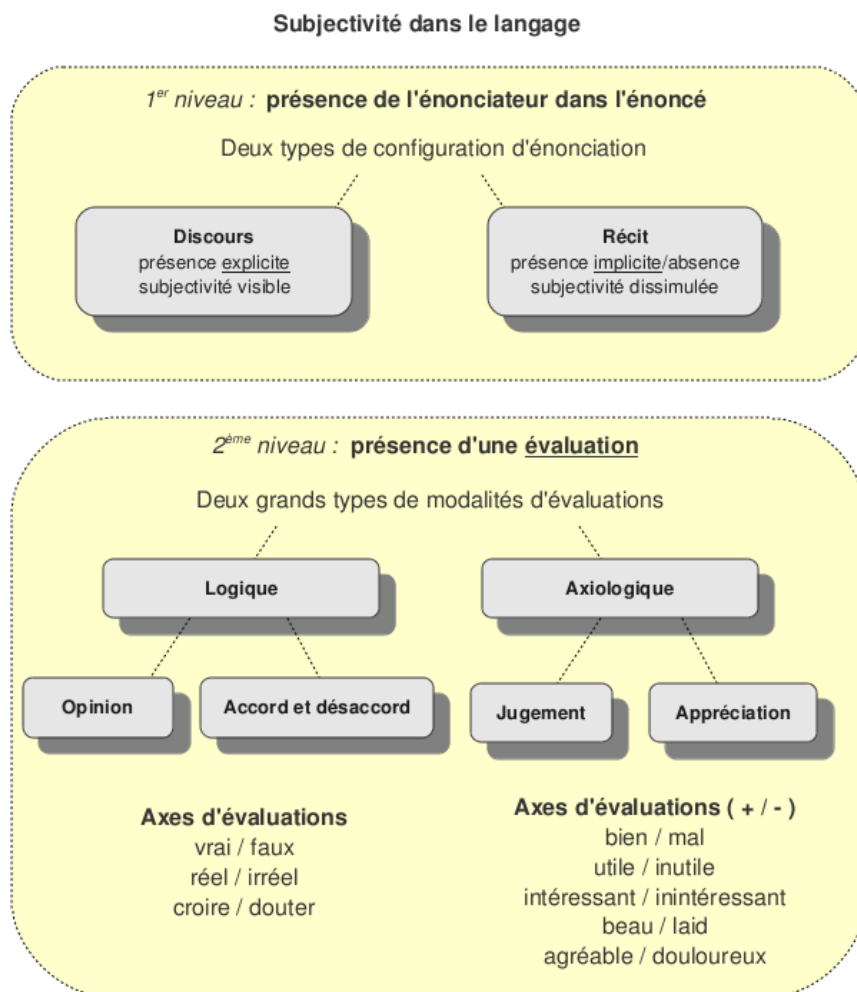


FIGURE 3.1 : Les deux niveaux de subjectivité dans le langage

Dans une seconde partie, nous nous intéressons en particulier aux évaluations. Charaudeau (1992) et Galatanu (2000) structurent le langage de l'évaluation en le scindant en plusieurs types de modalités énonciatives (§. 3.2). Après avoir défini le concept de modalité (§. 3.2.1), nous adaptons leurs travaux pour considérer qu'il existe deux grandes catégories de modalités pour exprimer une évaluation :

- les modalités logiques : l'opinion et l'accord/désaccord (§. 3.2.2) ;
- les modalités axiologiques : le jugement et l'appréciation (§. 3.2.3).

Cette étude linguistique sur la subjectivité et l'évaluation nous permet de définir l'évaluation et ses principaux traits à caractériser dans une optique de traitement automatique des langues (§. 3.3).

3.1 La subjectivité dans le langage : un phénomène énonciatif

La notion de **subjectivité dans le langage** a été introduite par Émile Benveniste (1966), puis a été notamment reprise, dans le même courant d'idée, par Anscombe et Ducrot (1976) ou Kerbrat-Orecchioni (1980). Pour en comprendre la définition, il est nécessaire de définir la problématique de l'énonciation.

3.1.1 La problématique de l'énonciation

Le courant linguistique de l'énonciation considère que tout énoncé est le pendant d'un événement unique et que la production de cet énoncé prend place dans une situation de communication qui suppose : un moment, un lieu, un énonciateur, un destinataire (définition 1). En résumé, l'énonciation est l'acte individuel de production et d'utilisation de la langue dans un contexte déterminé, ayant pour résultat l'énoncé. Les deux termes s'opposent comme la fabrication s'oppose à l'objet fabriqué.

Définition 1 (Énonciation) *L'énonciation est l'activité langagière exercée par celui qui parle au moment où il parle.* - p.18, (Anscombe et Ducrot, 1976). *L'énonciation est cette mise en fonctionnement de la langue par un acte individuel d'utilisation.* - p.80, (Benveniste, 1966).

Dans la tradition de la linguistique structurale née de Saussure (1916), la langue est un objet d'étude extérieur au monde, elle s'oppose à la **linguistique de l'énonciation** qui étudie le langage mis en situation par l'activité d'un énonciateur. Avec l'énonciation, c'est l'acte même de produire un énoncé et non simplement l'énoncé lui-même qui est étudié. Cependant, faute de pouvoir étudier directement l'acte de production, la problématique de l'énonciation consiste à identifier et décrire les traces de l'acte dans le produit, c'est-à-dire les lieux d'inscription des différents *constituants* du cadre énonciatif perceptibles dans l'énoncé (Kerbrat-Orecchioni, 1980). Parmi ces constituants, les énonciativistes s'intéressent notamment à :

- la place des protagonistes du discours (la place de l'énonciateur¹ et du destinataire²),
- la situation de communication et ses différents aspects (compétences linguistiques, idéologiques de l'énonciateur/du destinataire, encodage du message, décodage, etc.).

Dans cet optique, l'énonciateur est le constituant privilégié des théoriciens de l'énonciation. Ils cherchent à définir le mécanisme d'engendrement d'un texte et le surgissement dans l'énoncé de l'**énonciateur** (définition 2). Les linguistes précisent ainsi qu'il importe de distinguer rigoureusement ce qui est dit – l'énoncé – et la présence de l'énonciateur à l'intérieur de son propre discours – l'énonciation.

1. Celui qui est à l'origine de l'énoncé.

2. Celui à qui est destiné le message porté par l'énoncé.

Définition 2 (Problématique de l'énonciation) *Cette problématique consiste à rechercher les procédés linguistiques (embrayeurs, modalisateurs, termes évaluatifs, etc.) par lesquels l'énonciateur imprime sa marque à l'énoncé, s'inscrit dans le message (implicitement ou explicitement) et se situe par rapport à lui-même (distance énonciative) (Kerbrat-Orecchioni, 1980).*

De cette problématique naît la notion de subjectivité dans le langage qui intervient au niveau de la production de l'énoncé par un énonciateur et non pas simplement dans ce qui est dit. Nous en précisons la définition dans le paragraphe suivant.

3.1.1.1 Définition de la subjectivité

Nous adoptons la définition de la subjectivité dans le langage établie par Benveniste (1966). Cette définition a depuis été développée en la confrontant aux analyses en corpus. Elle reste néanmoins toujours validée par les travaux plus récents en linguistique de l'énonciation (Kerbrat-Orecchioni, 1980 ; Maingueneau, 1991).

Définition 3 (Subjectivité dans le langage) *La **subjectivité** [...] est la capacité du locuteur à se poser comme sujet dans et par le langage. Elle se définit, non par le sentiment que chacun éprouve d'être lui-même [...], mais comme l'unité psychique qui transcende la totalité des expériences vécues qu'elle assemble, et qui assure la permanence de la conscience. Or nous tenons que cette **subjectivité** [...] n'est que l'émergence dans l'être d'une propriété fondamentale du langage. Est ego qui dit ego. [...] la subjectivité se détermine par le statut linguistique de la **personne**. - p.259-260, (Benveniste, 1966).*

À travers cette définition, Benveniste amorce la remise en question de la notion de **personne** en linguistique. En effet, la définition grammaticale traditionnelle distingue six personnes représentées dans la langue par les pronoms : *je, tu, il, nous, vous, ils*. Or, ce que la grammaire appelle la *troisième personne (il(s), elle(s))* peut renvoyer à un objet, et fait normalement référence à un élément exclu de la communication. À l'inverse, seuls *je* et *tu* (et leurs extensions, *nous* et *vous*) sont de vraies personnes, puisqu'ils renvoient nécessairement à une personne présente dans l'acte de communication. Les linguistes opposent ainsi la non-personne (*il*) aux personnes (*je, tu*). La subjectivité dans le langage se détermine donc par des marqueurs de la présence du *je* et du *tu* selon la définition de Benveniste.

Dès lors, cette définition de la subjectivité oriente les linguistes vers l'identification et l'analyse des marqueurs de subjectivité, c'est à dire aux *traces*, laissées volontairement ou involontairement par l'énonciateur, incluant celui-ci dans son propre énoncé. Les énonciativistes préviennent que l'objectivité totale d'un énoncé est une illusion : un énoncé est nécessairement le résultat d'un acte personnel dans le choix des événements décrits dans le message, dans le choix de l'encodage lexical, etc. En revanche, c'est la visibilité de la présence de l'émetteur qui s'échelonne et qui peut être réduite à un minimum pour donner un effet d'objectivité (Olga Galatanu (2000) parle ainsi d'objectivation et de subjectivation). Selon cette idée, le courant linguistique de l'énonciation détermine deux types d'attitudes énonciatives : l'une tend à l'objectivité (ou donner une impression d'objectivité), le **récit**, et une autre admet la subjectivité en plaçant l'énonciateur dans son énoncé : le **discours**.

3.1.1.2 Deux attitudes énonciatives opposées : le discours et le récit

Deux ensembles majeurs de marques énonciatives ont été distingués. Ces deux systèmes énonciatifs distincts témoignent de deux attitudes locutives vis-à-vis de l'énoncé : le **discours** et le **récit**.

Le **discours** met en relation l'énoncé dans le contexte d'une situation d'énonciation. Il contient notamment des marques de personnes et de temps qui reflètent le moment où le discours est prononcé.

Ces marques sont appelées *embrayeurs*³ ou *déictiques*. Pour ce type d'énonciation, le temps de référence est le présent de l'indicatif construit par l'actualité de l'énonciateur, il alterne avec le passé composé ou l'imparfait et le futur proche ou le futur simple. L'énonciateur est inscrit dans l'énoncé par l'apparition des marques de personnes. L'exemple textuel 1 illustre ce type d'énonciation. De nombreux **embrayeurs** (*en quelque sorte, bref, ce, lisez, etc.*), surlignés en bleu, et des **mots évaluatifs** (*mensonge, victime, succès, etc.*), surlignés en rouge, sont présents tout au long du discours et montrent la présence de l'énonciateur. Ces deux types d'indices révèlent deux niveaux de subjectivité. Nous les décrivons dans les sections 3.1.2 et 3.1.3.

La vérité sur le blocus de Gaza avouée par les sionistes eux-mêmes

texte issu d'un blog personnel | 11.06.2010

Dans l'**archipel des mensonges** sionistes, il en est un qui **a la vie dure**, **c'**est que le blocus imposé à la bande de Gaza a pour finalité d'empêcher **ce** territoire peuplé surtout de réfugiés Palestiniens de s'armer et de continuer à tirer ses roquettes artisanales sur les localités sionistes avoisinantes. Une de **ces** localités est Sderot, anciennement Najd avant que sa population indigène en soit **chassée** pour faire de la place aux **squatters** sionistes. Sderot est **en quelque sorte** le Dresde de l'entité sioniste, une cité **martyre victime** d'une **avalanche de roquettes** artisanales tirées depuis la bande de Gaza. Sauf que, à la différence de Dresde ou de Gaza, les victimes à Sderot se comptent **sur les doigts de la main**. N'empêche que les visites organisées à Sderot par les autorités sionistes pour faire comprendre les difficultés endurées par ses habitants ont **un franc succès** et les visiteurs sont amenés à **compatir** devant une vitre brisée ou un toit endommagé. **Bref**, l'entité sioniste a le droit de se défendre et donc d'imposer un blocus à Gaza afin d'empêcher le Hamas de se procurer des armes éventuellement plus performantes que ses roquettes d'aéroclub. Cette vérité communément admise n'est cependant qu'un **mensonge de plus** de la part des sionistes et de leurs complices un peu partout. **Lisez** donc **ce** passage dans l'article où un membre du Département d'Etat (le service de Mme Clinton) explique ne pas être au courant de l'existence d'un document de l'Etat sioniste qui contredit **son propre mensonge**. **Ce** serait **presque comique** s'il ne s'agissait de questions **très graves**. **Ce** n'est pas par les journaux français que **vous** apprendrez **cela** **malheureusement**, mais par l'agence de presse américaine McClatchy. Cependant, sans être une agence marginale, McClatchy est **loin d'avoir le poids** des services de Reuters, de l'AFP ou même du New York Times. Cette agence suit un procès qui se déroule **en ce moment** dans l'entité sioniste où une organisation humanitaire a déféré l'**Etat voyou** devant ses propres tribunaux. Sommé de s'expliquer sur les motifs du blocus qu'il impose, l'Etat délinquant s'est exécuté pour reconnaître **nettement** que le blocus n'est pas lié à des impératifs sécuritaires mais à une volonté de mettre un terme au contrôle qu'exerce le Hamas sur la bande de Gaza.

Comment dit-on déjà en français ? **Ah oui** : **cela** va sans dire, mais ça **va mieux** en le disant ?

ILLUS. TEXTUELLE 1: Texte issu d'un blog évoquant le blocus de Gaza et illustrant le style d'énonciation de type *discours* et de quelques marqueurs subjectifs

Par opposition, le **récit** est en rupture avec la situation d'énonciation, dont il ne montre pas

3. Terme défini dans la section 3.1.2

les traces. Il convoque un mode de référence sans embrayeurs, les personnes ne sont pas absentes mais effacées. Émile Benveniste dit ainsi que *les événements semblent se raconter eux-mêmes* et que *personne ne parle ici*. L'énonciateur prend une grande distance par rapport à son énoncé, qui s'inscrit dans une visée d'objectivité. L'énoncé est alors caractérisé par l'occurrence de la **non personne** et du passé simple, temps de base de l'énonciation historique, qui s'associe à l'imparfait de l'indicatif. À l'intérieur d'un récit, le discours rapporté permet d'insérer les paroles des personnages. L'auteur peut les rapporter de trois manières : en style direct, en style indirect, en style indirect libre. Ce type d'énonciation, cherchant à donner l'illusion d'objectivité dans le langage, est notamment lié à quelques domaines de production d'énoncés comme le *journalisme de récit*, le *récit de fictions* ou encore la production d'*articles scientifiques*. L'exemple 2 donne une illustration d'énonciation qui tend vers le récit. Seules les citations, surlignées en bleu, contiennent des marqueurs subjectifs. L'énonciateur a explicitement pris soin de s'exclure de ces propos par l'usage de guillemets.

A Gaza, le secrétaire général de la Ligue arabe demande la fin du blocus
LEMONDE.FR avec AFP | 02.06.2010

Le secrétaire général de la Ligue arabe, Amr Moussa, a réclamé dimanche la levée du blocus imposé par Israël à la bande de Gaza et appelé à la réconciliation interpalestinienne, lors d'une visite inédite dans l'enclave palestinienne.

"Ce blocus doit être brisé. La position de la Ligue arabe est claire", a assuré M. Moussa après avoir traversé à Rafah la frontière entre l'Égypte et la bande de Gaza. "Non seulement les Arabes mais le monde entier doit soutenir le peuple palestinien contre le blocus de Gaza et face à ce qui se passe dans les territoires occupés, en particulier à Jérusalem-Est", a ajouté le secrétaire général de la Ligue arabe.

M. Moussa visite pour la première fois la bande de Gaza alors que la communauté internationale fait pression sur Israël pour qu'il allège le blocus du territoire palestinien. Il a été accueilli par des membres du mouvement islamiste palestinien Hamas, qui a pris le contrôle de la bande de Gaza il y a trois ans, et par des représentants des autres factions palestiniennes.

"LA RÉCONCILIATION PALESTINIENNE EST FONDAMENTALE"

Au cours de cette visite de quelques heures, M. Moussa doit rencontrer des responsables de l'ONU et d'ONG, le premier ministre du Hamas, Ismaïl Haniyeh, et les chefs des principaux mouvements palestiniens à Gaza. "La réconciliation [entre factions palestiniennes] est fondamentale, c'est une question de principe, c'est une question de volonté et pas simplement de signature", a plaidé le chef de la Ligue Arabe. "Il s'agit de volonté, de décision politique, c'est une position qui doit se traduire par un accord sur toutes les questions", a-t-il ajouté. [...]

La visite de M. Moussa, effectuée selon lui pour marquer sa "solidarité" survient à la suite du raid meurtrier israélien contre une flottille humanitaire internationale en route vers Gaza le 31 mai, qui a fait neuf morts parmi les passagers d'un navire turc et des dizaines de blessés.

ILLUS. TEXTUELLE 2: Texte issu d'un article de LeMonde.fr évoquant le blocus de Gaza et illulant le style d'énonciation de type *récit*

Dans la majorité des textes, le récit et le discours sont la plupart du temps entremêlés. Le linguiste allemand Weinrich (1989) parle également de *monde raconté* pour le récit et de *monde commenté* pour le discours. Ces expressions permettent de noter l'importance du discours dans les textes : des *commentaires* sont presque toujours insérés dans la narration des faits, qu'ils prennent

la forme d'adverbes de phrase, de noms de qualité ou d'adjectifs évaluatifs, qui constituent tous des éléments subjectifs, les traces d'une subjectivité énonciative. Dans les sections suivantes, nous définissons plus précisément les deux niveaux de subjectivité proposés par Kerbrat-Orecchioni (1980). Ils représentent :

- les traces de l'énonciateur dans son discours (par l'usage d'*embrayeurs*) ;
- ou les traces d'une évaluation de l'énonciateur (par l'usage de *mots subjectifs ou évaluatifs*⁴).

3.1.2 Un 1^{er} niveau de subjectivité : les traces de l'énonciateur

Il existe dans la langue une série de termes par lesquels un énonciateur se positionne dans son énoncé. C'est le cas de nombreuses unités linguistiques qui ne prennent sens qu'à l'occasion d'un acte d'énonciation particulier. Ces unités sont appelées *embrayeurs* (traduit de l'anglais *shifters*). Les embrayeurs ont pour fonction d'articuler l'énoncé selon la situation d'énonciation. Sous cette appellation sont classées toutes les unités linguistiques qui « *délimitent l'instance spatiale et temporelle coextensive et contemporaine de la présente instance de discours contenant je* » - p.253, (Benveniste, 1966). Un embrayeur est toujours en relation avec l'événement décrit dans le discours. Ainsi, lorsque le discours devient récit, des termes correspondants, qui ne sont plus des embrayeurs, sont utilisés. Par exemple, dans la transposition du discours direct en discours indirect, *je* devient *il*, *hier* devient *la veille*, etc. Les embrayeurs renvoient, soit :

- à l'énonciateur ou au(x) destinataire(s). Ils sont appelés **embrayeurs de personnes**.
- aux circonstances de l'énonciation, circonstances de lieu, ou circonstances de temps. Ils sont appelés **embrayeurs spatio-temporels**.

3.1.2.1 Les embrayeurs de personnes

La présence de l'énonciateur ou du destinataire peut se manifester par :

- la **terminaison des verbes** conjugués aux premières et deuxième personnes du présent et du pluriel (1), (2), (4) ;
- les **pronoms personnels** de la première personne (*je, me, moi, nous*, etc.) et de la deuxième personne (*tu, te, toi, vous*, etc.) (3), (4) ;
- les **pronoms possessifs** et les **adjectifs possessifs** de la première personne (*mon, ma, mes, notre, mien*, etc.) et ceux de la deuxième personne (*ton, tes, votre, tien*, etc.) (3).

- (1) Courez voir Slumdog Millionnaire !
- (2) Laissez Raymond faire son travail et soyez patient.
- (3) Mais ce n'est que mon avis qui n'engage que moi.
- (4) Comme vous pouvez vous en douter, une école de commerce coûte cher.

La double présence de l'énonciateur et du destinataire est également visible à travers :

- le **mode impératif**, qui permet d'exprimer l'ordre, la défense, le conseil ou la prière de l'énonciateur envers un ou des destinataires. Signe que ce mode implique nécessairement l'énonciateur et le destinataire, il n'y a que trois personnes de conjugaison à ce mode : la 2^{ème} du singulier et la 1^{ère} et 2^{ème} du pluriel (1), (2) ;
- l'**apostrophe**, qui permet d'interpeller un destinataire dans le cours d'un texte de façon rhétorique ou non. À l'écrit, cet emploi est marqué spécifiquement par le recours à la virgule. Les mots mis en apostrophe sont suivis d'une virgule lorsqu'ils débutent un énoncé, sont isolés entre deux virgules quand ils sont au centre de la phrase, ou enfin précédés d'une virgule lorsqu'ils

4. Kerbrat-Orecchioni (1980) emploie le terme de *subjectivème*.

la terminent. La virgule signale également le changement de tonalité et d'intonation qui marque l'apostrophe à l'oral (5) ;

- les phrases **interrogatives directes** qui permettent à l'énonciateur de demander une information au destinataire, et attendre ou non (dans le cas d'une question rhétorique) une réponse de la part de celui-ci (6) ;
- les phrases **exclamatives**, qui indiquent que l'énonciateur exprime ses sentiments et ses émotions. Elle contient explicitement ou implicitement une forte affectivité (joie, colère, surprise, effroi, enthousiasme, amour, haine...) (7).

(5) Veuillez recevoir, **Monsieur**, mes plus cordiales salutations.

(6) «Il faut que nos joueurs se ressaisissent», avait-elle lancé dimanche soir sur TF1. Mais n'est-ce pas trop tard ?

(7) La bêtise est au pouvoir ! Une abyssale, insondable, indécrottable, insupportable bêtise !

3.1.2.2 Les embrayeurs spatio-temporels

Les embrayeurs spatiaux font référence à l'énonciateur à travers les circonstances de lieu de l'énonciation. Il s'agit :

- de certains **adverbes de lieu** (*ici, là-bas, là-haut, au-dessus, derrière, etc.*) (8) ;
- de **compléments circonstanciels de lieu** (*à gauche, près de moi, etc.*) (9) ;
- des **présentatifs** *voici* et *voilà* (10).

(8) **Ici**, il n'y a aucune certitude.

(9) Les voiles **au loin**, descendant vers Harfleur - *Victor Hugo*

(10) **Voici** le nouvel iPhone.

Pour faire référence aux circonstances de temps de l'énonciation (le moment où le *je* parle), le français permet l'utilisation :

- d'**adverbes de temps** (*maintenant, hier, demain, actuellement, etc.*) ;
- et **compléments circonstanciels de temps** (*En ce moment, lundi prochain, le mois dernier, depuis une semaine, dans trois jours, etc.*) (11).

(11) **Demain**, vous voterez l'abolition de la peine de mort. - Discours de *Robert Badinter*, le 17 septembre 1981

Selon Maingueneau (1994) : « *Les embrayeurs temporels les plus essentiels sont les marques de temps attachées aux verbes* ». En effet, le présent indique qu'il y a coïncidence entre les événements dont *je* parle et le moment de l'énonciation. Est passé ce qui s'est passé avant ce moment de l'énonciation. Est futur ce qui est postérieur à ce moment de l'énonciation (12). Maingueneau rappelle que « *le passé simple en français contemporain n'indique pas le passé mais l'absence de relation avec la situation d'énonciation* ». Toutefois, l'utilisation en déclin du passé simple est souvent remplacée dans les faits par le passé composé.

(12) Je **serai** devant la porte du magasin de DVD dès l'aube.

Les embrayeurs de personnes et spatio-temporels caractérisent un premier niveau de subjectivité qui révèlent la présence explicite de l'énonciateur dans son énoncé. La présence de l'énonciateur peut également être dévoilée à travers les évaluations et les émotions exprimées dans l'énoncé. Celles-ci correspondent à un deuxième niveau de subjectivité que nous présentons ci-dessous.

3.1.3 Un 2^{ème} niveau de subjectivité : les traces d'une évaluation

La subjectivité décrite jusqu'ici, fondée sur l'expression de la personne révélée par le langage, peut être qualifiée de subjectivité au premier degré. Selon Kerbrat-Orecchioni (1980), il existe une deuxième forme de subjectivité dite évaluative et affective. Cette forme de subjectivité dénonce la réaction émotionnelle du sujet qui s'avoue implicitement ou explicitement comme source d'interprétation et d'évaluation d'un objet référent qu'il est censé décrire. Elle se manifeste par la sélection de modalisateurs⁵ et d'unités linguistiques évaluatives organisant le discours en termes de jugement de valeur, d'adhésion ou de rejet de la part du sujet. Kerbrat-Orecchioni (1980) regroupe ces unités sous l'intitulé de **subjectivèmes**. Cette catégorie se décompose elle-même en deux sous-catégories : les subjectivèmes *affectifs* et les subjectivèmes *évaluatifs*. Enfin, les subjectivèmes évaluatifs peuvent être axiologiques ou non axiologiques. Avant de définir plus précisément cette catégorie d'indices de subjectivité et d'en donner une liste d'exemples, Kerbrat-Orecchioni (1980) prend plusieurs précautions oratoires que nous adoptons également en les synthétisant ci-dessous :

- **toute unité lexicale est, en un sens, subjective.** La linguistique démontre que les « mots » sont tous des symboles interprétatifs des « choses » et que, de fait, la production d'un discours est le résultat d'un découpage particulier du monde référentiel, choisi, organisé selon les contraintes établies par la langue. Ainsi, « *aucun individu n'est libre de décrire la nature avec une impartialité absolue, mais contraint au contraire à certains modes d'interprétation alors même qu'il se croit le plus libre.* » – Lee Whorf (1969). Autrement dit, le simple fait de nommer passe par les filtres de la perception, de l'interprétation et de la catégorisation. Dans les exemples (13) (14) (15), le même référent est ainsi nommé à travers différents filtres subjectifs ;

(13) Ce **chien** aboie toute la journée.

(14) Ce **chiot** aboie toute la journée.

(15) Ce **clébard** aboie toute la journée.

- les « **unités lexicales ne sont pas intrinsèquement toutes égales** » précise toutefois Kerbrat-Orecchioni (1980). Elle ajoute que « *l'axe d'opposition objectif/subjectif n'est pas dichotomique, mais graduel. Les unités lexicales sont en elles-mêmes (en langue) chargées d'une dose plus ou moins forte de subjectivité* ». Dans la même idée Todorov (1970) montre ainsi que « *bon implique l'énonciateur d'une manière plus forte que jaune* » – p.7 (Fig. 3.2).



FIGURE 3.2 : Échelle de subjectivité intrinsèque aux unités lexicales – (Kerbrat-Orecchioni, 1980)

En tenant compte de ces précautions oratoires, toute la complexité consiste à faire l'inventaire des unités linguistiques signifiantes en leur affectant un indice de subjectivité. Certains de ces subjectivèmes sont dits **axiologiques**, notion que nous définissons dans le paragraphe suivant.

5. Notion définie dans la section 3.2

3.1.3.1 Définition de l'axiologie

Le terme **axiologie** (du grec *axios* signifie *ce qui a de la valeur*) renvoie à l'idée d'une préférence au travers des champs d'expériences humaines de :

- la morale ou l'éthique : correspondant à l'axe du *bien* vs. *mal*,
- la pragmatique : axe *utile* vs. *inutile* ou *important* vs. *dérisoire* ou *efficace* vs. *inefficace*,
- le cognitif ou l'intellectuel : axe *intéressant* vs. *inintéressant* ou *intelligent* vs. *idiot*,
- l'esthétique : axe *beau* vs. *laid* ou *sublime* vs. *banal*,
- l'hédonique ou l'affectif : axe *agréable* vs. *désagréable* ou *plaisir* vs. *souffrance*.

Il s'agit alors, selon les cas, de valoriser ou de dévaloriser à travers notamment quatre grandes catégories de marqueurs axiologiques :

- les **marqueurs lexicaux** : *imbécile* ou *génie*, *superbe* ou *affreux*, en liaison parfois avec les registres de langue (*demeuré* ou *baraqué*).
- les **suffixes**, généralement péjoratifs : *traîner* devient *traînasser* ou *traînailleur*, *parler* devient *parloter*, *jaune* devient *jaunâtre* ou *jaunasse*.
- des **termes objectifs** qui, en fonction du contexte discursif, se chargent de significations axiologiques diverses. Le verbe *parler* peut ainsi être valorisant dans un énoncé comme : *Il a parlé. Nous savons ce que nous devons faire*, ou dévalorisant dans une phrase comme : *Tu parles, tu parles, c'est tout ce que tu sais faire*.
- les **figures de style**, et en particulier la métaphore et la comparaison, qui peuvent se charger d'une forte valeur axiologique. *Il s'est battu comme un lion* s'oppose à *Il s'est enfui comme un rat*, ou *C'est une véritable vipère* s'oppose à *C'est véritablement un aigle*.

La connotation positive ou négative de ces marqueurs peut être déterminée par l'usage et est pas conséquent relative à des codifications sociales. Cet aspect rend complexe l'idée de structurer un vocabulaire selon son axiologie. Nous précisons ce point dans le paragraphe suivant.

3.1.3.2 Une structuration axiologique complexe au niveau lexical

Les dictionnaires notent que certains mots sont péjoratifs (*avare*), d'autres, au contraire, mélioratifs (*généreux*). Ainsi, selon Pupier (1998), il y a des mots de valeur intrinsèquement positive (*généreux*, *délicieux*) et d'autres de valeur intrinsèquement négative (*avare*, *mauvais*). D'autres mots semblent en revanche neutres : *table*⁶ est l'exemple classiquement donné par les linguistes.

Néanmoins, à côté de mots intrinsèquement positifs ou négatifs, il en existe une partie dont l'orientation dépend des préférences, voire de l'idéologie, de l'énonciateur. Kerbrat-Orecchioni (1980) donne ainsi les exemples des mots tels que *communisme*, *nationalisme*, *ordre*, *discipline*, *aristocrate* qui sont entièrement solidaires, en ce qui concerne leur connotation axiologique, de la spécificité du lieu idéologique d'où parle l'énonciateur. Les termes du vocabulaire politique sont d'ailleurs très sensibles à cette notion d'idéologie. Ainsi le mot *libéral*, quand il a rapport à l'appartenance à un parti politique ne semble pas avoir d'axiologie propre. Pourtant, ce mot, quand il s'oppose à *autoritaire*, a une connotation positive. Au contraire, il enclenche un stéréotype négatif dans un discours politique « de gauche ». On retrouve le même phénomène avec le mot *bourgeois* qui est fondé sur une sémantique neutre. Or, quand il s'agit de préjugé ou d'opinion, ce qui est *bourgeois* est souvent mal vu (16).

(16) C'est juste un petit **bourgeois** qui joue au révolutionnaire.

6. Bien que Pupier (1998) précise que la neutralité du mot *table* n'empêche pas que, même ce mot pourrait évoquer, pour telle ou telle personne, des choses agréables (un festin) ou désagréables (pour qui ferait une association mentale avec une table d'opération par exemple).

Malgré tout, il y a des effets de domination idéologique qui imposent telle ou telle axiologie à une unité lexicale donnée. Ainsi se sont créés des tabous terminologiques sur des mots comme *nazi*, *fasciste* et *raciste*, lesquels ne pourraient être revendiqués par un groupe qu'à des fins provocatrices. Dans cette catégorie, Kerbrat-Orecchioni (1980) cite les exemples suivants : *poujadisme*, *réformisme*, *électoratisme*, *racisme*, *sexisme*, *jeunisme*, etc. qui fonctionnent régulièrement comme des termes injurieux et peuvent donc être considérés comme marqués en langue. En résumé, si certaines unités lexicales ont une orientation axiologique stable, il faut aussi tenir compte de la dimension pragmatique et idéologique des usages pour circonscrire le vocabulaire évaluatif.

Comme nous venons de l'observer, dans la langue, l'axiologie est structurée au niveau lexical. Elle peut être également perceptible au niveau de la morphologie dérivationnelle des mots. Nous précisons cet aspect dans le paragraphe suivant.

3.1.3.3 Quelques caractéristiques morphologiques de l'axiologie

Selon les travaux de Pupier (1998), dans le français, l'axiologie est marquée par certains suffixes et certains préfixes. Pupier (1998) précise qu'il ne semble pas exister, en français, de suffixe ou de préfixe mélioratif, il s'agit principalement d'un processus de dévalorisation.

Les suffixes péjoratifs Il arrive que le trait évaluatif de certains subjectivèmes reçoive un support signifiant spécifique ; c'est ainsi le cas des termes péjoratifs possédant un des suffixes exprimants le mépris :

- en *-ard* ou *-arde* : *chauffard*, *vantard*, *fuyard*, *flemmard*, *cossard*, *fêtard*, *froussard*, *trouillard*, *revanchard*, *communard*, *ringard*, etc.
- en *-asse* : *vinasse*, *blondasse*, *pêtasse*, *bouillasse*, *flemasse*, *paperasse*, *crasse*, *mélasse*, etc. Il existe néanmoins quelques exceptions : *crevasse*, *cuirasse*, *terrasse*.
- en *-âtre* : *jaunâtre*, *bellâtre*, *marâtre*, *acariâtre*, etc.

Notons au passage la forte production de mots dérivés par analogie, en ajoutant ce type de suffixes à des mots neutres (*politicard*, *facebookard*, *soixante-huitard*, etc.) (17).

- (17) C'est comme les **politicards** : ca mange à tous les rateliers pour récolter des voix.

Par ailleurs, les noms en *-aille* sont neutres lorsqu'ils désignent des objets inanimés ou des êtres animés non humains (*brousaille*, *ferraille*, *grenaille*, *limaille*, *muraille*, *volaille*, etc.) et sont péjoratifs quand ils désignent des humains (*canaille*, *gueusaille*, *racaille*, *truandaille*, etc.). Enfin, les suffixes en *-âtre* sont péjoratifs, mais peuvent également être subjectif et faiblement péjoratifs pour exprimer une atténuation : *beigeâtre*, *blanchâtre*, *bleuâtre*, *brunâtre*, *grisâtre*, *jaunâtre*, *noirâtre*, *olivâtre*, *rosâtre*, *rougeâtre*, *roussâtre*, *verdâtre*, *violâtre*.

Les préfixes péjoratifs Parmi les préfixes du français, le préfixe *mé(s)-* apporte une connotation négative sur le radical. Par exemple : *méconnaître*, *mécréant*, *médire*, *médisant*, *mépriser*, *méprisant*, *mésentente*. Le préfixe *Mal-* fonctionne de la même façon. Plus nettement que pour *mes-*, on peut distinguer les radicaux auxquels il s'attache. Ces radicaux sont :

- soit neutres : *malfaçon*, *malfaisant*, *malformation*, *malheur*, *malodorant*, etc.
- soit positifs : *maladroit*, *malavisé*, *malhabile*, *malhonnête*, *malpoli*, *malsain*, etc.

Le préfixe *In-* (et ses variantes : *Im-*, *Il-*, *Ir-*) est le plus productif des préfixes dans le vocabulaire subjectif. Il fonctionne néanmoins différemment que les deux préfixes cités précédemment puisqu'il a plutôt tendance à inverser l'axiologie originale du radical. Par exemple :

- **radical négatif**, préfixal positif : *impartial, irréfutable, invulnérable, infaillible, incassable, increvable, indéchirable, indéréglable, indestructible, infroissable, inviolable, etc.*
- **radical positif**, préfixal négatif : *injouable, intolérable, invivable, insupportable, insatisfaisant, insuffisant, etc.*
- **radical neutre**, préfixal neutre : *immobile, impair, indirect, infini, inhabité, inexploré, etc.*

Ce préfixe a d'ailleurs été tellement productif qu'un certain nombre de mots en *In-* n'ont pas ou plus de mot simple correspondant, à l'image de : *inconvenant, indécis, indispensable, etc.*

Dans un autre registre, *pseudo-* (du grec *pseudēs* : *faux*) est un autre préfixe à orientation axiologique exclusivement négative. On ne peut pas l'accoler à un mot axiologiquement négatif. En revanche, il fonctionne bien avec des mots valorisés positivement (*pseudo héros, pseudo réussite*) ou neutre (*une pseudo science, pseudo médecine*) (18). Le préfixe *faux-* fonctionne de la même façon (*faux frère, faux derche, faux papier, etc.*).

- (18) [...] n'est pas une école d'ingénieur. Tu vas payer une fortune tes études pour un **pseudo diplômé**.

À l'image du morphème *pseudo-*, certaines constructions lexico-grammaticales fonctionnent comme des absorbants axiologiques négatifs. Par exemple, la structure [espèce + de + GN] impose une connotation négative au groupe nominal (GN) et rend compte de l'idéologie de l'énonciateur (20).

- (19) « **Espèce de** bachibouzouk » – *Le capitaine Haddock*, Tintin
 (20) « **Espèce de** remplaçant de prof de chimie » – *Robin Williams*, Hook

Nous renvoyons aux travaux de Pupier (1998) pour davantage d'analyses sur les dérivations morphologiques qui impliquent une connotation axiologique. Au delà de la morphologie et du lexique, l'axiologie d'une évaluation peut être perturbée par la structure grammaticale de l'énoncé. Nous évoquons plus précisément la complexité des structures de négation dans la partie suivante.

3.1.3.4 Effets des structures grammaticales de négation sur l'axiologie

Les **tournures négatives** sont un aspect grammatical central pour l'analyse de la subjectivité. En effet, une forme négative inversera la plupart du temps la valeur axiologique d'une évaluation. Notons tout de même les réserves posées par Pupier (1998) et Ducrot (1973) au sujet de l'asymétrie axiologique causée par la négation. Ainsi, si les énoncés de l'exemple (21) sont assez proches sémantiquement, dans l'exemple (22) les deux propositions ont un sens beaucoup plus éloigné.

- (21) Pierre **n'est pas** gentil \approx Pierre est méchant
 (22) Pierre est **pas** méchant \neq Pierre est gentil

Les formes négatives varient dans leur complexité. En langue écrite, l'adverbe de négation *ne* est généralement accompagné d'un des mots suivants : *pas, plus, point, jamais, aucun, aucunement, nul, nullement, personne, rien, etc.* En revanche, dans la langue courante, on emploie généralement *pas* (ou un autre mot de négation) sans *ne* (22). L'adverbe *ne* peut également être employé seul dans certaines expressions ou avec les verbes *avoir, savoir, pouvoir, ou cesser* par exemple (23).

- (23) Je **ne** peux me résoudre à acheter le Monde.

Par ailleurs, l'usage d'une forme négative avec un verbe à l'infinitif varie l'ordre syntaxique des mots (24), les adverbes *sans* ou *aucun* peuvent également être utilisés à la place du *ne* (25). Enfin, la structure *ni ... ni* est une autre forme négative régulière (25).

- (24) Vous nous donnez 4 raisons de **ne pas** aller voir ce film

- **explicite** : si l'énonciateur exprime une évaluation tout en laissant des traces de sa position ;
- **implicite** : si l'énonciateur exprime une évaluation en cachant sa présence.

Dans une optique de traitement automatique, les marqueurs de subjectivité de 2^{ème} niveau étant plus instables et plus nombreux, les évaluations de type implicite semblent plus délicates à détecter.

Dans la section suivante, nous nous intéressons aux modalités énonciatives de l'évaluation. Celles-ci rendent compte des possibilités de la langue pour exprimer la subjectivité de deuxième niveau. Ces modalités ont été étudiées notamment par Charaudeau (1992) pour le français. Avec la configuration d'énonciation, elles permettent de caractériser un autre aspect important de la subjectivité.

3.2 Différentes modalités énonciatives pour exprimer une évaluation

Dans la continuité des travaux de Kerbrat-Orecchioni (1980), Charaudeau (1992) caractérise différentes modalités énonciatives qui servent à exprimer une évaluation dans le français (subjectivité de deuxième niveau). Après avoir défini la notion de modalité (§. 3.2.1), une synthèse de son travail nous permet de distinguer deux types d'évaluations :

- les évaluations *logiques*, qui évaluent notamment si une chose est vraie ou fausse, réelle ou irréelle, etc. (§. 3.2.2) ;
- et les évaluations *axiologiques*, qui évaluent par exemple si une chose est bonne ou mauvaise, belle ou laide, agréable ou douloureuse, etc. (§. 3.2.3).

La présence et le rôle différent de ces évaluations est une deuxième caractéristique qu'il nous semble intéressant de prendre en compte dans nos traitements automatiques.

3.2.1 Définition de la modalité

Bally (1932) introduit pour la première fois en linguistique le terme de modalité. La modalité désigne l'attitude particulière de l'énonciateur dans l'activité d'énonciation. En effet, les linguistes soutiennent que l'énonciation correspond à la communication d'une pensée distincte d'une simple représentation du monde. Dans ce cadre, l'énonciateur participe activement à cette expression. D'après Bally (1932), énoncer « *c'est juger qu'une chose est ou n'est pas, ou estimer qu'elle est désirable ou indésirable, ou enfin désirer qu'elle ne soit ou ne soit pas. On croit qu'il pleut ou on ne le croit pas, ou on en doute, on se réjouit qu'il pleuve ou on le regrette, on souhaite qu'il pleuve ou qu'il ne pleuve pas* ». Selon cette définition, Bally décrit que tout énoncé est la combinaison de deux éléments :

- un *dictum* : le contenu représenté ;
- un *modus* : l'opérateur de modalité.

Dans (31), l'énonciateur inclut le dictum (*il + être + à l'heure*) dans une modalité de *volonté*. Dans (32), l'énonciateur active une modalité de *connaissance*, alors que dans (33) il s'agit d'une modalité d'*appréciation*.

(31) **Je veux** qu'il soit à l'heure.

(32) **C'est certain**, il sera à l'heure.

(33) **J'aime** quand il est à l'heure.

Charaudeau (1992) propose une liste de modalités énonciatives pour le français. Nous synthétisons son travail en nous intéressant aux modalités spécifiques à l'expression d'une évaluation. Nous retenons quatre modalités : l'opinion, l'accord/désaccord (modalités d'évaluations logiques), le jugement et l'appréciation (modalités d'évaluations axiologiques).

3.2.2 Les modalités logiques

Les modalités d'opinion et d'accord/désaccord sont les premières modalités qui permettent d'exprimer une évaluation selon Charaudeau (1992). Notons que ces modalités n'activent pas d'axiologie positive ou négative en elles-mêmes. En revanche, elles peuvent influencer sur (voire inverser) l'axiologie en évaluant un propos sur un axe vrai/faux (ou réel/irréel). Par exemple, dans (34), l'évaluation logique *je ne pense pas*, annule voire inverse l'axiologie positive du mot *légitime*.

(34) *je ne pense pas* que Domenech soit le sélectionneur légitime de cette équipe de France.

3.2.2.1 Opinion

L'opinion est la première modalité qui peut intervenir pour exprimer une évaluation. Charaudeau (1992) en propose la définition suivante :

Définition 4 (Opinion) *Il est présupposé un fait (ou une information) à propos duquel l'énonciateur explicite la place que celui-ci (ou celle-ci) occupe dans son univers de croyance. L'énonciateur évalue donc la vérité de son propos et révèle du même coup ce qu'est son point de vue. Ce point de vue est d'ordre intellectif, c'est à dire que, tout en gardant le propos à distance, l'énonciateur exprime une attitude de croyance plus ou moins certaine qui relève de la raison (je pense, je crois, je doute).*

L'opinion active un univers de croyance dans lequel l'énonciateur se positionne. Elle peut se spécifier en deux attitudes de croyance : l'une de **conviction**, l'autre de **supposition**. La conviction implique qu'il y a un doute sur la vérité du propos ; et c'est précisément parce qu'il existe un doute possible que l'énonciateur éprouve le besoin d'exprimer la certitude qu'il a sur la vérité du propos. Cette certitude est totale, mais elle lui est propre (35). La supposition implique qu'il n'y ait pas de certitude totale sur la vérité du propos. Dès lors, l'énonciateur peut être amené à exprimer son point de vue quant à son propre degré de certitude. On peut donc dire que la supposition varie sur un axe qui va d'une *certitude forte* (mais jamais totale) à un *pressentiment* (36)

(35) *J'ai la conviction* qu'il réussira.

(36) *J'imagine* qu'il n'est pas près de recommencer après ce qui lui est arrivé.

Dans les travaux de Charaudeau (1992), la notion de configuration d'énonciation est également présente. Il précise ainsi que l'énonciateur peut exprimer une opinion en étant **explicitement** présent dans son énoncé (37) ou dissimuler sa présence (**implicite**) (38).

(37) *Je suis persuadé* qu'on ne peut rien y faire.

(38) *C'est sûr*, on ne peut rien y faire.

3.2.2.2 Accord/Désaccord

La modalité d'accord/désaccord est relativement proche de celle de l'opinion. Charaudeau (1992) propose la définition suivante :

Définition 5 (Accord/Désaccord) *L'accord ou le désaccord présuppose qu'il a été adressé à l'énonciateur une demande de dire s'il adhère ou non à la vérité d'un propos ou d'un acte tenu par un autre (que cette demande ait été faite réellement ou non). L'énonciateur répond en exprimant qu'il adhère ou non au propos ou à l'acte tenu. Du même coup, il contribue à la validation de la vérité de celui-ci. Entre l'accord et le désaccord des nuances peuvent être spécifiées : accord total, accord approximatif, rectificatif.*

Par rapport à l'opinion, l'énonciateur se positionne ici par rapport à un univers de croyance déjà existant. Il a la possibilité d'aller dans le sens de cette croyance (39) ou de s'y opposer (40).

(39) **Oui**, certainement.

(40) **Je ne suis pas d'accord** avec ceux qui disent que tout est beau à l'ENSAM.

Outre *Oui/Non*, et *Je suis d'accord/pas d'accord avec*, divers mots et locutions (*certes, bien sûr, bien entendu, tout à fait, absolument pas, pas du tout, certainement pas, sûr que oui/non* ou encore *oui, mais...*) annoncent explicitement un accord ou un désaccord d'après Charaudeau (1992).

Les modalités d'évaluations logiques sont importantes à prendre en compte, en particulier si elles sont utilisées dans le contexte d'une évaluation axiologique. Elles peuvent en modifier l'intensité ou en modifier la polarité axiologique. Nous présentons ci-dessous deux modalités d'évaluations axiologiques proposées par Charaudeau (1992).

3.2.3 Les modalités axiologiques

Charaudeau (1992) définit deux modalités d'évaluations axiologiques pour le français : l'appréciation et le jugement. Celles-ci permettent à l'énonciateur d'adopter une attitude favorable (positive) ou défavorable (négative) par rapport à un propos, un fait ou un objet. Le vocabulaire associé à ces modalités est particulièrement vaste d'après Charaudeau.

3.2.3.1 Appréciation

Dans la classification de Charaudeau (1992) et en tenant compte des raffinements proposés par Galatanu (2000), l'appréciation est la modalité évaluative la plus subjective. L'appréciation regroupe toutes les attitudes où l'énonciateur se positionne dans les champs de valeurs du désir, de l'hédonique, de l'esthétique et du pragmatique. Charaudeau (1992) propose la définition suivante :

Définition 6 (Appréciation) *Il est présupposé un fait à propos duquel l'énonciateur dit quel est son sentiment. L'énonciateur évalue donc, non plus la vérité du propos, mais sa valeur, en révélant ses propres préférences. Cette évaluation est nécessairement polarisée. L'appréciation peut porter sur différents domaines de valeurs :*

- le **pragmatique**, qui qualifie positivement ou négativement le propos, au regard des valeurs utilitaires propres au locuteur : "je trouve utile/inutile que ..."
- l'**esthétique**, qui qualifie positivement ou négativement le propos, au regard des valeurs de l'esthétique propre au locuteur : "je trouve beau/laid que ..."
- l'**hédonique**, qui qualifie positivement ou négativement le propos, au regard des valeurs émotionnelles propres au locuteur : "je trouve heureux/malheureux que ..."

À l'image des modalités précédentes, l'énonciateur peut s'inclure explicitement (41) ou implicitement dans son propos (42).

(41) **Je suis content** que vous soyez là.

(42) Ce qui arrive est **un grand malheur**.

3.2.3.2 Jugement

La modalité de jugement définie par Charaudeau (1992) concerne un autre type d'évaluation axiologique. Si celle-ci est subjective, elle l'est moins que la modalité d'appréciation d'après Galatanu (2000) car le jugement se positionne dans les champs de valeurs de l'éthique, de la morale et de l'intellect. Selon Galatanu (2000), le jugement implique donc un peu moins l'énonciateur que l'appréciation car il se positionne également en référence à des normes externes (loi, textes religieux ou philosophiques, connaissances scientifiques, etc.). Charaudeau (1992) propose la définition suivante du jugement :

Définition 7 (Jugement) *Il est présupposé un fait à propos duquel le locuteur juge si ce fait est bon ou mauvais en déclarant son approbation ou réprobation. Le locuteur se donne l'autorité morale de celui qui peut juger. Par rapport à l'appréciation, la modalité de jugement porte sur les domaines de valeurs suivants :*

- *l'éthique ou la morale, qui qualifie le fait positivement ou négativement, au regard des valeurs de l'éthique ou de la morale propre au locuteur : je trouve bien/mal que ...,*
- *l'intellect, qui qualifie le fait positivement ou négativement, au regard des valeurs de l'intelligence propre au locuteur : je trouve intéressant/inintéressant que...*

L'énonciateur peut juger en se positionnant explicitement comme la source du jugement (43), ou au contraire masquer sa présence pour renforcer l'effet d'objectivité (44).

(43) Je condamne fermement ces violences.

(44) Ce raisonnement est idiot.

Dans un énoncé, toutes ces modalités évaluatives peuvent être activées dans différentes phrases, mais elles peuvent également se combiner. La combinaison de ces modalités et ses effets sur le discours évaluatif et argumentatif ont été étudiés par Galatanu (2000).

3.2.4 Combinaison de modalités

La théorie de Galatanu (2000) sur l'évaluation complète le modèle de Charaudeau en hiérarchisant les modalités sur une échelle de subjectivité. Lorsqu'un énonciateur organise son énoncé, il peut choisir d'objectiver ou de subjectiver son discours en activant certaines modalités ou en adoptant une configuration énonciative explicite ou implicite. Par exemple, le mot *mentir* correspond à une valeur appartenant à la modalité de jugement (éthique). En le combinant à d'autres modalités, ce mot peut intervenir dans des stratégies évaluatives ou argumentatives différentes. Dans (45), l'axiologie *a priori* négative de *mentir* est inversée par l'appréciation explicite positive qui l'englobe.

(45) J'aime sa façon de mentir.

Dans (47), la valeur de *mentir* est atténuée ou annulée par la modalité d'opinion (supposition faible). Dans (47), la modalité de jugement est intensifiée en étant englobée par une autre modalité de jugement explicite (*nous condamnons*). Dans un (48), l'évaluation *il ment* est également renforcée par une modalité de conviction (*il est évident*). À l'image de (49), la configuration implicite du jugement (énonciateur masqué) augmente sa force argumentative et s'impose presque aux destinataires comme l'énonciation d'une vérité.

(46) je doute qu'il mente.

(47) Nous condamnons ses mensonges.

(48) Il est évident qu'il ment.

(49) C'est un menteur.

Il ressort des travaux de Galatanu qu'il est nécessaire de distinguer ses différentes modalités de l'évaluation pour mieux recombinaer le sens évaluatif d'un énoncé. C'est dans cet optique que nous positionnons nos travaux d'analyses automatiques de la subjectivité.

3.3 Vers un modèle informatique du langage de l'évaluation

Notre proposition Nous proposons un modèle adapté des travaux en linguistique sur la subjectivité et l'évaluation (Kerbrat-Orecchioni, 1980 ; Charaudeau, 1992 ; Galatanu, 2000) pour nos traitements informatiques. Nous considérons que dans un énoncé, toute évaluation est un passage délimité par un caractère de début (**D**) et un caractère de fin (**F**). Ce passage correspond à une unique modalité d'évaluation (**M**). L'énonciateur peut exprimer son évaluation avec une configuration explicite ou implicite (**C**). Dans son contexte, l'évaluation peut exprimer une axiologie positive, négative, ambiguë ou neutre (**A**). En résumé, nous représentons chaque évaluation (**E**) par cinq caractéristiques principales :

$$E = \langle D, F, M, C, A \rangle$$

où chaque caractéristique prend comme valeur :

- D = position du caractère de début dans l'énoncé ;
- F = position du caractère de fin dans l'énoncé ;
- M = {opinion, accord-desaccord, jugement, appréciation} ;
- C = {explicite, implicite} ;
- A = {positif, négatif, ambigu, neutre} ;

Par ailleurs, nous nous intéressons à identifier l'objet sur lequel porte une évaluation. Nous employons le terme cible⁷ (**T**) pour parler de ce concept particulier. Chaque évaluation peut donc potentiellement avoir une sixième caractéristique qui correspond à la cible qu'elle évalue. Par exemple, dans l'énoncé (50), l'évaluation en bleue est modélisée comme suit : $\langle 1, 13, \text{appréciation, explicite, négatif, équipe de France} \rangle$.

(50) Je suis déçue par l'équipe de France.

Développer un corpus et un lexique de référence La perspective de développer des méthodes d'analyse automatique de l'évaluation à granularité fine nous amène à exprimer deux besoins principaux en ressources :

- un **corpus**, pour les phases d'entraînements et de tests de nos méthodes. Celui-ci nécessite de contenir des annotations de passages évaluatifs qui puissent servir de référence. Chacun de ces passages a également besoin d'être décrit selon les six caractéristiques de notre modèle. Un tel corpus ou un corpus approchant n'existe pas pour le français ;
- un **lexique** de mots ou d'expressions utilisés pour exprimer une évaluation. Pour chacune des entrées de ce lexique, des caractéristiques sur la modalité et l'axiologie qu'elles activent sont nécessaires.

Il n'existe pas de telles ressources pour le français. Le développement de deux ressources de ce type fait l'objet du chapitre suivant.

7. Terme anglais *Target*.

Développer des ressources pour le français

« *Data is the Oxygen of Science* » – Hans Rosling.

4.1	Un corpus annoté, multithématique et en français	52
4.1.1	Les blogs : un support riche en phénomènes évaluatifs	53
4.1.1.1	<i>Un rôle en rupture avec les médias « traditionnels »</i>	53
4.1.1.2	<i>Extraction de blogs thématiques personnels</i>	54
4.1.2	Annotation manuelle des objets textuels, des évaluations et de leur cible	55
4.1.2.1	<i>Les objets de l'énoncé</i>	55
4.1.2.2	<i>Les évaluations de l'énoncé</i>	56
4.1.2.2	<i>Accord inter-annotateurs : une tâche d'annotation complexe</i>	58
4.1.3	Le corpus Blogoscopie : résultats et statistiques	58
4.1.3.1	<i>Le corpus d'entraînement</i>	59
4.1.3.2	<i>Le corpus de test</i>	59
4.1.4	Travaux liés et discussion	60
4.2	Un lexique de mots évaluatifs pour le français	61
4.2.1	Du corpus annoté aux entrées lexicales	66
4.2.1.1	<i>Caractéristiques grammaticales</i>	62
4.2.1.2	<i>Caractéristiques sémantiques</i>	63
4.2.2	Le lexique de l'évaluation : résultats et statistiques	63
4.2.3	Travaux liés et discussion	65
4.3	Conclusion	66

En fouille d'opinions, et en particulier pour la problématique d'analyse à granularité fine des évaluations, deux types de ressources représentent un besoin critique pour développer et évaluer des méthodes : les corpus et les lexiques. Pour pallier le manque de ressources pour le français, nous avons développé dans le cadre du projet ANR Blogoscopie :

- le **corpus Blogoscopie** (§. 4.1). Celui-ci regroupe des billets de blogs d'utilisateurs distincts et les commentaires qui y sont associés. Il a été développé dans l'idée de représenter un large éventail de thématiques pour ne pas réduire la diversité du phénomène évaluatif étudié. Les différents objets évoqués dans chaque document, les passages évaluatifs et leur cible y ont été annotés manuellement. La description des évaluations suit le modèle présenté dans le chapitre 3 ;
- le **lexique de l'évaluation** (§. 4.2). Ce lexique a été développé manuellement en s'appuyant sur les annotations du corpus Blogoscopie. La structuration de ce lexique rend compte d'informations grammaticales et sémantiques s'appuyant également sur le modèle utilisé pour décrire les évaluations. La signification d'une entrée lexicale pouvant être potentiellement instable, la structure du lexique permet d'explicitier la nature de cette ambiguïté.

Nous précisons leur processus de création et nous comparons leurs caractéristiques à celles des ressources existantes du domaine. Le corpus Blogoscopie et le lexique de l'évaluation constituent des ressources essentielles pour nos travaux. Elles nous permettent d'aborder trois tâches hétérogènes liées à l'analyse à granularité fine des évaluations (chapitres 5, 6 et 7).

4.1 Un corpus annoté, multithématique et en français

Évolution des corpus en fouille d'opinions En fouille d'opinions, le choix des corpus, et en particulier leur disponibilité, est un facteur qui a fortement influencé la nature des travaux. Nous distinguons trois types de corpus :

- les corpus où les documents ont **une seule étiquette globale**. À l'origine de la fouille d'opinions, les textes de critiques ou les textes politiques sont des corpus disponibles « naturellement ». Dans ces corpus, chaque document possède une étiquette globale indiquant par exemple soit sa polarité axiologique, soit l'opinion politique de l'auteur. De nombreux corpus anglophones (TREC, NTCIR, etc.) ou francophones (DEFT'07, DEFT'09) existent dans cette catégorie pour aborder la problématique d'analyse d'« opinion » au grain du document (voir chapitre 2, §. 2.1) ;
- les corpus de documents où **les phrases sont étiquetées**. Dans ces corpus, une étiquette est attribuée à chaque phrase d'un document pour indiquer si elle est soit subjective ou objective, soit positive, négative ou neutre. Ces corpus, essentiellement anglophones, ont été développés manuellement ou à l'aide d'une amorce automatique pour pouvoir aborder cette problématique à plus grande échelle (Wiebe et Riloff, 2005) ;
- dans une troisième phase plus récente, quelques travaux ont développé des corpus de documents où **les évaluations et les cibles sont étiquetées**. Dans cette catégorie, nous connaissons deux corpus anglophones :
 - le corpus d'articles de presse MPQA (Wiebe et collab., 2006) où les évaluations sont délimitées et catégorisées selon un modèle théorique ;
 - le corpus de critiques de film de Zhuang et collab. (2006) où les passages évaluatifs sont uniquement reliés à leurs cibles.

Du point de vue de la disponibilité des corpus, la disparité est grande entre l'anglais et le français. Il n'existe aucun corpus en français correspondant à la troisième catégorie. La première motivation de notre travail consiste à pallier ce manque.

Développer un corpus multithématique pour le français Pour développer un corpus destiné à l'analyse à granularité fine des évaluations, nous nous intéressons en particulier aux textes issus de blogs. Contrairement aux articles de presse, aux critiques de films ou aux débats parlementaires, les blogs ont l'intérêt de ne pas être limités en genre. Ils couvrent un large panel de thématiques et de styles d'écritures. De ce point de vue, ils nous paraissent être un support textuel idéal pour prendre en compte la diversité des phénomènes évaluatifs de la langue.

Dans une première partie, nous motivons plus précisément le choix des blogs comme support textuel (§. 4.1.1). Nous présentons notre modèle d'annotation et la méthodologie appliquée (§. 4.1.2) pour développer le corpus Blogoscopie (§. 4.1.3). Finalement, nous donnons quelques points de comparaison avec le corpus MPQA (Wiebe et collab., 2006), son équivalent anglophone le plus proche (§. 4.1.4).

4.1.1 Les blogs : un support riche en phénomènes évaluatifs

Less blogs ont commencé à apparaître sur le web francophone au début de l'année 2003 et ont depuis fortement évolué. Ils se définissent désormais comme un nouveau média possédant un rôle et un style distinct des médias dits « traditionnels » (§. 4.1.1.1). Pour constituer notre corpus de travail, nous utilisons des textes fournis par la plateforme de blogs OverBlog¹. Plusieurs filtres permettent de retenir finalement 250 billets de blogs multithématiques et leurs commentaires associés (§. 4.1.1.2).

4.1.1.1 Un rôle en rupture avec les médias « traditionnels »

De façon synthétique, Fievet et Turrettini (2004) définissent un blog avant tout comme un format de publication : « [...] type de site web composé essentiellement d'actualités (ou billets) publiées au fil de l'eau et apparaissant selon un ordre antechronologique ». Chaque billet peut être commenté par les visiteurs ou l'auteur lui-même. Les billets et leurs commentaires sont archivés de sorte qu'il est possible de lire tout l'historique des publications d'un auteur.

Au delà de son format particulier, les blogs sont souvent le fait d'individus intervenant à titre personnel. C'est un point clé du phénomène des blogs. Beaucoup de blogueurs parlent à la première personne du singulier dans les textes qu'ils publient. Et contrairement aux médias traditionnels, la personnalité du blogueur, son style et ses opinions jouent un rôle essentiel dans l'intérêt de son blog. Par opposition, les médias traditionnels se définissent comme « une façon impersonnelle de faire circuler une information. Ils permettent de diffuser des informations vers un large nombre d'individus sans modifier, ni personnaliser l'information »². Le respect d'une déontologie qui conduit, en principe, à présenter des informations strictement objectives est remplacé dans la blogosphère par une liberté de ton, l'art de la critique, le fait de pouvoir ouvertement s'en prendre à ce qu'on n'aime pas, ou au contraire encenser ce que l'on adore.

Le blog est donc, généralement, le contraire de la recherche d'une information objective, mais constitue l'expression d'un parti pris affiché. La plupart des blogueurs ne cherchent pas à être objectif mais cette absence d'objectivité est remplacée par une forme de transparence qui est, elle, absente des médias traditionnels. Une grande quantité de blogs n'ont de raison d'être que par les commentaires qu'ils apportent sur l'actualité. S'ajoute à cela le fait qu'un nombre croissant de blogueurs dépassent leur rôle de commentateur pour empiéter sur celui d'analyste, de journaliste, voire d'enquêteur, et dans le même temps beaucoup de journalistes professionnels se mettent à « bloguer ». Il en résulte la présence d'une très grande variété d'expressions évaluatives, de la plus implicite à la plus explicite.

1. OverBlog - <http://www.over-blog.com/>

2. Wikipedia - <http://fr.wikipedia.org/wiki/Média>

4.1.1.2 Extraction de blogs thématiques personnels

Les blogs thématiques personnels Une analyse de la typologie des blogs nous permet de nous focaliser sur un type particulier de blogs : les *blogs personnels thématiques*. Ceux-ci sont plus intéressants pour notre travail car ils ont la particularité de traiter régulièrement d'un ou plusieurs concepts d'une même thématique (*le sport, la politique, la mode, les sorties culturels*, etc.). Le contenu de ces blogs présente simultanément deux parties :

- l'une relative à une actualité ;
- l'autre exprimant les évaluations de l'auteur (du billet ou d'un commentaire) sur cette actualité.

Ils s'opposent notamment aux purs *journaux intimes* qui ne présentent que l'actualité de l'auteur du blog ou aux *blogs de liens* qui ne contiennent que des liens hypertextes vers d'autres sites.

Quatre phases d'extraction Pour constituer notre corpus, des billets et des commentaires ont été extraits des blogs hébergés par la principale plateforme francophone : OverBlog. Par la suite, le terme *document* est utilisé pour parler d'un billet particulier et de ses commentaires. Dans un souci de représentativité, sur les 43 thématiques internes proposées par OverBlog à ces blogueurs pour classer leurs productions, nous en avons sélectionné 33 pour composer notre corpus. Ce premier corpus est scindé en quatre sous parties correspondantes à quatre stratégies d'extraction différentes :

- la première partie du corpus contient **76 billets** et **296 commentaires**. Ils ont été extraits pour représenter 33 des thématiques internes de la plateforme OverBlog (2 à 3 documents par thématique) : *actualité, artiste, blogzine, business, cinéma, collectionneur, consommation, console, croyance, décoration, détente, économie, enfants, familles, gastronomie, guides d'achats, internet, jeux, livres, loisirs, activités, maison, musique, philosophie, politique, religions, rêver, santé, sciences, sport, technologie, télévision, voyages* ;
- la deuxième partie du corpus est composée de **34 billets** et **162 commentaires**. Elle vise à renforcer 8 thématiques (4 documents par thématique) supposées contenir davantage d'expressions d'évaluations variées : *actualité, artiste, blogzine, business, cinéma, gastronomie, guide d'achats, livre* ;
- la troisième partie du corpus est composée de **90 billets** et **156 commentaires**. Ils ont été extraits à partir de 10 mots-clés correspondants à 10 sujets d'actualités (9 documents par sujet). Chaque document doit contenir au moins l'un des mots-clés suivants : *développement durable, beaujolais, grève SNCF, Harry Potter, le coeur des hommes 2, la loi LRU, nucléaire, Raymond Domenech, Vladimir Poutine* et la *Wii* ;
- la quatrième partie du corpus est composée de **50 billets** et **145 commentaires**. Ils ont été extraits à partir de 2 mots-clés (25 documents par mots-clés) : *sushi* et *Sarah Palin*. Cette quatrième partie vise en particulier à simuler un scénario applicatif réel. Elle est spécifiquement réservée aux tests (§. 4.1.3).

Constitution du corpus avant l'annotation manuelle Pour éviter d'avoir trop de documents liés aux élections présidentielles 2007 risquant de diminuer l'aspect multithématique, nous avons choisi de faire ces différentes phases d'extraction sur des documents publiés à des périodes comprises entre juin 2007 et juin 2008. Les documents ne contenant que des liens (ou peu de textes) ou très mal orthographiés ne sont pas considérés. Les documents issus d'un même auteur sont également filtrés.

Au final, 250 billets et 759 de leurs commentaires composent le corpus. Cela représente environ 115 000 mots répartis dans 250 documents. Après extraction de la plateforme de blogs et avant toute annotation manuelle, chaque document est composé d'une structuration XML minimale précisant notamment : *la date de publication du billet et des commentaires, le titre du billet, les mots-clés, l'url*

du blog, le pseudonyme des auteurs, etc. Un exemple de document issu de cette phase d'extraction est présenté dans l'annexe A.

4.1.2 Annotation manuelle des objets textuels, des évaluations et de leur cible

L'annotation du corpus Blogoscopie a démarré en juin 2007 au moment de la première phase d'extraction. Elle s'est poursuivie en continue pendant 18 mois pour annoter les documents issus des quatre phases d'extraction. Les blogs de la première phase d'extraction ont servi à tester un modèle d'annotation fixé par les participants du projet ANR Blogoscopie. Il a été quelque peu renforcé en tenant compte des cas d'accords et de désaccords rencontrés par trois annotateurs humains. Ce modèle d'annotation concerne deux catégories d'éléments :

- les **objets** réels ou abstraits présents dans un document (§. 4.1.2.1) ;
- les passages correspondant à des **évaluations** exprimées dans un document (§. 4.1.2.2).

Certains des objets d'un document peuvent être la *cible* d'une évaluation exprimée dans le même document. Ce lien sémantique particulier est pris en compte dans le modèle d'annotation. L'annotation de l'ensemble du corpus a été réalisée par une seule linguiste. Néanmoins, à partir d'un ensemble de documents de la première phase d'extraction, nous mesurons un accord inter-annotateurs pour évaluer la complexité de cette tâche (§. 4.1.2.3).

4.1.2.1 Les objets de l'énoncé

Le modèle d'annotation choisi suppose qu'un énoncé évoque deux types d'objets textuels :

- les concepts (**C**) : *le nucléaire, la politique, le cinéma, le rugby, etc.*
- les instances de concepts (**I**). Dans notre cadre, une instance de concepts est une entité nommée³. Ces entités nommées regroupent les noms de personnes physiques, morales, mythiques ou de fictions, les entités géographiques, etc : *EDF, Barack Obama, Catherine Deneuve, Raymond Domenech, etc.*

Les anaphores nominales Deux objets textuels peuvent faire référence au même objet du monde. Par exemple, (1) et (2) présentent l'annotation⁴ de deux passages d'un même document. Dans ces deux passages :

- l'instance I1 (*Berlin*) et le concept C3 (*gouvernement allemand*) sont un même objet réel ;
- l'instance I2 (*Sarkozy*) et l'instance I4 (*Nicolas Sarkozy*) sont un même objet réel ;
- le concept C1 (*accord*) et le concept C3 (*mémorandum*) sont un même objet réel.

(1) <I1>Berlin</I1> critique l'<C1>accord</C1> sur le <C2>nucléaire</C2>
<I2>Sarkozy</I2>-<I3>Kadhafi</I3>

(2) Le <C3>mémorandum</C3> sur le <C1>nucléaire</C1> signé par
<I4>Nicolas Sarkozy</I4> pendant sa visite à <I5>Tripoli</I5> a suscité jeudi
des protestations du <C3>gouvernement allemand</C3>

Pour représenter ces relations entre les objets textuels, un identifiant est associé à chaque concept. Les concepts renvoyant au même objet réel partagent le même identifiant. Les instances reliées à un concept possède également un attribut permettant d'indiquer à quels concepts elles font référence. Enfin, les anaphores pronominales n'ont quant à elle pas été annotées.

3. Telles qu'elles sont définies dans les conférences MUC (MUC, 1995) et par Fourour et Morin (2009).

4. Dans ce manuscrit, nous allégeons quelque peu l'annotation des exemples. Une description plus précise et plus technique des balises XML utilisées est disponible dans le manuel utilisateur du corpus.

Trois niveaux d'importance Au sein d'un document, le modèle d'annotation précise l'importance d'un objet selon trois niveaux. Chaque objet peut être dit :

- « concerné » (**C**), si le sujet général du document concerne principalement cet objet. Un même document peut avoir plusieurs objets « concernés » ;
- « associé » (**A**), si cet objet est une sous-partie ou une caractéristique d'un objet « concernés ». Un même objet concerné peut avoir plusieurs objets associés ;
- « non associé » (**N**), si cet objet ne concerne pas directement le sujet général du document.

Pour illustrer ces différents types d'objets, nous donnons plusieurs exemples issus du corpus. L'exemple (3) contient un concept concerné (**CC**) (*armement nucléaire*). Cet objet est un des concepts principaux du document. Il possède l'identifiant C1. Le concept C1 est rattaché à un concept associé (**CA**) (*le démentellement*).

(3) Je décrète le <CA cc="C1">démentellement</CA> de l'<CC id_c="C1">armement nucléaire</CC>.

Certains concepts ne sont pas associés au sujet général du document (**CN**). Ils sont néanmoins pris en compte (en particulier si un passage évaluatif porte sur ce concept⁵). Dans (4), le concept *histoire* ne fait pas parti des sujets principaux du document, mais l'évaluation *surréaliste* porte sur cet objet. Le principe est identique pour les instances non associés (**IN**) (5).

(4) Vous trouverez un compte-rendu détaillé de cette <CN>histoire</CN> **surréaliste** sur ce document.

(5) Meme <IN>Pierre Menes</IN>, ne comprend pas son attitude et ses choix et defend <IN>Trézeguet</IN> !!! **je commence a l'aimer** ce <IN>Pierre</IN> !

Tous les concepts ou les instances de concepts ne sont pas toutes les cibles d'une évaluation. Ils sont néanmoins annotés en vue d'analyses futures. Dans le paragraphe suivant, nous présentons le modèle adopté pour annoter les évaluations.

4.1.2.2 Les évaluations de l'énoncé

Le modèle d'annotation choisi pour décrire les évaluations est issu de la théorie de Charaudeau sur les modalités de l'évaluation (voir chapitre 3, §. 3.2). Nous distinguons les modalités logiques et axiologiques de l'évaluation. Seules les évaluations axiologiques sont reliées à un objet particulier sur lequel porte l'évaluation. Cet objet particulier est appelé *cible*.

Évaluations de modalité axiologique et relation avec leur « cible » Le modèle d'annotation tient compte de deux modalités axiologiques : les appréciations (**A**) et les jugements (**J**). Pour chaque évaluation axiologique, il précise également :

- sa configuration d'énonciation : implicite (**I**), explicite (**E**), voire explicite et exclamatif simultanément (**EE**) ;
- son axiologie : favorable (**F**) ou défavorable (**D**) ;
- la présence d'ironie : oui ou non ;
- et la forme textuelle de son objet cible ;

5. Nous précisons l'annotation des évaluations dans la section suivante (§. 4.1.2.2).

Dans l'exemple (6), les passages évaluatifs *pose un énorme problème* et *irresponsable* sont considérés comme des évaluations de modalité *appréciation*, de configuration *implicite* et d'axiologie négative ou *défavorable* (AID).

- (6) Cet accord **<Appreciation type="AID" forme="accord">**pose un énorme problème**</Appreciation>** de prolifération nucléaire et se situe dans la droite ligne de la politique française d'exportation **<Appreciation type="AID" forme="politique française d'exportation">**irresponsable**</Appreciation>** de sa technologie nucléaire.

L'exemple (7) illustre l'annotation d'une évaluation de type *jugement explicite exclamatif favorable* (JEEF).

- (7) C'est cela la loi Pécresse!!! Cette loi s'inscrivant parfaitement dans le projet des jeunes populaires et de Nicolas Sarkozy, **<Jugement type="JEEF" forme="loi">**nous la soutenons!!!**</Jugement>**

Par ailleurs, la modèle d'annotation permet d'indiquer l'objet évalué par chaque évaluation axiologique. Dans (7), l'évaluation *nous la soutenons!!!* porte sur un objet du monde réel. Cet objet est référé dans le texte par plusieurs objets textuels : les groupes nominaux *loi Pécresse*, *loi* et l'anaphore pronominale *la*. Selon la méthodologie adoptée, l'anaphore nominale la plus proche du passage évaluatif est choisie comme cible (ici, le mot *loi*). Dans (6), la cible de l'évaluation *pose un problème* est un objet référé par le mot *accord*. La cible de l'évaluation *irresponsable* est un objet référé par l'expression *politique française d'exportation*.

Évaluations de modalité logique Le modèle d'annotation tient compte de deux modalités logiques : les opinions (O) et les accords/désaccords (AD). Conformément à la théorie de Charaudeau, ces évaluations ne possèdent pas d'axiologie. En revanche leur modalité se spécialise en plusieurs sous-modalités possibles :

- les opinions ont une sous-modalité parmi les suivantes : *conviction*, *supposition forte*, *supposition moyenne*, *supposition faible* ou un *pressentiment* ;
- pour les accords/désaccords : *accord total* ou *approximatif*, *rectificatif* ou *désaccord total*.

Dans l'exemple (8), le passage évaluatif *je pense* correspond à une modalité d'opinion, de sous-modalité *supposition moyenne* (SM) et de configuration explicite (OSME). Dans (9), le passage évaluatif *Je ne suis pas convaincue* correspond à une opinion de conviction (C) explicite (OCE).

- (8) Mais ce qui plait le plus dans le coeur des hommes, comme dans beaucoup de comédies de moeurs, **<Opinion type="OSME">**je pense**</Opinion>**, c'est [...].
- (9) **<Opinion type="OCE">**Je ne suis pas convaincue**</Opinion>** que la fidélité doit être la condition absolue de la pérennité d'un couple. Et du même coup, **<Opinion type="OSME">**je pense**</Opinion>** qu'un mensonge est parfois salutaire.

Le choix de ce modèle d'annotation soulève a posteriori de nombreuses questions. Nous discutons des choix effectués et des évolutions à envisager dans la conclusion de cette thèse (voir chapitre 8). Néanmoins, ce modèle nous semble une amorce utile pour pouvoir développer et évaluer des méthodes

d'analyse à granularité fine des évaluations. Dans le paragraphe suivant, nous donnons une estimation de la complexité de la tâche d'annotation du corpus selon ce modèle.

4.1.2.3 Accord inter-annotateurs : une tâche d'annotation complexe

Pour estimer la complexité de cette tâche d'annotation manuelle, nous mesurons l'accord inter-annotateurs en utilisant deux métriques :

- la métrique *agr* définie par Wiebe et collab. (2006). Celle-ci évalue l'intersection existante entre les annotations posées par deux annotateurs (équation 4.1). A (respectivement B) est l'ensemble des annotations posées par l'annotateur a (respectivement b) ;
- le coefficient κ de Fleiss (1971). Cet indice est utilisé pour mesurer l'accord entre plusieurs annotateurs sur la catégorisation des objets et des évaluations annotées.

$$agr(a|b) = \frac{A \cap B}{|A|} \quad (4.1)$$

Après une phase de normalisation du modèle menée sur 12 documents du corpus, trois linguistes ont simultanément annotés 64 autres documents issus de la première phase d'extraction. Le tableau 4.1 présente l'accord observé selon les métriques choisies et le type d'annotation. Le score *agr* figurant dans ce tableau est la moyenne arithmétique des accords entre les trois annotateurs. Les résultats obtenus montrent que l'intersection des annotations entre deux annotateurs est en moyenne assez faible : 45 % pour les objets et 52 % pour les évaluations. Parmi les objets, les concepts associés provoquent le plus de désaccords. En revanche, lorsque les annotateurs s'accordent sur une annotation, ils sont également pratiquement toujours d'accord sur les traits à associer à cette annotation (indice κ de 94 % à 97 %). L'aspect critique concerne donc le fait d'annoter ou ne pas annoter un passage, plutôt que sa catégorisation.

	<i>agr</i>	κ
Objets	45 %	94 %
<i>Concept Concerné</i>	62 %	-
<i>Concept Associé</i>	38 %	-
<i>Instance</i>	54 %	-
Évaluations	52 %	97 %

TABLE 4.1 : Accord inter-annotateurs pour l'annotation des objets et des évaluations

Dans le paragraphe suivant, nous présentons des statistiques sur le corpus Blogoscopie obtenu suite à la phase d'annotation manuelle.

4.1.3 Le corpus Blogoscopie : résultats et statistiques

Le corpus Blogoscopie est le résultat de l'annotation manuelle des 250 documents (billets et commentaires) extraits précédemment. Il contient environ 115 000 mots et 11 776 passages annotés :

- 6 876 objets textuels distincts ;
- 5 900 passages évaluatifs.

Dans une optique expérimentale, ce corpus est segmenté en deux parties : l'une réservée à l'entraînement des méthodes et l'autre réservée aux tests. Nous présentons les caractéristiques statistiques de ces deux sous-parties du corpus dans les paragraphes suivants.

4.1.3.1 Le corpus d'entraînement

Le corpus d'entraînement est constitué de 200 billets et 614 commentaires. Ces documents correspondent aux trois premières phases d'extraction. Ils contiennent :

- 6 876 objets textuels : 5 666 concepts et 1 210 instances de concepts. Nous étudions plus précisément la répartition et la position de ces objets dans le chapitre 7 (§. 7.2) ;
- 4 945 passages évaluatifs annotés.

La répartition de ces évaluations est présentée dans les tableaux 4.2, 4.3 et 4.4. D'après le tableau 4.2, nous observons que :

- les évaluations de modalités **axiologiques** sont présentes à 4 306 reprises. L'écart important entre le nombre de jugements et le nombre d'appréciations nous semble dû à un fort taux de désaccord entre les annotateurs sur ces deux modalités. La nuance entre jugement et appréciation semblent délicates en contexte. L'annotateur humain a dès lors privilégié la modalité d'appréciation ;
- les évaluation de modalités **logiques** sont moins souvent présentes : 639 exemples.

Nous observons également :

- qu'environ 75 % des 4 945 passages évaluatifs ont une configuration implicite (Tab. 4.3) ;
- et 60 % des évaluations axiologiques ont une polarité positive (Tab. 4.4).

Nombre d'évaluations	Appréciation	Jugement	Opinion	Accord/Désaccord
Corpus d'entraînement	4 163	143	270	369
Corpus de test	851	-	60	44
<i>Sarah Palin</i>	376	-	32	10
<i>Sushi</i>	475	-	28	34

TABLE 4.2 : Répartition des modalités d'évaluations dans le corpus Blogoscopie

NB d'évaluations	Implicite	Explicite	NB d'évaluations	Positif	Négatif
Entraînement	3 936	1 009	Entraînement	2 593	1 713
Corpus de test	685	270	Corpus de test	526	325
<i>Sarah Palin</i>	349	69	<i>Sarah Palin</i>	171	205
<i>Sushi</i>	336	201	<i>Sushi</i>	355	120

TABLE 4.3 : Répartition des configurations d'énonciations dans le corpus Blogoscopie

TABLE 4.4 : Répartition des axiologies dans le corpus Blogoscopie

Suite aux observations issues du corpus d'entraînement, il a été décidé de ne pas faire de distinction entre jugement et appréciation lors de l'annotation du corpus de test.

4.1.3.2 Le corpus de test

Le corpus de test est constitué de 50 billets et 145 commentaires. Ces documents correspondent à la quatrième phase d'extraction. Dans ce corpus, seules les évaluations sont annotées (955 passages évaluatifs), aucun objet textuel n'a été annoté (voir tableaux 4.2, 4.3 et 4.4).

Les tendances obtenues sur le corpus de test sont assez similaires à celles du corpus d'entraînement :

- 89 % des évaluations du corpus de test sont axiologiques (851 appréciations contre 104 évaluations logiques) ;
- 72 % des évaluations ont une configuration d'énonciation implicite ;
- 62 % des évaluations axiologiques ont une polarité positive.

La partie entraînement du corpus Blogoscopie et un manuel d'utilisation sont disponibles sous licence LGPLLR et téléchargeables à l'adresse suivante : <http://www.lina.univ-nantes.fr/Ressources.html>. Un document annoté du corpus Blogoscopie est fourni en annexe B. Dans le paragraphe suivant, nous repositionnons les caractéristiques de ce corpus par rapport à son équivalent anglophone le plus proche.

4.1.4 Travaux liés et discussion

À notre connaissance, le travail d'annotation manuelle des passages évaluatifs le plus important est celui effectué par Wiebe et ses collègues pour l'anglais : le corpus MPQA (Wiebe et collab., 2006). Nous synthétisons brièvement leur travail et le comparons à notre propre travail.

Le corpus MPQA Le corpus MPQA est composé d'environ 130 000 mots provenant de 535 articles de presses écrits en anglais et publié entre juin 2001 et mai 2002. Ces articles proviennent de 187 journaux de pays anglophones différents⁶. Ce corpus a été collecté durant l'été 2002 et annoté pendant 15 mois par 5 annotateurs. Les articles de presse concernent dix sujets d'actualités géopolitiques : *la crise économique en Argentine, Guantanamo, la ratification du protocole de Kyoto, les élections présidentielles au Zimbabwe, etc.*

Selon leur modèle d'annotation, les notions d'évaluation, d'opinion, d'émotion ou de sentiment sont considérées comme des points de vue et sont regroupées dans une seule catégorie (*private state*) (Quirk, Greenbaum, Leech et Svartvik, 1985 ; Banfield, 1982). Les auteurs définissent la notion de point de vue comme étant le contraire d'une observation factuelle objective. Les passages textuels correspondants à des points de vue sont délimités dans le corpus et sont notamment enrichis par quatre types d'informations :

- la source : la personne physique ou morale qui exprime le point de vue ;
- la cible : l'objet qui est concernée par le point de vue ;
- l'intensité : elle peut être faible, moyenne, forte ou extrême ;
- la polarité axiologique : elle peut être positive, négative ou neutre.

(10) « the report is **full of absurdities** », said *Xirao-Nima*.

Leur modèle d'annotation tient compte du genre journalistique du corpus. Les textes contiennent en effet beaucoup de discours rapporté pour lequel il est nécessaire d'identifier la personne source (10). Les objectifs des travaux de Wiebe sont identiques aux nôtres sur plusieurs aspects. Dans les deux cas, il s'agit de :

- collecter les différentes façons d'exprimer une évaluation dans la langue à travers l'étude et l'annotation d'un corpus ;

6. Le corpus est librement disponible à l'adresse suivante : <http://nrrc.mitre.org/NRRC/publications.htm>

- décrire les évaluations à partir d'un modèle théorique. Wiebe utilise des théories anglaises (Quirk et collab., 1985), alors que nous utilisons des travaux français (Charaudeau, 1992) pour décrire l'expression des évaluations dans la langue.

Les deux approches partagent également les choix suivants :

- les évaluations peuvent être constituées d'un ou de plusieurs mots, de toutes catégories grammaticales possibles ;
- certaines évaluations ont une axiologie et d'autres sont neutres ;
- certaines évaluations ont un trait ironique ;
- les évaluations axiologiques sont rattachées à une cible.

Les deux modèles d'annotations diffèrent en revanche fondamentalement sur la théorie linguistique utilisée pour décrire les évaluations. Celle utilisée par Wiebe s'intéresse essentiellement à l'intensité et à l'axiologie des évaluations alors que la nôtre décrit l'axiologie et différentes caractéristiques énonciatives (modalité, configuration d'énonciation) de l'évaluation. Par ailleurs, dans le corpus Blogoscopie, nous n'avons pas pris en compte certaines caractéristiques présentes dans le corpus MPQA (*source, agent, patient, etc.*). Cette différence s'explique par la nature du corpus et par les objectifs applicatifs visés. Le corpus Blogoscopie est construit avec l'objectif applicatif principal de pouvoir identifier les évaluations exprimées autour d'une même cible dans des thématiques diverses. Sur les blogs, la source d'une évaluation est la plupart du temps l'auteur du billet ou du commentaire. Cette caractéristique ne nous a donc pas semblé primordiale. En revanche, l'aspect multithématique est un point davantage mis en avant dans le corpus Blogoscopie par rapport au corpus MPQA.

Dans la section suivante nous évoquons un second besoin, celui de décrire le vocabulaire évaluatif dans un lexique. À l'image des travaux de Wiebe qui utilise le Subjectivity Lexicon (Wilson, Wiebe et Hoffmann, 2005), notre approche nécessite le développement d'un lexique pour le français.

4.2 Un lexique de mots évaluatifs pour le français

La connaissance du vocabulaire évaluatif d'une langue constitue un pré-requis pour les méthodes de fouille d'opinions. Les caractéristiques de ce vocabulaire sont généralement stockées dans des lexiques. Pour l'anglais, de telles ressources sont nombreuses, SentiWordNet (Baccianella, Esuli et Sebastiani, 2010) et le Subjectivity Lexicon (Wilson et collab., 2005) sont les références les plus souvent citées. En revanche, le lexique des sentiments (Mathieu, 2006) est à notre connaissance la seule ressource pour le français.

Nos objectifs nous amènent à développer un lexique pour le français, le *lexique de l'évaluation*. Ce travail a démarré en 2007 dans le cadre du projet ANR Blogoscopie et s'inscrit parallèlement à la création du lexique des sentiments. Notre démarche vise à structurer le vocabulaire évaluatif du français selon ses caractéristiques grammaticales et sémantiques (modalité, axiologie, ambiguïté, etc.). Dans la suite de ce chapitre, nous présentons le processus de création et les caractéristiques du lexique de l'évaluation (§. 4.2.1). Nous présentons des données statistiques sur ses entrées (§. 4.2.2) et dans une dernière partie, nous le comparons au lexique des sentiments et aux lexiques anglophones (§. 4.2.3).

4.2.1 Du corpus annoté aux entrées lexicales

Notre objectif consiste à développer un lexique francophone structurant le vocabulaire utilisé pour exprimer une évaluation. Pour cela, nous attachons une importance particulière à décrire ce vocabulaire selon sa (ou ses) signification(s) en contexte. Nous nous appuyons sur une observation

des valeurs prises par chaque entrée lexicale dans le corpus Blogoscopie (Fig. 4.1). La structuration de ce lexique tient compte des traits grammaticaux (§. 4.2.1.1) et des valeurs sémantiques prises en contexte par chaque entrée (§. 4.2.1.2) D'un point de vue technique, les entrées du lexique sont structurées dans un format de données XML.

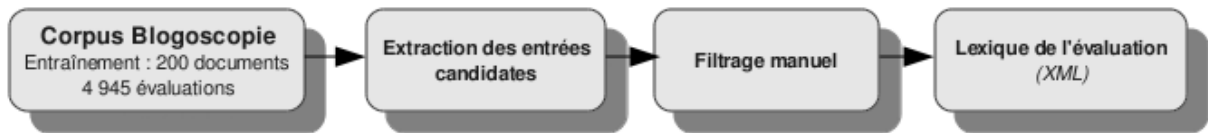


FIGURE 4.1 : Développement du lexique de l'évaluation à partir du corpus Blogoscopie

Extraction et filtrage des candidats Les 200 documents de la partie entraînement du corpus Blogoscopie sont utilisés comme unique ressource pour constituer le lexique. Une première phase automatique consiste à annoter la classe grammaticale et le lemme de chaque mot du corpus à l'aide de l'étiqueteur LemmaNG⁷. Les 4 945 évaluations du corpus sont ensuite extraites tout en conservant les annotations posées dans le corpus. Les étapes suivantes sont entièrement réalisées manuellement, il s'agit de :

- parcourir un fichier regroupant, ligne par ligne, les 4 945 évaluations du corpus Blogoscopie ;
- identifier les mots ou les expressions qui portent spécifiquement le sens évaluatif ;
- ajouter les entrées au lexique en précisant les caractéristiques grammaticales et sémantiques rencontrées dans le contexte du corpus Blogoscopie (voir §. 4.2.1.1 et §. 4.2.1.2) ;
- effectuer une phase de validation. Cette phase a pour objectif d'identifier les entrées du lexique potentiellement ambiguës (voir §. 4.2.1.2).

Les caractéristiques grammaticales et sémantiques associées à chaque entrée du lexique sont précisées dans les paragraphes suivants.

4.2.1.1 Caractéristiques grammaticales

Les entrées du lexique peuvent être des unigrammes de mots (*abominable*, *abandon*, *accepter*, etc.) ou des formes complexes (*au rabais*, *du jamais vu*, *il y en a marre*, etc.). Ces formes complexes sont soit des syntagmes, soit des propositions. Les entrées lexicales ne sont pas limitées en termes de classes grammaticales. Dans le lexique, chaque entrée est lemmatisée et possède une classe grammaticale (11) ou une suite de classes grammaticales si l'entrée est complexe (12). Les exemples d'entrées lexicales données dans ce manuscrit sont présentées dans un format allégé⁸.

- (11) **Entrée** id="ADJ_désopilant", ambiguïté="faux"
 • **Grammatical** lemme="désopilant", classe="ADJ"
 • **Sémantique** [...]
- (12) **Entrée** id="SV_faire_un_bien_fou", ambiguïté="faux"
 • **Grammatical** lemme="faire un bien fou", classe="VER DET NOM ADJ"
 • **Sémantique** [...]

7. Outil interne de Sinequa.

8. Une description plus précise et plus technique des balises XML utilisées est disponible dans le manuel utilisateur du lexique.

Les entrées des exemples (11) et (12) sont considérées comme non ambiguës. L'ambiguïté d'une entrée est l'une des caractéristiques sémantiques associées à chaque entrée lexicale. Nous les précisons ci-dessous.

4.2.1.2 Caractéristiques sémantiques

Dans le contexte du corpus d'entraînement Blogoscopie, chaque entrée peut être rencontrée une ou plusieurs fois. Si une entrée est associée à plusieurs modalités ou plusieurs axiologies différentes, cette entrée est considérée comme ambiguë. L'ambiguïté peut porter sur la modalité, sur l'axiologie ou sur les deux. Par exemple, l'entrée *exceller* (13) est non ambiguë car elle est toujours associée à une modalité d'appréciation et à une axiologie positive. En revanche, *fou* (14) possède une ambiguïté axiologique.

- (13) **Entrée** id="VER_exceller", ambiguïté="faux"
 • **Grammatical** lemme="exceller", classe="VER"
 • **Sémantique** modalité="appréciation", axiologie="positif"
- (14) **Entrée** id="ADJ_fou", ambiguïté="vrai", type_ambiguïté="axiologie"
 • **Grammatical** lemme="fou", classe="ADJ"
 • **Sémantique** modalité="appréciation", axiologie="positif#négatif"

De la même façon, selon les exemples rencontrés dans le corpus, le verbe *nier* (15) possède une ambiguïté sur la modalité. Celle-ci correspond soit à une appréciation, soit à une opinion.

- (15) **Entrée** id="VER_nier", ambiguïté="vrai", type_ambiguïté="modalité"
 • **Grammatical** lemme="nier", classe="VER"
 • **Sémantique** modalité="Appréciation#Opinion", axiologie="négatif"

Pour chaque entrée du lexique, nous conservons les différentes références au corpus. Ainsi, pour l'exemple *novateur*, nous conservons l'information que cet adjectif apparaît dans une évaluation (*la plus novatrice depuis longtemps!*) du document *Wii-7701912-7701941.xml* issu du corpus Blogoscopie. L'évaluation porte sur l'objet textuel *console*. Cette dernière information est ajoutée en prévision de traitements futurs visant à désambigüiser une entrée lexicale selon la cible de l'évaluation. Un extrait du lexique de l'évaluation au format XML est fourni dans l'annexe C.

- (16) **Entrée** id="ADJ_novateur", ambiguïté="faux"
 • **Grammatical** lemme="novateur", classe="ADJ"
 • **Sémantique** modalité="Appréciation", axiologie="positif"
 • **Concordance** fichier="Wii-7701912-7701941.xml" cible="console"
 évaluation="la plus novatrice depuis longtemps" !

4.2.2 Le lexique de l'évaluation : résultats et statistiques

Dans sa version initiale⁹, le lexique de l'évaluation est constitué de 984 entrées. Le tableau 4.5¹⁰ donne le détail du nombre d'entrées selon leur catégorie grammaticale et leurs caractéristiques sémantiques.

9. L'enrichissement automatique du lexique de l'évaluation fait l'objet du chapitre 6.

10. CCO : Conjonction de coordination – CSU : Conjonction de subordination – SN : Syntagme nominal – SV : Syntagme verbal – SA : Syntagme adjectival.

Caractéristiques grammaticales Les adjectifs représentent la moitié des entrées du lexique (50,1 %), devant les verbes (19,5 %), les noms (16,9 %) et les adverbes (6,1 %). Nous notons également que 2 % des entrées sont des interjections (*grr, lol, miam, oups, snif, tada, wow, youpi*, etc.). Sur les blogs, ces interjections sont particulièrement variables et sont des indicateurs très précis de la présence de l'énonciateur (subjectivité de premier degré). Elles pourraient dès lors faire l'objet d'une étude spécifique pour renforcer cet aspect.

Par ailleurs, nous observons que seules 4,1 % des entrées du lexique ne sont pas des unigrammes de mots. Il s'agit d'un des points faibles des lexiques du domaine. Nous présentons plus largement cette problématique dans le chapitre 6. Enfin, selon les données du lexique de l'évaluation, les modalités logiques sont exprimées en particulier à partir :

- de verbes : 41,9 % des marqueurs de modalités logiques ;
- d'adverbes : 34,4 % des marqueurs de modalités logiques.

catégorie	nombre d'entrées	+ ambiguïtés	axiologique	logique	positif	négatif
Adjectif	493	26	505	18	267	238
Nom	166	3	167	2	88	79
Adverbe	60	9	28	49	17	11
Verbe	192	15	153	60	77	76
Interjection	20	0	20	0	13	7
CCO	3	0	0	3	0	0
CSU	8	0	0	8	0	0
Déterminant	1	0	1	0	1	0
Préposition	1	0	1	0	0	1
SN	1	0	1	0	1	0
SV	24	0	23	1	13	10
SA	1	0	1	0	0	1
Phrase	14	0	12	2	8	4
Total	984	53	912	143	485	427

TABLE 4.5 : Caractéristiques des entrées du lexique de l'évaluation

Caractéristiques sémantiques Du point de vue des caractéristiques sémantiques, nous remarquons qu'une ambiguïté a été identifiée pour 5,4 % des entrées du lexique. En particulier, 24 adjectifs (2,5 %) ont une axiologie variable selon le contexte (*brave, délicat, fou, incroyable, sévère*, etc.). Les entrées du lexique sont très majoritairement liées aux modalités d'appréciation et de jugement (modalités axiologiques) :

- 92,7 % des entrées ont une signification évaluative axiologique (*magnifique, ne pas mâcher ses mots, fastoche*, etc.) ;
- 14,5 % des entrées ont une signification évaluative logique (*constater, croire, effectivement, être vrai*, etc.) ;
- 7,2 % des entrées ont donc une modalité ambiguë mêlant axiologie et logique (*comprendre, conseiller, impossible*, etc.).

En revanche, parmi les entrées axiologiques, la proportion d'entrées positives et négatives est relativement équilibrée : 53,2 % des entrées axiologies sont positives, 46,8 % sont négatives.

Le lexique de l'évaluation n'a pas fait l'objet d'une mesure d'accord inter-annotateurs ou d'un autre protocole d'évaluation hors-contexte. Néanmoins, dans le chapitre 5, nous évaluons en contexte

ce lexique sur son aspect qualitatif et quantitatif. Dans le paragraphe suivant, nous comparons ces caractéristiques à celles de lexiques francophones ou anglophones.

4.2.3 Travaux liés et discussion

À l'exception du Subjectivity Lexicon (voir chapitre 6, §. 6.1), les ressources existantes n'ont pas été évaluées quantitativement et qualitativement sur une tâche d'analyse à granularité fine des évaluations. Il est alors complexe d'estimer quelles sont les ressources réellement les plus efficaces. Nous donnons néanmoins des repères pour comparer ces ressources en s'appuyant sur plusieurs critères : le nombre d'entrées, les traits grammaticaux et sémantiques pris en compte et leur processus de création. Nous considérons le lexique des sentiments (Mathieu, 2006), le SentiWordnet (Baccianella et collab., 2010) et le Subjectivity Lexicon (Wilson et collab., 2005) comme points de comparaison. Leurs caractéristiques sont regroupées dans le tableau 4.6.

lexique	de l'évaluation	des sentiments	SentiWordNet	Subjectivity
Langue	français	français	anglais	anglais
Nombre d'entrées	984	950	115 000	8 000
Type de création	manuel	manuel	automatique	auto. et manuel
Traits grammaticaux				
<i>Classe grammaticale</i>	toutes	verbe et nom	toutes	toutes
<i>Forme complexe</i>	partiellement	non	non	non
Traits sémantiques				
<i>Axiologie</i>	oui	oui	oui	oui
<i>Classe sémantique</i>	non	oui	non	non
<i>Intensité</i>	non	oui	non	oui
<i>Modalité</i>	oui	non	non	non
<i>Représentation des ambiguïtés</i>	oui	non	oui	non

TABLE 4.6 : Caractéristiques des lexiques francophones et anglophones existants

Caractéristiques grammaticales Les lexiques français développés manuellement ont un nombre d'entrées comparables (984 pour le lexique de l'évaluation et 950 pour le lexique des sentiments). Le lexique des sentiments ne contient en revanche que des verbes et des noms. Néanmoins, celui-ci est plus riche que le lexique de l'évaluation sur ces deux catégories grammaticales. Il contient :

- 600 verbes contre 192 pour le lexique de l'évaluation ;
- 350 noms contre 166 pour le lexique de l'évaluation.

De part leur mode de création automatique ou semi-automatique, la taille des lexiques anglophones est beaucoup plus importante. Le SentiWordNet contient 115 000 des synsets¹¹ de WordNet (Fellbaum, 1998), alors que le Subjectivity Lexicon possède 8 000 mots. Nous précisons néanmoins que les entrées du Subjectivity Lexicon ne sont pas lemmatisées. Ainsi *doubt* et *doubts* comptent pour deux entrées. Les ressources anglophones et le lexique des sentiments ne contiennent que des unigrammes de mots. Cet aspect est un facteur limitant des ressources existantes que nous abordons plus précisément dans le chapitre 6. Quelques formes complexes sont présentes dans le lexique de l'évaluation mais elles sont limités à 4,1 % du nombre d'entrées totales.

11. Une signification particulière d'un mot.

Caractéristiques sémantiques D'un point de vue sémantique, les ressources existantes francophones ou anglophones nous semblent complémentaires. Elles représentent l'axiologie positive, négative ou neutre d'une entrée lexicale, mais ont ensuite chacune leurs spécificités. Le lexique de l'évaluation et SentiWordNet représente de façon différente l'ambiguïté d'une entrée lexicale : de façon binaire pour le lexique de l'évaluation et de façon graduée pour SentiWordNet. Par exemple, *dingue* est soit positif, soit négatif dans le lexique de l'évaluation. Dans SentiWordNet, le mot *crazy* est à 62,5 % positif et 37,5 % négatif. L'ambiguïté n'est pas représentée dans les autres ressources.

Le lexique des sentiments et le Subjectivity Lexicon associent une étiquette à chaque entrée pour estimer son degré d'intensité (élevé, moyen ou faible). Par exemple, *exaspération* a une intensité élevée selon le lexique des sentiments. Le lexique des sentiments est le seul à scinder le vocabulaire évaluatif en 33 classes sémantiques (*tristesse, étonnement, indifférence, intérêt, amusement, etc.*) et à relier les classes par des relations d'opposition ou d'intensité. Le lexique de l'évaluation est quant à lui le seul à distinguer le vocabulaire évaluatif selon leur modalité d'énonciation.

Ces caractéristiques spécifiques à chaque ressource suggèrent d'établir des protocoles pour évaluer et comparer les ressources en contexte. Les résultats mèneraient, en particulier pour le français, à réunir les points forts de chaque ressource et à mieux estimer leurs lacunes. Une observation manuelle des ressources anglophones montrent notamment que de part leur création automatique elles contiennent beaucoup d'entrées susceptibles de provoquer du bruit (voir chapitre 6).

4.3 Conclusion

Dans ce chapitre, nous avons présenté les ressources que nous allons utiliser dans nos travaux, le corpus Blogoscopie et le lexique de l'évaluation. Elles ont été développées dans le cadre de cette thèse et du projet ANR Blogoscopie. Ces ressources répondent à deux besoins permettant d'amorcer les travaux d'analyse à granularité fine des évaluations pour le français et dans un cadre multithématique :

- le corpus Blogoscopie contient 250 textes issus de blogs francophones et abordant des thématiques variées. 5 900 passages évaluatifs et 6 876 objets textuels y ont été annotés manuellement. Le modèle utilisé pour annoter les évaluations tient notamment compte des travaux de Charaudeau (1992) présentés dans le chapitre 3 ;
- le lexique de l'évaluation contient 984 entrées, principalement des unigrammes de mots, de catégories grammaticales variées. Sa taille est équivalente à celle d'une autre ressource lexicale disponible pour le français, le lexique des sentiments. Les caractéristiques de ces lexiques nous semblent complémentaires. La comparaison et l'union de ces deux lexiques est une perspective envisageable non abordée dans cette thèse. L'évaluation qualitative et quantitative, en contexte, du lexique de l'évaluation fait elle l'objet d'une partie du chapitre 5.

D'un point de vue informatique, ces deux ressources permettent d'aborder une première problématique : la délimitation et la catégorisation automatique des passages évaluatifs exprimés dans un texte. Dans ce contexte, le corpus Blogoscopie offre des données d'apprentissages et une annotation de référence pour évaluer les méthodes. Cette problématique fait l'objet du chapitre 5.

Délimiter et catégoriser les évaluations

5.1	Analyse à granularité fine des évaluations	68
5.1.1	Un double objectif : délimiter et catégoriser	68
5.1.1.1	<i>Détecter et délimiter les évaluations</i>	69
5.1.1.2	<i>Catégoriser les traits sémantiques et énonciatifs</i>	70
5.1.2	Modéliser l'évaluation en articulant lexique, grammaire et sémantique	72
5.1.2.1	<i>Des structures linguistiques spécifiques à l'évaluation</i>	72
5.1.2.2	<i>Automatiser le processus d'acquisition des structures évaluatives</i>	72
5.2	Acquisition de structures évaluatives	75
5.2.1	Extraction des passages évaluatifs à partir d'un corpus annoté	75
5.2.2	Transformation en structures évaluatives généralisées	76
5.2.2.1	<i>Règles de substitution</i>	77
5.2.2.2	<i>Règles d'ajout ou de suppression</i>	78
5.2.3	Annotation de méta-données sur les structures évaluatives	79
5.3.4	Résultats et discussion	80
5.3	Délimitation et catégorisation des évaluations	82
5.3.1	Méthode par structures évaluatives	82
5.3.1.1	<i>Détection et délimitation</i>	82
5.3.1.2	<i>Catégorisation sémantique et énonciative</i>	82
5.3.2	Méthode lexicale (<i>baseline</i>)	83
5.3.2.1	<i>Détection et délimitation</i>	84
5.3.2.2	<i>Catégorisation sémantique et énonciative</i>	84
5.4	Expérimentation : intérêts et limites des structures évaluatives	85
5.4.1	Protocole d'évaluation	85
5.4.2	Une bonne délimitation des évaluations	86
5.4.3	Une couverture trop faible du vocabulaire évaluatif	88
5.4.4	Une bonne catégorisation des traits sémantiques et énonciatifs	89
5.4.5	Un intérêt pour résoudre les effets du contexte sur l'axiologie	90
5.4.6	Une bonne adaptation à des corpus multi-thématiques	91
5.5	Conclusion	93

En fouille d'opinions, les travaux existants ont majoritairement analysé l'opinion au grain document, cherchant à caractériser leur subjectivité ou leur polarité (voir chapitre 2, §. 2.1). Pour structurer plus finement les données d'opinions, des travaux proposent plus récemment de catégoriser les phrases selon leur subjectivité ou leur polarité (voir chapitre 2, §. 2.2). Pour pallier les limites de ces approches, améliorer la qualité des analyses et ainsi mieux répondre aux besoins applicatifs récents, nous envisageons deux nouveaux objectifs :

- diminuer le grain d'analyse et délimiter des passages évaluatifs intraphrastiques ;
- caractériser plusieurs aspects sémantiques et énonciatifs de ces évaluations (voir chapitre 3, §. 3.3) ;

Pour réaliser ces deux objectifs (§. 5.1), notre démarche s'appuie sur l'idée de Thompson et Hunston (2000) selon laquelle il existe de nombreuses structures linguistiques spécifiques qui contraignent le sens d'une évaluation dans son contexte. Une première étape consiste alors à acquérir automatiquement ces structures évaluatives en proposant une méthode d'extraction à partir d'un corpus d'entraînement en français (§. 5.2). À partir de ces structures et d'un lexique, nous proposons une méthode pour délimiter et catégoriser les passages évaluatifs (§. 5.3). Nous en mesurons l'intérêt et les limites à travers une expérimentation sur un corpus de tests. Nous comparons la méthode par structures évaluatives à une méthode lexicale « *baseline* » (§. 5.4). Enfin, nous discutons des résultats obtenus et des perspectives de ce travail (§. 5.5).

5.1 Analyse à granularité fine des évaluations

L'analyse à granularité fine des évaluations se scinde en deux phases : (i) détecter et délimiter correctement les évaluations d'un document, (ii) catégoriser différents aspects sémantiques et énonciatifs des évaluations détectées lors de la première phase. La réussite de la phase de catégorisation repose en grande partie sur la phase de détection et délimitation. Nous présentons plus précisément ce double objectif dans la première partie de cette section (§. 5.1.1). Dans une deuxième partie, pour atteindre ces objectifs, nous présentons l'idée de modéliser les évaluations par des structures linguistiques articulant des constituants lexicaux, grammaticaux et sémantiques (§. 5.1.2).

5.1.1 Un double objectif : délimiter et catégoriser

L'exemple du billet de blog (1) montre que les évaluations véhiculées par un même document sont souvent très hétérogènes. Dans cet extrait, des appréciations positives (en bleu), des appréciations négatives (en rouge) et des opinions (en gris) structurent l'argumentation globalement autour d'un seul sujet, *le blocage d'un lycée au cours d'une grève*. Néanmoins, les passages évaluatifs :

- concernent des cibles différentes : *coupe de cheveux, blocus, question, situation des jeunes, manifestation, police, etc.* ;
- sont exprimées par des sources différentes : l'auteur du billet, *Nora, une dame, etc.* ;
- ont des modalités d'évaluations différentes : opinion (conviction ou supposition) ou appréciation ;
- sont énoncés avec des configurations différentes : explicite ou implicite ;
- ont une axiologie différente : positif ou négatif.

Pour pouvoir mieux structurer le discours évaluatif exprimé dans ce document, notre objectif consiste donc d'une part, à détecter et délimiter ces différents passages évaluatifs et d'autre part, à catégoriser leurs caractéristiques sémantiques et énonciatives.

- (1) **BILLET : *Ma semaine lycéenne en mode blocus*** | Jeudi 21/10/2010 |
 Vendredi dernier, nous sommes nombreux à nous retrouver « bloqués » devant le lycée [...]. Ceux qui s'aventurent à vouloir passer par-dessus la grille, ont de grandes chances de se prendre un oeuf sur le crâne. Comme j'aime bien ma coupe de cheveux ce jour-là, je fais demi-tour... [...] je pars de chez moi à 8 heures en me disant que j'ai peu d'espoir d'atterrir en salle 104 [...] Miracle, il n'y a pas un chat devant le portail. [...] « T'as vu, ils nous ont saoulés avec leur blocus bidon, là, beaucoup de bruit pour rien du tout ! » Mardi 19, jour de grève nationale. Le bus passe, ouf ! A 8 heures, le lycée n'est pas encore bloqué [...]. En attendant mon tram, je rencontre Nora, agacée par cette histoire : « C'est plus une manif', à ce stade, c'est un massacre ! [...] » Deux élèves de la classe nous font des signes, elles sont dans le bâtiment où l'on devait avoir cours. Comment ont-elles fait pour entrer ? Bonne question. Notre prof est descendu pour essayer de nous faire entrer nous aussi. En vain ! [...] Des affiches sont collées sur les grilles : « Les jeunes dans la misère et les vieux dans la galère ». [...] nous attendons avec toujours un peu d'espoir que ça se débloque, mais tu parles, c'est de pire en pire. La police est présente pour assurer la sécurité. [...] Une dame d'un certain âge vient nous faire part de son avis : « Vous avez raison, c'est très bien ce que vous faites parce que c'est votre avenir qui est en jeu [...]. En tout cas, bon courage pour la manifestation, continuez, vous avez raison, faut se défendre, allez les jeunes, moi je suis avec vous ! » [...] - extrait de <http://yahoo.bondyblog.fr>

COMMENTAIRE : Je ne sais pas si tous les lycéens qui sont dans la rue protestent contre la réforme : j'imagine que pour beaucoup, c'est une simple occasion de ne pas aller en cours mais ce qui est sûr, c'est qu'une majorité est dotée de conscience politique. Pour ma part, je suis convaincue qu'il y a un malaise bien plus grand, et que la réforme n'a été que le déclencheur. Une inquiétude importante était déjà présente depuis bien longtemps dans l'esprit de tous concernant l'entrée sur le marché du travail. Ne reste que l'espoir d'être entendus...

5.1.1.1 Détecter et délimiter les évaluations

Notre premier objectif consiste à détecter et délimiter les différentes évaluations d'un document. De façon empirique, nous considérons qu'une évaluation a pour granularité minimale le syntagme non récursif (ou chunk) nominal ou verbal au sens de Abney (1991). Dans l'exemple (2), où les chunks apparaissent entre « | », l'évaluation englobe tout le syntagme *est vraiment incompetent* et non pas uniquement le mot évaluatif *incompétent*.

- (2) | Domenech | [...] | il | est vraiment **incompétent** |.

L'évaluation peut également être constituée de plusieurs chunks successifs. En particulier, il est important de considérer comme faisant parti de l'évaluation les tournures de négation qui sont exprimées dans les chunks voisins (3) (4).

- (3) | Domenech | **n'est pas** | un bon entraîneur |, il | ne transmet rien |.
 (4) | Domenech | dans une publicité **géniale** |. | **Difficile** | de ne pas | le | trouver **sympathique** | après cette publicité |!

Nous fixons également que les bornes d'une évaluation se superposent toujours aux bornes fixées par une segmentation en syntagmes non récursifs. En d'autres termes, la délimitation des passages

évaluatifs repose sur une segmentation syntaxique minimale. Par ailleurs, certains passages évaluatifs apparaissent dans un voisinage proche, s'ils ont une modalité, une configuration d'énonciation ou une axiologie différente, il ne s'agit pas de la même évaluation selon Charaudeau (1992). Ainsi, dans les exemples (5) et (6)¹, il faut pouvoir délimiter chacune des évaluations présentes par des annotations différentes. L'identification des relations entre ces différentes évaluations est une étape ultérieure hors du cadre de cette thèse. En revanche, la catégorisation de ces passages évaluatifs est le second objectif de l'analyse des évaluations à granularité fine.

- (5) Je suis persuadé que l'on va droit à l'échec.
Opinion (conviction) AIN
- (6) La démarche d'Obama est certes intéressante, mais elle ne peut réellement être efficace si celle-ci [...].
Accord AIP. Rectificatif AIN

5.1.1.2 Catégoriser les traits sémantiques et énonciatifs

Le second objectif consiste à catégoriser les évaluations selon les trois caractéristiques sémantiques et énonciatives que nous avons définies dans le chapitre 3 :

- la modalité ;
- la configuration d'énonciation ;
- et l'axiologie.

L'axiologie d'une évaluation est une caractéristique particulièrement sensible au contexte. Les tournures de négation en sont le premier exemple. Celles-ci sont rarement prises en compte dans les méthodes à grains « larges ». Au grain document, les méthodes attribuent le plus souvent une polarité positive ou négative à une unité linguistique (mot, expression, n-gramme) sans considérer le contexte dans lequel elle peut être instanciée. Pour ce grain, les tournures de négation ne sont classiquement pas traitées car elles sont considérées comme marginales par rapport à l'ensemble des évaluations contenues dans le document. Les cas bruités dûs aux tournures de négation n'affecteront pas ou peu la réponse globale du système. En revanche, au grain phrase, plusieurs travaux (Ku et collab., 2008 ; Wilson, 2008) tiennent compte de la présence d'un marqueur de négation dans le contexte pour inverser la polarité axiologique locale. Néanmoins, les exemples (7) et (8) montrent que les mécanismes d'inversions de polarité sont parfois plus complexes que la simple présence d'un marqueur de négation. Ainsi, dans (7) et (8), les mêmes marqueurs sont présents dans le même ordre (*peu d'espoir*) mais les mots voisins rendent leur sens opposé (*avoir* ou *avec un*).

- (7) J'ai peu d'espoir [...] → négatif
- (8) [...] avec un peu d'espoir [...] → positif

Par ailleurs, des cas de doubles négations ou de « fausses » négations trompent les méthodes actuelles (9) (10). Pour des analyses à granularité fine, le traitement de la portée des marqueurs de négation représente un défi non résolu à l'heure actuelle.

- (9) Loin de n'être qu'une litanie vulgaire, le texte est tout simplement exquis [...].
- (10) Je ne peux pas ne pas m'indigner de voir les agissements du ministre [...].

1. AIN = Appréciation Implicite Négative, AIP = Appréciation Implicite Positive.

Effets du contexte sur l'axiologie d'une évaluation Les tournures de négation ne sont pas les seuls cas où le contexte modifie l'axiologie d'un passage. Comme évoqué notamment par Thompson et Hunston (2000), certaines constructions linguistiques sont utilisées de manière créative et peuvent rendre évaluatif un mot ou une expression à la base non évaluatif (11), ou neutraliser son sens axiologique (12). Ces types de constructions peuvent se combiner avec une tournure de négation, à l'image de (13), rendant la catégorisation de l'axiologie localement complexe.

(11) C'est un réflexe terriblement latin que de différer la solution des problèmes.
le contexte rend **latin** évaluatif

(12) Les sims 2 : vous allez adorer les détester !
le contexte annule la polarité de **détester**

(13) Ashley était loin de n'être qu'une ravissante idiote.
le contexte tend à rendre **ambigu** le passage

Nous considérons que le contexte d'une évaluation a quatre effets possibles. Il peut :

- **conserver** l'axiologie activée par le vocabulaire. Par exemple, dans *Domenech est **incompétent***, le vocabulaire est négatif et l'évaluation est négative ;
- **inverser** l'axiologie activée par le vocabulaire. Dans *Domenech n'est pas un **bon** entraîneur*, le vocabulaire est positif, mais l'évaluation est négative ;
- **activer** une axiologie. Dans *C'est terriblement latin*, le vocabulaire n'a pas une axiologie stable connue, mais l'évaluation est rendue axiologique de part sa construction. Néanmoins, le contexte ne permet pas de déterminer la polarité de l'axiologie. Celle-ci dépend de l'idéologie de l'énonciateur ;
- **rendre ambiguë** l'axiologie. Dans *C'est une ravissante idiote*, le vocabulaire est positif et négatif, l'évaluation est ambiguë. Dans (14), le vocabulaire est négatif, mais la construction tend à rendre l'évaluation ni positive, ni négative.

(14) La puncture de quelques points, c'est jamais plus désagréable qu'une piqûre de moustique !

Ainsi, les approches uniquement basées sur un vocabulaire de mots évaluatifs ne permettent pas de délimiter et de catégoriser correctement ce type de passages évaluatifs. Le sens d'une évaluation semble contraint à la fois par sa construction grammaticale, par le lexique qu'elle utilise et les traits sémantiques activés par le lexique. De notre point de vue, les méthodes existantes ont donc essentiellement la limite de ne pas tenir compte simultanément de ces trois niveaux de caractéristiques linguistiques :

- **lexicales** (ou **morphologiques**) : *quels sont les lemmes ou les expressions qui activent une évaluation ? voire quelles sont les formes ou les affixes qui activent une évaluation ?*
- **grammaticales** : *quelles sont les constructions qui activent, désactivent ou modifient le sens d'une évaluation ?*
- **sémantiques** : *quelles sont les traits sémantiques qui influent sur la modalité, la configuration d'énonciation et l'axiologie d'une évaluation ?*

Dans la section suivante, nous proposons d'articuler ces trois niveaux pour modéliser l'évaluation à un grain fin.

5.1.2 Modéliser l'évaluation en articulant lexique, grammaire et sémantique

L'idée de représenter les évaluations en articulant et en ordonnant des marqueurs lexico-grammaticaux et sémantiques est présente dans les travaux anglophones de Thompson et Hunston (2000). Les auteurs montrent qu'il existe des structures linguistiques spécifiques à l'expression d'une évaluation. Nous adoptons ce cadre en vue d'une démarche automatique permettant d'acquérir automatiquement les structures évaluatives du français. Cette méthode d'acquisition de structures évaluatives est présentée plus en détails dans la section 5.2. Ces structures sont à la base de la méthode permettant de détecter, délimiter et catégoriser les évaluations (§. 5.3).

5.1.2.1 Des structures linguistiques spécifiques à l'évaluation

Thompson et Hunston (2000) ont entrepris un travail manuel pour montrer qu'il existe, dans la langue, des structures linguistiques utilisées spécifiquement pour exprimer une évaluation. Selon leur définition, ces structures ordonnent des constituants lexico-grammaticaux ou sémantiques. Ils ont menés en corpus une étude des contextes lexicaux et grammaticaux dans lesquels occurrent certains adjectifs ou groupes nominaux évaluatifs afin d'en extraire des régularités. Ils identifient ainsi une liste d'une trentaine de structures d'évaluation pour l'anglais, dont (15) et (16) sont des exemples.

(15) $\underbrace{[\text{cible évaluée}] + [\text{verbe transitif}] + [\text{adjectif évaluatif}]}$
They are interesting, They feel guilty, etc.

(16) $\underbrace{\text{It} + [\text{verbe transitif}] + [\text{groupe nominal évaluatif}] + \text{to} + [\text{cible évaluée}]}$
It was a damn nuisance to have to reboot my DSL, etc.

Dans (15) et (16), [*verbe transitif*] ou [*adjectif évaluatif*] sont des exemples de constituants évaluatifs définis par Thompson et Hunston (2000). En anglais, les *verbes transitifs* sont les verbes comme *to feel, to be, to seem, to become*, etc. Ce type de verbes permet d'introduire fréquemment une évaluation qui portent sur le sujet de la phrase.

D'une façon plus générale, Thompson et Hunston (2000) s'inscrivent dans une démarche de démonstration de faisabilité et non dans une optique exhaustive ou computationnelle. Il s'agit pour eux d'amasser suffisamment d'indices pour juger s'il est possible ou non d'identifier des structures évaluatives. Leurs conclusions renforcent l'idée qu'il existe bien un grand nombre de structures identifiables qui influent sur le sens évaluatif d'un énoncé, mais leur diversité et leur complexité représentent un verrou pour couvrir l'ensemble des phénomènes évaluatifs de la langue. Nous nous positionnons dans ce cadre pour proposer une démarche automatique pour acquérir les structures évaluatives du français.

5.1.2.2 Automatiser le processus d'acquisition des structures évaluatives

Nous poursuivons la démarche entreprise par Thompson et Hunston (2000), mais dans une optique computationnelle et pour le français. À l'image de leurs travaux, il s'agit de :

- définir une liste des différents constituants pouvant intervenir pour exprimer une évaluation ;
- mesurer l'articulation et l'ordre de ces constituants en corpus afin d'en généraliser des structures évaluatives.

Nous présentons les différents types de constituants que nous considérons dans notre travail ci-dessous.

Cinq types de constituants évaluatifs Nous considérons empiriquement qu'il existe cinq catégories principales de constituants impliquées dans une structure évaluative :

- les *marqueurs de personne* (voir chapitre 3, §. 3.1.2) ;
- le *vocabulaire évaluatif* (voir chapitre 3, §. 3.1.3) ;
- les *marqueurs de négation* ;
- les *marqueurs d'intensité* ;
- et *les mots* qui n'appartiennent pas aux quatre premières catégories.

Ces classes de constituants peuvent contenir des mots ou des expressions de catégories grammaticales différentes. Par exemple, *compétent* est un adjectif et *tenir le bon bout* est une expression verbale, ils sont tous les deux des constituants de type *vocabulaire évaluatif*. Pour modéliser ces constituants, nous empruntons la notion de structure de traits définie dans les grammaires d'unification (Shieber, 1986, 1992). Tous les constituants évaluatifs sont définis par une structure de traits unique représentable sous forme de matrice (Fig. 5.1). Cette structure de trait possède trois traits principaux (lexical, grammatical, sémantique) non atomiques conduisant à trois sous structures de traits :

- les traits forme (*forme*) et lemme (*lem*) pour le niveau lexical ;
- le trait partie du discours (*pos*) pour le niveau grammatical ;
- et le type de marqueur (*type*), la modalité (*modalité*), la configuration d'énonciation (*config*) et l'axiologie (*axiol*) pour le niveau sémantique.

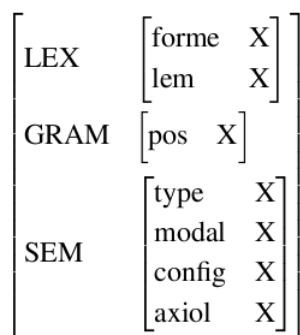


FIGURE 5.1 : Structure de traits d'un constituant évaluatif

Comme le précise Shieber (1992), l'utilisation d'une structure de traits va de pair avec la construction de dictionnaires électroniques ou de lexiques. Dans notre cas, toutes les entrées du lexique de l'évaluation sont considérés comme des constituants de type *vocabulaire évaluatif*. Les marqueurs de négation, les marqueurs d'intensité et les marqueurs de personne proviennent de listes définies manuellement et reportées exhaustivement ci-dessous.

- **Marqueurs de négation** [16] : *ne, point, guère, jamais, nulle part, plus, nullement, aucunement, aucun, nul, pas, personne, rien, difficile de, loin de, sans.*
- **Marqueurs d'intensité** [14] : *tellement, très, énormément, beaucoup, véritable, largement, particulièrement, spécialement, totalement, terriblement, clairement, incroyablement, infiniment, carrément.*
- **Marqueurs de personne** [12] : *moi, je, me, mon, toi, tu, te, ton, nous, notre, vous, votre.*

Les différents lexiques permettent de compléter la valeur des traits sémantiques de chaque constituant. Selon cette idée, une structure évaluative est une suite ordonnée de constituants. Par exemple, l'évaluation contenue dans (17) est composée de trois constituants ordonnés : *pas* (marqueur de négation) + *plus* (mot) + *utiles* (vocabulaire évaluatif). Les structures de traits de chaque constituant sont représentées dans la figure 5.2.

- (17) Reader Touch de Sony : **pas plus utiles** que ça également sauf si vous voulez vous déplacer avec plus de 350 livres.

<table style="border-collapse: collapse; width: 100%;"> <tr> <td style="padding: 5px;">LEX</td> <td style="padding: 5px;">[forme 'pas']</td> </tr> <tr> <td style="padding: 5px;"></td> <td style="padding: 5px;">[lem 'pas']</td> </tr> <tr> <td style="padding: 5px;">GRAM</td> <td style="padding: 5px;">[pos 'adv']</td> </tr> <tr> <td style="padding: 5px;">SEM</td> <td style="padding: 5px;">[type 'neg']</td> </tr> </table>	LEX	[forme 'pas']		[lem 'pas']	GRAM	[pos 'adv']	SEM	[type 'neg']	<table style="border-collapse: collapse; width: 100%;"> <tr> <td style="padding: 5px;">LEX</td> <td style="padding: 5px;">[forme 'plus']</td> </tr> <tr> <td style="padding: 5px;"></td> <td style="padding: 5px;">[lem 'plus']</td> </tr> <tr> <td style="padding: 5px;">GRAM</td> <td style="padding: 5px;">[pos 'adv']</td> </tr> <tr> <td style="padding: 5px;">SEM</td> <td style="padding: 5px;">[type 'mot']</td> </tr> </table>	LEX	[forme 'plus']		[lem 'plus']	GRAM	[pos 'adv']	SEM	[type 'mot']	<table style="border-collapse: collapse; width: 100%;"> <tr> <td style="padding: 5px;">LEX</td> <td style="padding: 5px;">[forme 'utiles']</td> </tr> <tr> <td style="padding: 5px;"></td> <td style="padding: 5px;">[lem 'utile']</td> </tr> <tr> <td style="padding: 5px;">GRAM</td> <td style="padding: 5px;">[pos 'adj']</td> </tr> <tr> <td style="padding: 5px;">SEM</td> <td style="padding: 5px;">[type 'evaluation']</td> </tr> <tr> <td style="padding: 5px;"></td> <td style="padding: 5px;">[modal 'appreciation']</td> </tr> <tr> <td style="padding: 5px;"></td> <td style="padding: 5px;">[config 'implicite']</td> </tr> <tr> <td style="padding: 5px;"></td> <td style="padding: 5px;">[axiol 'positif']</td> </tr> </table>	LEX	[forme 'utiles']		[lem 'utile']	GRAM	[pos 'adj']	SEM	[type 'evaluation']		[modal 'appreciation']		[config 'implicite']		[axiol 'positif']
LEX	[forme 'pas']																															
	[lem 'pas']																															
GRAM	[pos 'adv']																															
SEM	[type 'neg']																															
LEX	[forme 'plus']																															
	[lem 'plus']																															
GRAM	[pos 'adv']																															
SEM	[type 'mot']																															
LEX	[forme 'utiles']																															
	[lem 'utile']																															
GRAM	[pos 'adj']																															
SEM	[type 'evaluation']																															
	[modal 'appreciation']																															
	[config 'implicite']																															
	[axiol 'positif']																															

FIGURE 5.2 : Constituants associés à l'évaluation « *pas plus utiles* »

Dans l'exemple (17), le passage évaluatif *pas plus utiles* tend à annuler (ou rendre ambigu) la polarité du mot évaluatif *utile*. Dans le cadre d'un processus automatique, pour pouvoir identifier et catégoriser un maximum de passages évaluatifs qui se comportent de la même façon que cet exemple, il est nécessaire de s'abstraire le plus possible du lexique utilisé. L'idée est alors de généraliser la structure de traits de ces différents constituants. Nous présentons cette idée dans le paragraphe suivant.

Généraliser la structure de traits des constituants Pour généraliser les exemples observés, il s'agit de ne retenir que les informations lexicales, grammaticales ou sémantiques qui vont permettre de détecter un passage évaluatif et d'en désambigüer la signification dans son contexte. Par exemple, selon nous, le passage évaluatif « *pas plus utiles* » possède les mêmes constituants et la même structure évaluative que les passages évaluatifs (18) et (19) :

- **constituant 1** : la négation (*pas*) est remplacé par un autre adverbe de négation (*jamais*) ;
- **constituant 2** : l'adverbe *plus* n'est jamais modifié ;
- **constituant 3** : un mot du vocabulaire évaluatif (*utile*) est remplacé par un autre mot du vocabulaire évaluatif ayant ou non la même axiologie (*désagréable*, *mauvais*).

- (18) La puncture de quelques points, c'est **jamais plus désagréable** qu'une piqûre de moustique !

- (19) [Le Nutella] c'est **pas plus mauvais** que beaucoup d'autres aliments. Si vous avez envie de Nutella, surtout ne vous privez pas !

Dans les trois exemples présentés, la construction particulière des passages évaluatifs tend à annuler la polarité du mot évaluatif activé. De part cette observation, nous pouvons induire que la suite de constituants évaluatifs de la figure 5.2 possède en fait une structure évaluative sous-jacente, modélisable par la structure 5.3. Dans cette figure :

- les traits *lexicaux* du constituant 1 ont été supprimés. Seuls les traits [GRAM : pos : adv] et [SEM : type : neg] (adverbe de négation) contraignent le sens de ce passage évaluatif ;
- tous les traits du constituant 2 ont été conservés ;
- les traits *lexicaux* du constituant 3 ont été supprimés, ainsi que le trait *sémantique* axiologie. En d'autres termes, seul le fait que le troisième constituant soit un adjectif évaluatif d'appréciation implicite est une donnée importante à conserver pour représenter la structure évaluative sous-jacente des exemples cités.

D'après les exemples observés, la structure évaluative représentée par la figure 5.3 tend à annuler la polarité axiologique portée par le troisième constituant de la structure évaluative. L'ajout de méta-données sur les structures évaluatives pour caractériser leur signification est un point que nous développons dans la section suivante. Nous proposons une procédure pour acquérir automatiquement ces structures évaluatives à partir d'exemples issus d'un corpus annoté et de quelques règles de généralisation définies à la main.

$$\left[\begin{array}{l} \text{GRAM} \left[\begin{array}{l} \text{pos} \quad \text{'adv'}$$

FIGURE 5.3 : Structure évaluative généralisée (*pas plus utiles, jamais plus mauvais, etc.*)

5.2 Acquisition de structures évaluatives

Pour analyser les évaluations, une première étape consiste à acquérir les structures linguistiques utilisées fréquemment ou spécifiquement pour exprimer une évaluation. Si ces structures semblent fortement dépendantes du lexique, nous supposons que certains constituants ou certains traits peuvent être généralisables. Dans cette section, nous proposons une procédure automatique pour accomplir cet objectif. Celle-ci se scinde en trois étapes :

- l'extraction d'exemples de passages évaluatifs connus à partir d'un corpus annoté (§. 5.2.1) ;
- la transformation de ces exemples en structures évaluatives, à partir de règles (§. 5.2.2) ;
- l'annotation de méta-données pour caractériser ces structures évaluatives (§. 5.2.3) .

Enfin, nous présentons des données chiffrées sur les structures évaluatives obtenues (§. 5.2.4). Celles-ci sont à la base de la méthode proposée pour détecter, délimiter et catégoriser les évaluations (§. 5.3).

5.2.1 Extraction des passages évaluatifs à partir d'un corpus annoté

Pour identifier quelles sont les structures linguistiques utilisées pour exprimer une évaluation, notre approche nécessite un corpus où les passages évaluatifs ont été annotés. Nous utilisons pour cela le corpus d'entraînement Blogoscopie (par la suite, *le corpus*) (voir chapitre 4, §. 4.1). Ce corpus contient 200 documents et représente 4 945 exemples de passages évaluatifs annotés manuellement. La première phase consiste à compléter la structure de traits de tous les mots du corpus. Plusieurs processus de pré-traitements sont initialement appliqués au corpus :

- une segmentation en mots ;
- une lemmatisation ;
- un étiquetage grammatical (partie du discours) ;
- une projection des entrées des lexiques de l'évaluation, des marqueurs de personne, de négation et d'intensité. La projection de ces lexiques s'effectue à partir des lemmes de chaque mot.

Cette chaîne de traitement est assemblée à partir du cadre UIMA dédié à la structuration automatique de documents et développé par Apache et IBM (Ferruci et Lally, 2004). Celui-ci assure l'inter-opérabilité de chaque module traitement, la structuration et la réutilisation des annotations posées. La lemmatisation et l'étiquetage grammatical est réalisé par une encapsulation du TreeTagger

(Schmid, 1994) dans UIMA. La projection des lexiques utilise un composant développé dans UIMA à cette occasion.

À la suite de ces pré-traitements, nous disposons d'un corpus où figure des annotations posées manuellement (voir chapitre 4, §. 3.1) et des annotations posées automatiquement. Les structures de traits obtenues par l'annotation automatique sont comparables avec la structure de traits des constituants évaluatifs définis précédemment (§. 5.1.2.2). Pour chacun des 4 945 exemples de passages évaluatifs, nous collectons automatiquement la suite des constituants et leur structure de traits. À partir des annotations manuelles, chaque exemple est également rattaché à :

- sa modalité ;
- sa configuration d'énonciation ;
- et son axiologie.

Par exemple, le passage évaluatif *je l'aime* rencontré dans le corpus est composé de trois constituants (Fig. 5.4). Ce passage correspond à une évaluation de modalité **appréciation**, de configuration **explicite** et d'axiologie **positive**.

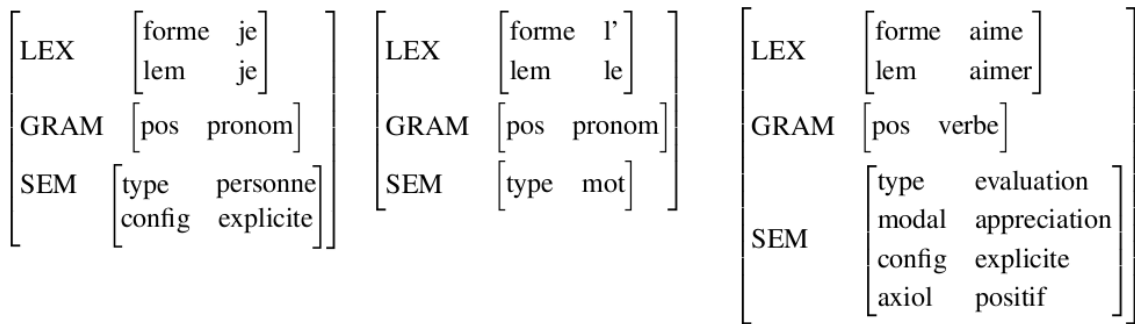


FIGURE 5.4 : Structure évaluative non généralisée issue du corpus d'entraînement

L'étape suivante consiste à généraliser la structure de traits des différents constituants d'une évaluation pour obtenir une structure évaluative généralisée grâce à un ensemble de règles définies manuellement.

5.2.2 Transformation en structures évaluatives généralisées

La phase de généralisation se rapproche partiellement des travaux de Karttunen (1984) sur la généralisation de deux structures de traits. Il définit la généralisation comme étant l'opération permettant de prendre en compte certains traits dont les valeurs ne sont pas incompatibles et d'ignorer les autres. Dans notre cadre, nous souhaitons ignorer les traits qui contraignent trop le lexique utilisé pour exprimer une évaluation. En d'autres termes, à partir d'un exemple d'évaluation, nous souhaitons relâcher certaines contraintes lexicales, grammaticales ou sémantiques qui n'ont pas d'effet sur la nature d'une structure évaluative. L'objectif recherché est de pouvoir plus tard détecter, délimiter et catégoriser des exemples non rencontrés dans le corpus d'entraînement, mais qui se comportent de façon analogue. Par exemple, l'évaluation de (20), issue du corpus d'entraînement, permet d'analyser l'évaluation (21), jamais rencontrée dans le corpus d'entraînement, car leur structure est analogue. Cette structure conserve la polarité axiologique du mot évaluatif activé (malgré la présence du marqueur de négation *non* dans le contexte) et correspond à une évaluation de type appréciation explicite.

- (20) Non seulement j'aime beaucoup J.K Rowling [...] – appréciation explicite positive
- (21) Non seulement nous haïssons le repassage [...] – appréciation explicite négative

Nous définissons un processus pour généraliser des exemples d'évaluations à partir de quatre règles. Ces règles sont guidées par des observations et des choix empiriques, nous rediscutons de cet aspect et de ces limites dans les perspectives de ce chapitre (§. 5.5). Ces règles reposent sur les deux types d'opérations suivantes :

- **substitution** : une ou plusieurs valeurs de traits d'un constituant peuvent être substituées par d'autres valeurs. Il est donc possible d'ignorer ces traits ;
- **ajout ou suppression** : un ou plusieurs constituants peuvent être ajoutés ou supprimés sans que cela n'affecte le sens de la structure évaluative. Il est donc possible d'ignorer ces constituants.

Nous présentons ces règles et donnons pour chacune d'elle des exemples d'évaluations.

5.2.2.1 Règles de substitution

Règle 1 : « Substitution d'un mot par un mot de la même classe sémantique et de la même catégorie grammaticale »

La première règle de substitution concerne les quatre catégories de constituants suivantes :

- les marqueurs de personne ;
- les expressions ou les mots évaluatifs ;
- les marqueurs de négation ;
- les marqueurs d'intensité.

Nous considérons intuitivement qu'une entrée d'un lexique peut être remplacée par une autre entrée du même lexique si elles partagent la même catégorie grammaticale. Le sens global de la structure évaluative n'étant pas affectée par cette substitution, la forme et le lemme d'un mot sont des traits qui peuvent être ignorés par les structures évaluatives pour les quatre types de constituants concernés. En revanche, nous conservons tous les autres traits. En appliquant cette règle, après généralisation, les exemples (22) et (23) sont modélisés par la même structure évaluative représentée dans la figure 5.5.

(22) Il n'y a **pas de défaite plaisante**.

(23) Il n'y a **jamais de défaite encourageante**.

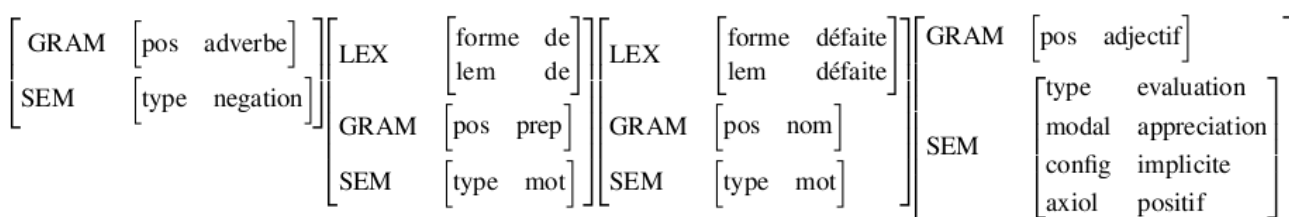


FIGURE 5.5 : Généralisation des mots de la même classe sémantique et grammaticale

Règle 2 : « Substitution d'un mot évaluatif axiologisé par un mot évaluatif ayant une axiologie différente »

La deuxième règle de substitution concerne uniquement les mots évaluatifs. Dans les évaluations contenant un seul mot évaluatif axiologisé (positif ou négatif), ce mot peut être remplacé par un autre mot évaluatif d'une axiologie inverse, mais de catégorie grammaticale identique. La nature de la structure évaluative n'est pas affectée par cette substitution, l'axiologie du mot évaluatif est donc

ignorée. Après application des règles 1 et 2, les exemples (24) et (25) sont modélisés par la même structure évaluative (Fig. 5.6).

(24) Leproux n'éprouve pas le moindre regret pour Heinze.

(25) Il ne ressent pas le moindre plaisir à revenir en France.

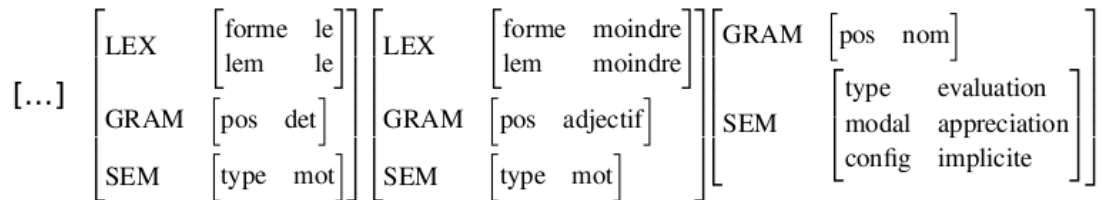


FIGURE 5.6 : Généralisation de l'axiologie d'un mot évaluatif

En plus de ces deux règles de substitution, nous utilisons deux règles d'ajout ou de suppression pour généraliser les exemples observés.

5.2.2.2 Règles d'ajout ou de suppression

Règle 3 : « Ajout ou suppression d'un marqueur d'intensité »

La première règle d'ajout ou de suppression concerne uniquement les marqueurs d'intensité. Ces marqueurs d'intensité peuvent être ajoutés ou supprimés à différentes positions dans une évaluation sans que cela n'affecte la structure évaluative. Ce type de marqueur est donc ignoré. Les exemples (26) et (27) sont ainsi modélisés par la même structure évaluative :

[VER : être] + [DET : un] + [NOM : facteur] + [NOM : mot évaluatif d'appréciation implicite].

(26) L'école est un facteur de stress.

(27) l'islam politique turc est véritablement un facteur de discorde.

Règle 4 : « Ajout ou suppression de pronoms adverbiaux et de pronoms personnels »

Enfin, la seconde règle d'ajout ou de suppression concerne uniquement les mots de type pronoms adverbiaux (*en*, *y*) ou pronoms personnels. De la même façon, ces mots sont potentiellement présents à différentes positions d'une évaluation. S'ils sont ignorés, le sens de la structure évaluative n'en est pas affecté. Les exemples (28) et (29) sont modélisés par la même structure évaluative :

[PRO : personne explicite] + [PRO : negation] + [PRO : personne explicite] + [VER : mot évaluatif d'appréciation] + [PRO : negation].

(28) Jossinet : « Je ne m'en lasse pas ».

(29) Thierry Henry : « Je ne me plains pas ».

Nous présentons les résultats obtenus après l'application de ces règles de généralisation dans la section 5.2.4. Avant cela, nous associons des méta-données à chacune des structures. Ces méta-données seront ensuite utilisées pour détecter et catégoriser des passages évaluatifs inconnus. Nous précisons ce point dans la sous-section suivante.

5.2.3 Annotation des méta-données sur les structures évaluatives

L'annotation des structures évaluatives consiste à ajouter des informations sur chaque structure à partir du corpus d'entraînement. L'objectif est de permettre ultérieurement de détecter et catégoriser des passages évaluatifs. Pour chaque structure, il s'agit de déterminer si elle :

- exprime une certaine modalité parmi : appréciation, opinion ou accord/désaccord ;
- exprime une certaine configuration d'énonciation parmi : explicite ou implicite ;
- a un certain effet sur l'axiologie parmi : conserver, inverser, activer, annuler/rendre ambiguë.

Déterminer les aspects sémantiques et énonciatifs les plus probables La probabilité qu'une structure modélise une des trois modalités est donnée par l'équation 5.1.

$$P(s, \text{modalite}) = \frac{\text{NB}(s, \text{modalité})}{\text{NB}(s, \text{modalité}) + \text{NB}(s, \text{autre modalité})} \quad (5.1)$$

où :

- $\text{NB}(s, \text{modalite})$ est le nombre de fois où la structure s correspond une modalité particulière ;
- $\text{NB}(s, \text{autre modalite})$ est le nombre de fois où la structure s correspond à une autre modalité ;
- $P(s, \text{modalite})$ est la probabilité que la structure s corresponde à la modalité considérée.

De la même façon, la configuration d'énonciation la plus probable pour une structure s_i est déterminée avec l'équation 5.2.

$$P(s, \text{config}) = \frac{\text{NB}(s, \text{config})}{\text{NB}(s, \text{config}) + \text{NB}(s, \text{autre config})} \quad (5.2)$$

Déterminer l'effet de la structure sur l'axiologie de l'évaluation Enfin, nous avons remarqué qu'une structure pouvait avoir quatre types d'actions sur la polarité axiologique d'une évaluation. Elle peut conserver, inverser, activer ou annuler/rendre ambiguë la polarité axiologique du constituant évaluatif présent dans la structure. À l'aide du corpus de référence et de l'équation 5.3, nous déterminons statistiquement quel est le cas le plus fréquent. La structure s :

- **conserve** la polarité axiologique si les passages évaluatifs concernés par s dans le corpus de référence ont la même polarité axiologique que le constituant évaluatif présent dans la structure ;
- **inverse** la polarité axiologique si les passages évaluatifs concernés par s dans le corpus de référence ont une polarité axiologique inverse ;
- **active** la polarité axiologique si les passages évaluatifs concernés par s dans le corpus de référence ont une polarité axiologique alors qu'il n'y a pas de constituant évaluatif axiologique ;
- **annule ou rend ambiguë** la polarité axiologique si :
 - la structure s correspond à davantage de passages non annotés comme évaluatifs dans le corpus de référence ;
 - ou s contient plusieurs constituants évaluatifs dont l'axiologie est différente.

$$P(s, \text{action}) = \frac{\text{NB}(s, \text{action})}{\text{NB}(s, \text{action}) + \text{NB}(s, \text{non-action})} \quad (5.3)$$

où :

- $\text{NB}(s, \text{action})$ est le nombre de fois où s réalise l'action action dans le corpus de référence ;
- $\text{NB}(s, \text{non-action})$ est le nombre de fois où s réalise une autre action dans le corpus de référence ;
- $P(s_i, \text{action})$ est la probabilité que la structure s ait une action action sur l'axiologie.

Par ailleurs, un identifiant est attribué à chaque structure évaluative. L'objectif est de pouvoir suivre lors d'expérimentations quel est le nombre de structures mis en jeu, quelles sont les structures les plus fréquentes, quelles sont les structures qui provoquent du bruit, etc. Dans la sous-section suivante, nous présentons les résultats obtenus par l'extraction et la généralisation des structures évaluatives.

5.2.4 Résultats et discussion

À partir des 4 945 passages évaluatifs, la méthode d'extraction et de généralisation obtient 2 031 structures évaluatives. À première vue, nous observons que le pouvoir de généralisation des règles présentées précédemment n'est pas très élevé (en moyenne environ 1 structure pour 2,4 évaluations). Les évaluations exprimées dans le corpus d'entraînement ont donc des formes relativement distinctes par rapport aux capacités des règles de généralisation utilisées. La figure 5.7 montre que si certaines structures sont assez fréquentes dans le corpus d'entraînement, environ un tiers des structures ne sont présentes qu'une ou deux fois. De ce point de vue, la taille du corpus d'entraînement nous semble un aspect critique. La figure 5.8 montre le nombre de structures extraites en fonction du nombre de passages évaluatifs traités. La droite en rouge présente sur ce graphique est un repère indiquant le nombre de structures évaluatives obtenues sans la phase de généralisation. Avec la généralisation, la pente de la courbe décroît après environ 1 200 exemples, mais le nombre de structures continuent ensuite d'augmenter régulièrement. Il serait intéressant d'avoir un plus grand nombre d'exemples d'évaluations pour suivre le comportement de cette courbe, mais le problème du coût de création d'un tel corpus d'exemples demeure.

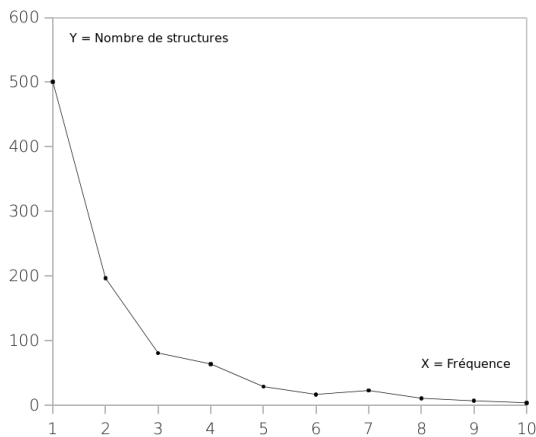


FIGURE 5.7 : Nombre de structures en fonction de leur fréquence dans le corpus

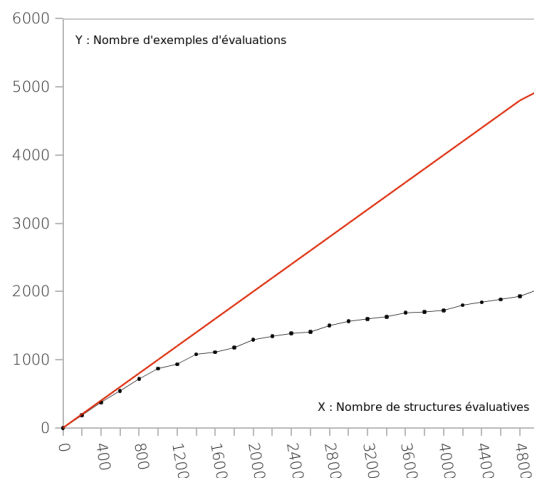


FIGURE 5.8 : Évolution du nombre de structures par rapport au nombre d'exemples d'évaluations

La répartition des structures évaluatives selon leur catégorie la plus probable est donnée dans les tableaux 5.1, 5.2 et 5.3. Ces tableaux montrent notamment que les appréciations implicites sont particulièrement riches en structures évaluatives différentes (1 536 structures). Le tableau 5.3 montre qu'à partir des 4 945 exemples d'évaluations, la méthode a extrait 870 structures qui conservent la polarité axiologique, 563 structures qui l'inversent, 391 structures qui l'activent et 34 qui la rend ambiguë ou qui l'annule. Notons que le nombre important de structures évaluatives qui activent une

polarité axiologie est influencé par la couverture du lexique de mots évaluatifs utilisé. Le lexique de l'évaluation n'étant pas exhaustif, à chaque fois qu'une évaluation contient un mot évaluatif absent du lexique, la structure est considérée comme « activant » une polarité. Nous rediscutons de cet aspect ultérieurement.

modalité	opinion	appréciation	accord/désaccord
Nombre de structures	95	1 858	78

TABLE 5.1 : Structures évaluatives acquises selon leur modalité

configuration	explicite	implicite
Nombre de structures	495	1536

TABLE 5.2 : Structures évaluatives acquises selon leur configuration d'énonciation

action	conserve	inverse	active	annule/ambiguë
Nombre de structures	870	563	391	34

TABLE 5.3 : Structures évaluatives acquises selon leur action sur l'axiologie

Dans cette section, nous avons proposé une méthode pour modéliser les évaluations sous forme de structures évaluatives. Ces structures généralisent les exemples observés dans un corpus d'entraînement. Si le taux de généralisation nous semble à première vue assez bas, nous observons néanmoins plusieurs cas intéressants de structures évaluatives très proches permettant de distinguer des évaluations ayant un sens opposé. Par exemple, les structures suivantes ont un sens opposées, alors qu'elles utilisent des marqueurs quasiment identiques.

- [ADV : neg] + [VER : être] + [DET : ce] + [ADV : neg] + [ADJ : mot évaluatif d'appréciation]
- [DET : ce] + [ADV : neg] + [VER : être] + [ADJ : mot évaluatif d'appréciation]

La première structure n'inverse pas la polarité (voir l'exemple (30)), alors que la seconde l'inverse (voir l'exemple (31)).

(30) **N'est-ce pas** intéressant d'avoir des gens qui donnent des avis contraires [...].

(31) **Ce n'est pas** intéressant de cotiser.

En s'appuyant sur les structures évaluatives du français apprises, nous proposons dans la section suivante une méthode pour détecter, délimiter et catégoriser les évaluations d'un corpus non annoté. Nous comparons cette méthode à une approche basée uniquement sur le lexique. L'évaluation de ces deux méthodes sur un corpus de test (§.5.4) nous permet de discuter de l'intérêt et des limites de ces structures évaluatives.

5.3 Délimitation et catégorisation des évaluations

Notre objectif est de délimiter et catégoriser les passages évaluatifs d'un corpus non annoté. Dans cette section, nous présentons deux méthodes pour ces tâches :

- notre méthode principale qui utilise les structures évaluatives acquises précédemment et des ressources lexicales (par la suite, *méthode par structures évaluatives*) (§. 5.3.1) ;
- une méthode lexicale (*baseline*), utilisant uniquement des ressources lexicales, qui puisse nous servir de point de comparaison lors de nos expérimentations (§. 5.3.2).

En amont de ces deux méthodes, plusieurs tâches de pré-traitements sont réalisées sur le corpus :

- une segmentation en mots ;
- une lemmatisation des mots ;
- une catégorisation grammaticale des mots ;
- un découpage du texte en syntagmes non récursifs ;
- une projection des entrées du lexique de l'évaluation, des marqueurs de négation, de personne et d'intensité.

Pour ces différents pré-traitements, nous avons recours aux mêmes composants que précédemment et à une ré-implémentation de l'algorithme de chunking proposé par Vergne et Giguet (1998) adapté à UIMA. À la suite de ces pré-traitements, chaque mot du corpus possède une structure de traits comparable avec la structure de traits des constituants évaluatifs définis dans les sections précédentes (§. 5.1.2.2).

5.3.1 Méthode par structures évaluatives

Pour détecter, délimiter et catégoriser les évaluations, nous proposons une méthode qui exploite le lexique de l'évaluation et l'ensemble S des structures évaluatives acquises précédemment (§. 5.2). Dans cette méthode, nous distinguons deux phases :

- la détection et la délimitation des évaluations ;
- la catégorisation des aspects sémantiques et énonciatifs des évaluations.

5.3.1.1 Détection et délimitation

Pour détecter et délimiter les évaluations, nous appliquons un algorithme de recherche de sous-chaîne pour parcourir l'ensemble d'une phrase à analyser. Lorsqu'une sous-chaîne c possède des traits compatibles avec une structure s ($s \in S$), la méthode pose une annotation sur les bornes des syntagmes non récursifs qui englobent c . Nous considérons que les passages évaluatifs sont tous d'un niveau intra-phrastique. Toutes les phrases d'un texte à analyser sont parcourues une fois, de façon itérative (algorithme 1).

D'après cet algorithme, aucune annotation d'évaluation ne peut se superposer à une autre. Dans le cas où plusieurs structures évaluatives correspondent à des passages qui se chevauchent, la priorité est donnée à la structure la plus grande. Lorsqu'un passage est annoté comme étant une évaluation, l'identifiant de la structure évaluative qui a servi à détecter ce passage est conservé. Cette donnée est utilisée pour catégoriser le passage évaluatif.

5.3.1.2 Catégorisation sémantique et énonciative

La catégorisation des passages évaluatifs délimités à l'étape précédente se scinde en 3 aspects : la modalité, la configuration d'énonciation et l'axiologie. Le principe de catégorisation de la modalité et de la configuration d'énonciation est simple : soit s , la structure ayant permis de détecter et délimiter une évaluation e . À partir des méta-données de s , la modalité et la configuration d'énonciation la plus

Algorithme 1 Détection des évaluations

ENTRÉES: S : Ensemble des phrases du corpus ; P : Ensemble des structures d'évaluation connues
entier t : taille de la phrase courante (en nombre de mots)
entier c_1 : curseur identifiant la position du mot de début dans la phrase
entier c_2 : curseur identifiant la position du mot de fin dans la phrase

Pour tout $s \in S$ **faire**
 $c_1 \leftarrow 0$
 $c_2 \leftarrow t - 1$
Tant que $c_1 \leq t - 1$ **faire**
 Tant que $c_1 \leq c_2$ **faire**
 booleen $\text{find} \leftarrow \text{match}(s(c_1, c_2), P)$ /*Existe t-il une structure compatible pour le passage
 (c_1, c_2) ?*/
 Si find **Alors**
 $\text{annotation}(c_1, c_2)$ /*Annotation du passage (c_1, c_2) : Évaluation */
 $c_1 \leftarrow c_2$ /*Ré-initialisation du curseur de début à la suite de l'évaluation détectée*/
 $c_2 \leftarrow t - 1$ /*Ré-initialisation du curseur de fin*/
 Sinon
 $c_2 \leftarrow c_2 - 1$ /*Déplacement du mot de fin : priorité aux plus grandes chaînes */
 Finsi
 fin Tant que
 $c_1 \leftarrow c_1 + 1$ /*Déplacement du mot de début : priorité au sens de lecture */
fin Tant que
fin pour

probable sont celles qui sont conservées pour catégoriser e . Deux traits sont ajoutés à l'annotation e pour indiquer sa modalité et sa configuration d'énonciation.

Pour la catégorisation de l'axiologie, les structures n'indiquent pas directement la polarité d'une évaluation mais indiquent seulement si :

- elles conservent la polarité du mot évaluatif contenu dans le passage ;
- elles inversent la polarité du mot évaluatif contenu dans le passage ;
- elles rendent ambiguë la polarité du passage ;
- elles activent une polarité pour ce passage sans pouvoir déterminer la polarité (ambiguë).

Soit s , la structure ayant permis de détecter et délimiter une évaluation e . Si e contient un constituant évaluatif **positif** (respectivement **néгатif**) :

- et si s_i **conserve** la polarité alors l'évaluation e est **positive** (resp. **néгатive**) ;
- et si s_i **inverse** la polarité alors l'évaluation e est **néгатive** (resp. **positive**) ;
- et si s_i rend **ambiguë** la polarité du passage alors l'évaluation e est **ambiguë** ;
- et si s_i **active** une polarité alors l'évaluation e est **ambiguë**.

Un trait est ajouté à l'annotation e pour indiquer son axiologie. Pour pouvoir mieux évaluer l'intérêt et les limites de cette méthode par structures évaluatives, nous proposons d'établir une baseline avec une méthode utilisant uniquement des ressources lexicales et quelques règles de catégorisation.

5.3.2 Méthode lexicale (*baseline*)

Les tâches de délimitation et de catégorisation des évaluations n'ayant jamais été traitées à granularité fine pour le français, il n'existe pas de niveau de référence auquel comparer notre méthode. Néanmoins, selon nous, pour pouvoir mieux évaluer l'intérêt de la méthode par structures évaluatives,

un point de comparaison s'avère nécessaire. Pour cela, nous proposons une méthode baseline (par la suite, *méthode lexicale*). Celle-ci reprend la distinction entre détection et délimitation des évaluations d'une part, et catégorisation d'autre part.

5.3.2.1 Détection et délimitation

Pour détecter et délimiter les évaluations, la méthode lexicale s'appuie sur la projection du lexique de l'évaluation et la segmentation d'un texte en syntagmes non récursifs. Cette méthode consiste à considérer que chaque syntagme qui contient un mot du lexique de l'évaluation est un passage évaluatif. Une annotation e est alors posée sur ce syntagme par la méthode de base.

(32) | Je suis | véritablement **admirative** | de l'univers | créé | par J.K Rowling |.

5.3.2.2 Catégorisation sémantique et énonciative

Modalité et configuration d'énonciation La catégorisation de la modalité et de la configuration d'énonciation par la méthode lexicale sont effectuées à partir de la procédure à base de règles suivantes :

- s'il existe au moins un marqueur d'appréciation dans l'annotation e alors l'évaluation est une *appréciation* ;
- sinon, s'il existe un marqueur d'opinion dans e , l'évaluation est une *opinion* ;
- sinon, s'il existe un marqueur d'accord/désaccord dans e , l'évaluation est un *accord/désaccord*.

L'ordre de préférence entre les appréciations, les opinions et les accord/désaccord est fixé de cette façon de part la plus grande fréquence des appréciations, puis des opinions, dans le corpus d'entraînement Blogoscopie. Le principe est identique pour la configuration d'énonciation :

- l'évaluation est explicite si elle contient au moins un marqueur de personne ou si la phrase se termine par un point d'exclamation ;
- sinon l'évaluation est implicite.

Axiologie Pour la catégorisation de l'axiologie, nous proposons deux stratégies :

- **stratégie A** : si l'évaluation contient un mot du lexique de l'évaluation positif (respect. négatif) alors l'évaluation est positive (respect. négative). Sinon l'évaluation est ambiguë.
- **stratégie B** : si l'évaluation contient au moins un marqueur de négation, alors l'axiologie obtenue par la stratégie A est inversée. Si l'évaluation était ambiguë, elle reste ambiguë.

Dans cette section, nous avons proposé deux méthodes pour détecter, délimiter et catégoriser les évaluations exprimées dans un corpus. La première utilise des structures évaluatives et des ressources lexicales. Nous la mettons en regard avec une méthode baseline plus simple fondée sur les mêmes ressources lexicales et des règles de catégorisation. Dans la section suivante, nous expérimentons ces deux méthodes sur un corpus de test et discutons des résultats obtenus, de l'intérêt et des limites de la première méthode.

5.4 Expérimentation : intérêts et limites des structures évaluatives

Afin de mesurer l'intérêt de la méthode par structures évaluatives proposée pour détecter, délimiter et catégoriser les évaluations, nous l'expérimentons sur un corpus de test. Nous utilisons le corpus de test Blogoscopie qui regroupe : 25 documents contenant le nom de *Sarah Palin* et 25 documents évoquant les *sushis*. Ce corpus représente 955 passages évaluatifs annotés manuellement. L'annotation manuelle de ce corpus sert de référence pour évaluer notre méthode et la méthode lexicale baseline. Nous présentons tout d'abord les mesures choisies pour évaluer les méthodes (§. 5.4.1), puis nous discutons des résultats obtenus en montrant les intérêts et les limites de la méthode par structures évaluatives. Nous évoquons également notre participation au DÉfi Fouille de Textes 2009 (DEFT'09) avec cette méthode pour montrer sa portabilité à d'autres types de corpus et à d'autres thématiques.

5.4.1 Protocole d'évaluation

Nous souhaitons évaluer la méthode par structures évaluatives en comparaison avec la méthode de base sur deux points :

- la détection et la délimitation des passages évaluatifs (§. 5.4.1.1) ;
- et la catégorisation de la modalité, de la configuration d'énonciation et de l'axiologie des évaluations (§. 5.4.1.1).

5.4.1.1 Évaluer la détection et la délimitation des évaluations

Le premier aspect que nous évaluons est la capacité des méthodes à détecter et délimiter les passages évaluatifs. Nous distinguons deux niveaux :

- la **détection** : la méthode automatique réussit à détecter un passage évaluatif, mais les bornes sont potentiellement incorrectes ;
- la **délimitation** : la méthode automatique réussit à détecter un passage évaluatif et les bornes sont correctes.

Détection Nous évaluons la détection des passages évaluatifs avec la mesure de précision (5.4) et de rappel (5.5), par rapport à une annotation de référence.

$$\text{Précision} = \frac{\text{nb d'évaluations correctement détectées}}{\text{nb d'évaluations détectées automatiquement}} \quad (5.4)$$

$$\text{Rappel} = \frac{\text{nb d'évaluations correctement détectées}}{\text{nb d'évaluations annotées dans la référence}} \quad (5.5)$$

Dans nos expérimentations, pour être considéré bien détecté, un passage évaluatif annoté automatiquement doit au moins recouvrir un syntagme appartenant à l'annotation de référence. Par exemple, dans (33), bien que les bornes de l'annotation automatique soit inexactes, celle-ci recouvre une partie de l'annotation de référence. Elle est donc considérée comme correctement détectée.

(33) | j'ai | malgré tout | une véritable reconnaissance | envers J.K Rowling | car |
automatique
référence
 elle a permis | à mon fils | de s'intéresser | aux livres |.

Délimitation Nous évaluons ensuite la délimitation des passages évaluatifs également avec la mesure de précision (5.6) et de rappel (5.7), par rapport à une annotation de référence.

$$\text{Précision} = \frac{\text{nb d'évaluations correctement délimitées}}{\text{nb d'évaluations délimitées automatiquement}} \quad (5.6)$$

$$\text{Rappel} = \frac{\text{nb d'évaluations correctement délimitées}}{\text{nb d'évaluations annotées dans la référence}} \quad (5.7)$$

Pour être considérée bien délimitée, une évaluation annotée automatiquement doit :

- soit recouvrir exactement le ou les syntagme(s) contenant l'annotation de référence (34) ;
- soit recouvrir les syntagmes contenant la référence avec une certaine tolérance à l'erreur : l'annotation automatique peut contenir un syntagme de plus ou un syntagme en moins (35).

Nous justifions la tolérance à l'erreur par le fait que les annotations manuelles de référence n'ont pas toujours respectées les mêmes critères pour borner les évaluations. Prendre un syntagme de plus ou un syntagme de moins n'entraîne pas nécessairement de conséquences négatives. Dans le cas contraire, les erreurs seront rendues visibles à l'étape de catégorisation. Tout autre cas est considéré erroné.

$$(34) \quad | \text{ la Wii } | \text{ risque vite } | \text{ de devenir } \underbrace{\text{introuvable}}_{\substack{\text{référence} \\ \text{automatique}}} |.$$

$$(35) \quad | \text{ J'avoue } | \underbrace{\text{ je suis pas } | \text{ vraiment fan } | \text{ [..] }}_{\substack{\text{automatique} \\ \text{référence}}} | \text{ en ce qui concerne } | \text{ les livres } |.$$

5.4.1.2 Évaluer la catégorisation des évaluations

Le deuxième aspect que nous évaluons est la capacité des méthodes à catégoriser les trois caractéristiques sémantiques et énonciatives des évaluations : la modalité, la configuration d'énonciation et l'axiologie. Pour évaluer ces trois catégorisations, nous ne tenons pas compte des évaluations non détectées ou des évaluations qui n'auraient pas dû être détectées, pour lesquelles la tâche de catégorisation est d'avance un échec. La méthode par structures évaluatives et les méthodes de base sont évaluées par des scores de précision et de rappel. Dans les équations 5.8 et 5.9, c_i représente une modalité, une configuration ou une axiologie particulière.

$$\text{Précision}(c_i) = \frac{\text{nb de } c_i \text{ correctement catégorisées}}{\text{nb de } c_i \text{ catégorisées automatiquement}} \quad (5.8)$$

$$\text{Rappel}(c_i) = \frac{\text{nb de } c_i \text{ correctement catégorisées}}{\text{nb de } c_i \text{ annotées par la référence}} \quad (5.9)$$

À partir de ces mesures, nous présentons dans les sections suivantes les résultats de la méthode par structures évaluatives et les méthodes de base.

5.4.2 Une bonne délimitation des évaluations

Le premier aspect que nous évaluons est la capacité des structures évaluatives à détecter une évaluation et à délimiter correctement ses bornes dans l'énoncé. Les résultats de la méthode par structures et de la méthode baseline sont résumés dans le tableau 5.4.

<i>DÉTECTION</i>	Précision	Rappel
Méthode par structures	91,5 % (495/541)	51,8 % (495/955)
Méthode baseline	91,1 % (470/516)	49,2 % (470/955)
<i>DÉLIMITATION</i>	Précision	Rappel
Méthode par structures	88,4 % (478/541)	50,1 % (478/955)
Méthode baseline	68,6 % (354/516)	37,1 % (354/955)

TABLE 5.4 : Résultats de la détection et de la délimitation des évaluations

Détection Concernant la détection des passages évaluatifs, la méthode baseline et la méthode par structures évaluatives ont des résultats très proches. En effet, la méthode par structures est basée sur le lexique, il n'est donc pas surprenant qu'une approche lexicale un peu plus naïve puisse avoir des résultats voisins pour la simple détection des évaluations. Les structures évaluatives permettent néanmoins de détecter 25 évaluations supplémentaires. Ces exemples correspondent à des évaluations qui sont activées non pas par le lexique, mais par une structure particulière. Dans (36), l'évaluation est détectée grâce à la structure évaluative [V :être] + [ADV :intensité :littéralement] + [V].

(36) [...] | les sushis | sont littéralement arrivés | 2h30 après l.

└──────────────────┘
automatique
└──────────────────┘
référence

Globalement, le lexique est assez peu ambigu (la précision est d'environ 91 %) dans le corpus de test. Les erreurs de précision proviennent essentiellement de quelques entrées du lexique de l'évaluation qui peuvent également être non évaluatives dans certains contextes ou qui peuvent être considérées comme évaluatives à cause d'erreurs lors des pré-traitements (37) (38).

(37) A San Diego (California) mon beau père roulait pour McCain.

(38) Bon, évidemment tout est conçu pour les professionnels.

Dans le cas de (38), le mot *bon* est toujours étiqueté comme un adjectif par le TreeTagger. Cette erreur d'étiquetage grammatical se propage au niveau de la projection du lexique évaluatif. Dans (37), l'absence de tiret entre beau et père induit une erreur lors du découpage en mots. Ces erreurs sont limitées à quelques entrées lexicales, mais elles se répètent assez fréquemment. Les mots *beau* et *bon* représentent environ 15 % des erreurs de précision sur le corpus de test. Des traitements de désambiguïsation spécifiques aux entrées lexicales fréquemment bruyantes sont une perspective envisageable.

Sur le simple aspect de la détection, la méthode par structures ne se démarquent pas beaucoup plus qu'une simple approche lexicale. En revanche, la délimitation des évaluations est un aspect sur lequel la méthode par structures apporte une amélioration.

Délimitation La méthode par structures obtient une précision de 88,4 % sur la délimitation des passages évaluatifs. Par comparaison, le score de la méthode baseline chute à 68,6 %. Les structures évaluatives permettent donc de mieux délimiter les passages par rapport à la méthode baseline. Cette dernière se contente de délimiter l'évaluation en posant ses bornes sur celles du syntagme englobant un mot évaluatif. Cela implique des erreurs illustrées par les exemples (39), (40) et (41). Dans ces exemples, la méthode par structures évaluatives englobe davantage de syntagmes liés à l'évaluation. En particulier, les structures permettent de considérer les tournures de négation des évaluations de (39) et (40).

- (39) [...] ça | n'a pas | l'air | d'être très facile | tout ces nouveaux mots québécois |.
-
- (40) | je me | lève | à 6h15 | ce qui est loin | de me mettre | d'humeur euphorique |.
-
- (41) | La visite | du temple | [...] nous | avons envie | d'y rester | des heures | [...].
-

5.4.3 Une couverture trop faible du vocabulaire évaluatif

Les mesures de rappel obtenues par les méthodes offrent un deuxième regard sur le tableau 5.4. La faiblesse des scores obtenus pour la simple détection (de 49,2 % à 51,8 %) montre que la moitié des évaluations ne sont pas détectées. Ceci s'explique par plusieurs facteurs :

- un facteur **morphologique** : la nature des blogs et le langage sms, font que de nombreuses variations orthographiques sont possibles pour certains mots (42) (43) et que celles-ci ne sont pas prises en compte ;

(42) j'adôore

(43) Le serveur a semblé alluciné par notre demande

- un facteur **lexical** : beaucoup d'adjectifs ou de noms évaluatifs n'existent pas dans le lexique de l'évaluation initial. Les structures ne possèdent aucune information sémantique sur ceux-ci et ne les considèrent donc pas. Ces adjectifs ou ces noms sont assez spécifiques aux sujets évalués et sont moins fréquents dans la langue que les mots du lexique de l'évaluation. Par exemple, *télégénique*, *énergique*, *puritaine*, *discrédité*, *tarée*, *regalade* sont des mots évaluatifs qu'il aurait été intéressant de détecter pour ce qu'ils révèlent des cibles évaluées (44) (45). Néanmoins, nous ne pouvons pas considérer tous les adjectifs d'un texte comme étant évaluatifs (*un produit scalaire*, *une voiture rouge*), la mesure de précision chuterait fortement. Il faut donc pouvoir détecter des expressions ou des mots évaluatifs inconnus ou plus rares (voir chapitre 6) ;
- un facteur **discursif** : les internautes choisissent d'évaluer à partir de structures qui correspondent la majorité du temps à des stéréotypes sociaux, à des expressions figées ou à des figures de styles particulièrement nombreuses. Pour les mêmes raisons que précédemment, elles n'existent pas dans le lexique de l'évaluation : *sa fibre écolo*, *pitbull aux lèvres rouges*, *fardeau*, *bon débarras*, *à tomber par terre* (46), etc.

(44) J'ai eu droit à des liguini au basilic, des tortellone au chèvre [...] des lasagnes aux épinard. **Beaucoup de regalades en perspective !**

(45) Venant d'une tarée d'Alaska qui défoncent des élans au sniper ça me paraissait crédible.

(46) Les ingrédients sont tellement bon que le résultat est à tomber par terre.

La question de la couverture du lexique de l'évaluation et la richesse du vocabulaire évaluatif est la question centrale du chapitre 6. Notons enfin que sur les 2 031 structures d'évaluations acquises, seules 368 ont été activées dans le corpus test de 50 documents. Il serait intéressant d'affecter une mesure d'importance sur chacune des structures en fonction de leur fréquence d'utilisation dans un corpus de plus grande taille et de mesurer un indice de confiance pour chacune d'elle. Néanmoins, le coût et la complexité de construction d'un tel corpus demeure.

5.4.4 Une bonne catégorisation des traits sémantiques et énonciatifs

À partir des évaluations **correctement détectées** par les deux méthodes, nous nous intéressons à mesurer la catégorisation de la modalité et de la configuration d'énonciation des évaluations. Pour la modalité, les résultats de la catégorisation sont récapitulés dans le table 5.5.

Méthode	Modalité	Précision	Rappel
Méthode par structures	Opinion	85,1% (48/56)	100,0% (48/48)
	Appréciation	100,0% (400/400)	99,2% (400/403)
	Accord/Désaccord	100,0% (39/39)	88,6% (39/44)
Méthode baseline	Opinion	61,5% (32/52)	82,1% (32/39)
	Appréciation	98,2% (389/396)	98,7% (389/394)
	Accord/Désaccord	100,0% (22/22)	59,5% (22/37)

TABLE 5.5 : Résultats de la catégorisation de la modalité

Modalité d'une évaluation À partir des résultats du tableau 5.5, nous observons qu'il y a assez peu d'ambiguïté pour catégoriser la modalité d'une évaluation. Une approche lexicale de base est déjà suffisante pour catégoriser correctement une bonne partie des modalités. Les structures évaluatives permettent néanmoins d'améliorer ces scores. Elles distinguent mieux les modalités d'accord/désaccord et d'opinion. Par ailleurs, les passages qui contiennent des entrées lexicales de différentes modalités ont tendance à tromper la méthode baseline, mais sont bien traitées grâce aux structures évaluatives. Par exemple, dans (47), l'adjectif *bien* ne correspond pas au marqueur d'appréciation positif, ceci induit en erreur la méthode baseline et non la méthode par structures.

(47) Je pense bien que cette psychose est née de [...].
} opinion

Avec la méthode par structures, nous observons également que :

- sur les 60 opinions annotées manuellement dans le corpus, 48 ont été correctement détectées et catégorisées. Soit un rappel de 80 % ;
- sur les 851 appréciations annotées manuellement, 400 ont été correctement détectées et catégorisées. Soit un rappel de 47 % ;
- sur les 44 accords/désaccords annotés manuellement, 39 ont été correctement détectées et catégorisées. Soit un rappel de 86,7 %.

La tâche critique concerne donc la détection des évaluations dont la modalité est l'appréciation.

Configuration d'énonciation d'une évaluation À partir des résultats du tableau 5.6, nous observons qu'il y a également assez peu d'ambiguïté pour catégoriser la configuration d'énonciation d'une évaluation. Les évaluations explicites sont en particulier reconnaissables grâce à l'utilisation d'embrayeurs de personnes (*je, nous, mon, etc.*) et de verbes d'appréciations explicites (*aimer, détester, apprécier, etc.*). Lorsque ces marqueurs ne sont pas présents, il s'agit très fréquemment d'appréciations implicites.

Méthode	Configuration énonciative	Précision	Rappel
Méthode par structures	Explicite	95,2 % (177/186)	97,3 % (177/182)
	Implicite	98,4 % (304/309)	97,1 % (304/313)
Méthode baseline	Explicite	94,2 % (130/138)	73,0 % (130/178)
	Implicite	85,5 % (284/332)	97,3 % (284/292)

TABLE 5.6 : Résultats de la catégorisation de la configuration d'énonciation

D'après le tableau 5.6, la méthode de base permet, proportionnellement, de bien rappeler les évaluations implicites (rappel : 97,3 %). En revanche, les évaluations explicites ont tendance à être classées implicites (rappel : 73,0 %). La méthode par structures corrige ce déséquilibre. La principale raison est qu'en délimitant mieux les évaluations, les structures évaluatives permettent de détecter les marqueurs de personne qui apparaissent dans un contexte un peu plus éloigné. Par ailleurs, dans le paragraphe précédent, nous avons observé que la tâche critique concernait essentiellement la détection des appréciations. À travers la catégorisation de configuration d'énonciation, nous observons les faits suivants :

- sur les 614 appréciations implicites annotées manuellement dans le corpus, seulement 239 ont été correctement reconnues et catégorisées. Soit un rappel de 38,9 % ;
- sur les 237 appréciations explicites annotées manuellement, 161 ont été correctement reconnues et catégorisées. Soit un rappel de 67,9 %.

Pour les appréciations explicites, le rappel de 67,9 % s'explique en partie par l'utilisation fréquente du langage SMS ou de formes d'expressions orales (*waahhh!!, oufff!*, etc.) dont les retranscriptions écrites sont très variables. En revanche, la détection des appréciations implicites apparaît particulièrement critique. Son rappel très bas dépend principalement de la faible couverture du lexique de l'évaluation. Nous discutons de cet aspect dans la conclusion de ce chapitre.

5.4.5 Un intérêt pour résoudre les effets du contexte sur l'axiologie

Un aspect important de la méthode par structures évaluatives est sa capacité à ré-ajuster la polarité axiologique d'une évaluation par rapport à son contexte. Le tableau 5.7 montre les résultats obtenus par la méthode par structures en comparaison avec la méthode baseline. Celle-ci se scinde en deux stratégies A et B (§. 5.3.2). Ces scores sont établies uniquement à partir des évaluations de type appréciation correctement détectées puisque ce sont les seules qui possèdent une axiologie.

À partir du tableau 5.7, nous observons que la méthode baseline A, qui ne traite pas du tout les marqueurs de négation, obtient une précision de 68,4 % à 83,5 % selon la polarité. La méthode baseline B, traitant la négation de façon relativement naïve, obtient une précision allant de 87,9 % à 90,1 %. La méthode par structures (93,3 % à 96,5 %) permet donc d'améliorer la catégorisation axiologique des passages évaluatifs en analysant plus finement l'organisation des différents marqueurs de négation. Par exemple, (48) est mal catégorisé par la baseline A et (49) est mal catégorisé par la baseline B, alors que les structures évaluatives catégorisent correctement ces passages.

<i>Méthode</i>	Axiologie	Précision	Rappel
Méthode par structures	Positif	93,3 % (266/285)	90,8 % (266/293)
	Négatif	96,5 % (83/86)	77,8 % (83/107)
	Ambiguë ou inconnue	29 occurrences	-
Baseline A	Positif	83,5 % (259/310)	89,6% (259/289)
	Négatif	68,4 % (52/76)	48,6% (52/107)
	Ambiguë ou inconnue	10 occurrences	
Baseline B	Positif	90,1 % (272/302)	95,1 % (272/286)
	Négatif	87,9 % (73/84)	68,2 % (73/107)
	Ambiguë ou inconnue	10 occurrences	

TABLE 5.7 : Résultats de la catégorisation axiologique

(48) Les énergies renouvelables [...] ne se sont pas avérées efficaces .

baseline A : positif / structure et référence : négatif

(49) [...] ne s'endort jamais sans pleurer .

baseline B : positif / structure et référence : négatif

Les résultats de la méthode par structures montrent néanmoins que le rappel des appréciations négatives est plus faible que celui des appréciations positives. Autrement dit, les appréciations négatives ont tendance à être catégorisées comme positives. Nous expliquons ces erreurs à travers deux facteurs :

- un facteur **syntactique** : certaines structures inversives n'ont pas été apprises durant l'entraînement, notamment lorsque les marques de négation sont particulièrement éloignées de l'évaluation ;
- un facteur **discursif** : de nombreux cas d'évaluations ironiques occurent au sujet de Sarah Palin (*voilà un bel exemple, toujours aussi passionnant*) et des sushis (*merci de m'avoir foutu la gerbe*). Ici, la sémantique lexicale (*bel, passionnant, merci*) induit en erreur la catégorisation axiologique.

Il est d'ailleurs intéressant de noter que les structures inversives et ironiques sont plus souvent activées pour évaluer défavorablement que favorablement. Nous constatons également que :

- sur les 526 appréciations positives annotées manuellement, 266 ont été correctement reconnues et catégorisées (rappel : 50,6%),
- sur les 325 appréciations négatives annotées manuellement, 83 ont été correctement reconnues et catégorisées (rappel : 25,5%).

Les appréciations négatives sont plus difficiles à traiter dans le corpus test. En plus des deux facteurs précédemment évoqués, nous pouvons en ajouter un troisième par l'utilisation de figures de styles métaphoriques qui perturbe la détection des appréciations négatives. Les appréciations portant sur Sarah Palin, en particulier celles qui sont défavorables, utilisent très souvent des figures de styles métaphoriques (ex : *se faire lyncher, une clown, bouc émissaire, pitbull aux lèvres rouges*), ce qui a pour effet de limiter leur détection automatique.

5.4.6 Une bonne adaptation à des corpus multi-thématiques

La méthode par structures évaluatives a également été appliquée dans le cadre de la campagne d'évaluation DEFT'09. En 2009, deux tâches principales sont proposées dans DEFT :

- la catégorisation globale d'un texte en deux classes : subjectif ou objectif ;

- la détection et la délimitation des passages subjectifs d'un texte.

La catégorisation globale d'un texte en trois classes (positif, négatif, intermédiaire) avait déjà fait l'objet de DEFT'07. En 2009, l'analyse de la subjectivité à granularité fine est donc présentée par les organisateurs comme un défi nouveau et ambitieux pour tenter de mieux répondre aux problématiques du domaine de la fouille d'opinions. Cette deuxième tâche est formulée ainsi par les organisateurs :

Détection des passages subjectifs d'un texte : un texte peut être segmenté en passages objectifs, qui donnent des faits, ou le thème du texte, et en passages subjectifs qui délivrent une opinion, un sentiment, sur ces faits, concernant ce thème. Cette tâche consiste donc à repérer les passages subjectifs d'un texte, que ce texte soit globalement subjectif ou objectif. Un passage peut aller d'un mot (par exemple un modifieur) à plusieurs phrases. - <http://deft09.limsi.fr/>

Cette tâche va dans le sens de nos hypothèses et de nos travaux. Par rapport à nos propres expérimentations, il s'agit de délimiter des passages subjectifs sans les catégoriser sémantiquement. Nous réutilisons donc les structures évaluatives acquises pour identifier les passages subjectifs. Ce travail a nécessité quelques adaptations mineures que nous ne détaillons pas ici, nous renvoyons pour cela à (Vernier, Monceaux et Daille, 2009).

Corpus Le corpus francophone utilisé dans DEFT'09 pour cette tâche est constitué à partir de deux sources distinctes : le journal *Le Monde* et la retranscription écrite de débats au parlement européen. Ce corpus se compose de textes non annotés divisés en deux sous corpus :

- un corpus « *Parlement* ». Celui-ci comprend 12 913 interventions de parlementaires, tenues au Parlement européen entre 1999 et 2004 ;
- un corpus « *Journal* ». Celui-ci comprend 16 787 articles du Monde sur la période 2003/2006 (éditorial, brève, article généraliste).

Protocole d'évaluation Contrairement au corpus Blogoscopie, les corpus utilisés pour DEFT'09 ne disposent pas d'annotation de référence des passages subjectifs. Les organisateurs évoquent le coût important pour créer manuellement ce type de corpus de référence. Pour évaluer l'efficacité des méthodes, ils choisissent alors de retenir le principe de constitution des données de référence par vote majoritaire. Dit autrement, les données de référence sont constituées a posteriori, à partir des résultats sur lesquelles une majorité de participants sont tombés d'accord. Ce principe a déjà été expérimenté dans des campagnes d'évaluation d'analyseurs syntaxiques (Paroubek, Robba, Vilnat et Ayache, 2008). L'évaluation des logiciels par rapport à cette référence donne une bonne idée des écarts entre les logiciels, sans pour autant donner un classement fiable entre participants d'après Grouin et collab. (2009).

Résultats Deux équipes ont participé à cette tâche et ont soumis chacune trois fichiers de résultats. Le faible nombre de participants limite d'emblée la portée du protocole d'évaluation fixé précédemment. Chaque soumission comporte à la fois le corpus des débats parlementaires et le corpus des articles de journaux. Le tableau 5.8 rassemble les résultats de ces soumissions. Les données de référence pour chaque corpus résultent de l'accord majoritaire entre les six soumissions. Plus une soumission est proche de ce sous-ensemble, plus le F-mesure est élevée. Les résultats pour chaque soumission marquent donc avant tout son écart par rapport à cet accord.

Notre première soumission (*LINA - S1*) est celle qui utilise la méthode par structures évaluatives pour détecter les passages subjectifs. Sans pouvoir réellement évaluer la pertinence des résultats, nous observons néanmoins que notre méthode obtient des résultats analogues aux expérimentations précédentes sur les blogs : un score de précision d'environ 92,8 % et un score de rappel assez bas

(environ 52,4 %). Nos deux autres soumissions font varier quelques paramètres. En particulier, la soumission 3 étend la soumission 1 au grain phrase. Elle obtient alors un meilleur rappel car nous avons anticipé le fait que les équipes concurrentes proposeraient des méthodes au grain phrase.

Corpus	Méthode	F-Mesure	Précision	Rappel
Presse	LINA - S1	0,670	0,928	0.524
Presse	LINA - S2	0,623	0,623	0.623
Presse	LINA - S3	0,863	0,808	0.926
Presse	LIPN - S1	0,777	0.701	0.871
Presse	LIPN - S2	0,714	0.929	0.579
Presse	LIPN - S3	0,775	0.699	0.869
Débats parlementaires	LINA - S1	0,648	0.805	0.543
Débats parlementaires	LINA - S2	0,648	0.804	0.543
Débats parlementaires	LINA - S3	0,909	0.903	0.916
Débats parlementaires	LIPN - S1	0,799	0.806	0.791
Débats parlementaires	LIPN - S2	0,678	0.816	0.580
Débats parlementaires	LIPN - S3	0,797	0.805	0.789

TABLE 5.8 : Résultats de la détection des passages subjectifs (DEFT'09)

Sans pouvoir réellement discuter davantage de ces résultats, ceux-ci laissent supposer que les structures évaluatives acquises et le lexique utilisé ne sont pas spécifiques à un corpus ou à une thématique. En revanche, la question de la couverture du vocabulaire évaluatif est une nouvelle fois posée sur les corpus DEFT'09 (rappel de 52,4 %). Les ressources lexicales pour le traitement du français nécessitent d'être améliorées.

5.5 Conclusion

5.5.1 Synthèse

Dans ce chapitre, nous proposons une méthode pour détecter, délimiter et catégoriser automatiquement les différents passages évaluatifs d'un énoncé. Cette méthode s'appuie sur des ressources lexicales et sur des structures linguistiques liées à l'expression des évaluations. Ces structures ordonnent des constituants en articulant des traits lexicaux, grammaticaux et sémantiques. Nous utilisons une méthode intermédiaire à base de règles pour acquérir automatiquement ce type de structures à partir d'un corpus annoté manuellement.

En comparaison avec une méthode basée uniquement sur le lexique, les expérimentations menées sur un corpus de test montrent l'intérêt de ces structures évaluatives sur plusieurs aspects. Elles permettent en particulier de mieux délimiter les passages évaluatifs. Ceci a pour conséquence :

- d'améliorer la catégorisation de la modalité et de la configuration d'énonciation des évaluations ;
- de prendre plus finement en compte les effets du contexte sur l'axiologique d'une évaluation.

Par ailleurs, ces structures ont été appliquées à des corpus d'autres thématiques lors de la campagne d'évaluation DEFT'09. Les résultats similaires obtenus semblent montrer que cette méthode est ré-utilisable pour d'autres type de corpus. De plus, le lexique de mots évaluatifs utilisé dans ces différentes expérimentations ne semble pas engendrer beaucoup d'ambiguïtés. Lors de nos expérimentations sur les blogs et dans le cadre de DEFT'09, nous observons que le lexique de l'évaluation provoque un bruit évalué de 7,2 % à 8,9 % selon les corpus. À notre connaissance, il

s'agit du seul lexique francophone de mots évaluatifs évalué directement en contexte pour l'analyse à granularité fine des évaluations.

Mis à part plusieurs mots qui génèrent du bruit (*bon, beau, etc.*), les entrées du lexique de l'évaluation sont relativement stables sur les corpus analysés. Néanmoins, les expérimentations menées rendent compte d'une limite majeure beaucoup plus importante selon nous : la couverture du lexique de l'évaluation est critique pour détecter les appréciations implicites. En particulier, nous observons que les mots évaluatifs moins fréquents (*télogénique, puritaine, etc.*), les expressions (*à tomber par terre, etc.*) et les stéréotypes culturels (*pouvoir trouver un emploi, énergie propre, etc.*) sont « oubliées » lors des analyses.

Par ailleurs, la faiblesse de la couverture du lexique semble renforcer un déséquilibre entre la proportion d'appréciations positives détectées et la proportion d'appréciations négatives détectées. En d'autres termes, le lexique négatif est beaucoup moins couvert. Ceci peut sans doute s'expliquer par le fait qu'un énonciateur active un vocabulaire plus rare ou des structures plus complexes (voire ironique) pour évaluer négativement. Nous abordons ce point dans les perspectives.

5.5.2 Perspectives

Amélioration de la couverture des lexiques de mots évaluatifs L'amélioration de la couverture du vocabulaire évaluatif est la perspective principale de ce travail. Celle-ci est mise en avant par les résultats obtenus dans ce chapitre : seulement environ 50 % des évaluations sont détectées. Dans le cadre d'une analyse des évaluations à granularité fine, nous pensons qu'il est nécessaire de réexplorer les méthodes existantes dans le domaine pour extraire et catégoriser des expressions ou des mots évaluatifs d'une langue (voir chapitre 2, §. 2.3). En effet, la plupart des mots évaluatifs rares, des expressions évaluatives ou de stéréotypes culturels ne se trouvent pas dans les ressources lexicales existantes francophones ou anglophones. Il faut donc pouvoir considérer ce type de mots ou d'expressions. Pour autant, les analyses à granularité fine étant particulièrement sensibles aux erreurs, il faut pouvoir enrichir les ressources lexicales sans y ajouter de bruits. Nous développons cette problématique dans le chapitre 6 de cette thèse.

Désambiguïsation des mots évaluatifs dans leur contexte Parmi les autres perspectives de ce travail, nous avons évoqué la désambiguïsation en contexte de certains mots évaluatifs. Ceux-ci peuvent être instables selon la thématique ou le concept qu'ils ciblent. Par exemple, dans (50), l'adjectif *fondamental* active une évaluation, alors que dans (51), l'emploi du mot *fondamentale* n'est pas évaluatif.

(50) C'est **fondamental** de finir dans les deux premiers de notre groupe.

(51) Huawei ouvre un centre de **recherche fondamentale** en France

Néanmoins, selon nous, cette problématique ne peut pas être abordée avant de pouvoir identifier automatiquement sur quoi porte un mot évaluatif ou une évaluation. Or, il s'agit là d'une autre problématique complexe non résolue à l'heure actuelle. Nous l'abordons et proposons des méthodes pour traiter cette question dans le chapitre 7.

Amélioration de la modélisation des structures évaluatives Par ailleurs, les choix empiriques effectués pour modéliser les structures évaluatives et leurs constituants nécessiteraient d'être évalués plus précisément et d'être confrontés à d'autres types de représentations. En particulier, il nous semblerait intéressant de mesurer l'apport des analyses syntaxiques en dépendances pour modéliser les structures évaluatives et ses constituants. En effet, nous avons remarqué que certaines structures évaluatives pouvaient être généralisées en une seule en considérant leur construction syntaxique.

Nous nous sommes essentiellement replacés dans la continuité des travaux de Thompson et Hunston (2000) qui ne prennent pas en compte la dimension linguistique des relations syntaxiques. Leurs travaux constituent à notre connaissance le point de référence le plus proche du notre. De plus, à l'heure actuelle, l'intégration d'outils d'analyses syntaxiques pour le français à des chaînes de traitements linguistiques s'avère assez coûteuse en temps de calcul et complexe, notamment pour traiter des phrases longues. Prendre en compte la dimension syntaxique de la langue reste néanmoins une perspective majeure de ce travail.

Création de corpus annoté par amorce automatique Une dernière perspective importante issue du travail présenté dans ce chapitre est la possibilité de créer des corpus annotés à partir d'une amorce automatique. En effet, le développement manuel de corpus annotés est reconnu comme un aspect particulièrement critique dans le domaine. Dès lors, bien qu'imparfaite, la méthode et les ressources lexicales que nous proposons détecte, délimite et catégorise les évaluations avec une précision intéressante. Elles permettraient alors d'annoter un gros volume de données par amorce et de s'intéresser ensuite à des traitements automatiques pour corriger les cas qui posent des problèmes. Cette idée rejoint notamment celle des travaux de Riloff, Wiebe et Wilson (2003) qui utilisent un corpus de phrases subjectives et objectives développé par une amorce automatique.

Dans le chapitre suivant, nous présentons la problématique d'amélioration automatique de la couverture des lexiques de mots évaluatifs. Comme nous l'avons observé ici, les lexiques existants sont limités en ce qui concerne les mots rares, les expressions et certains stéréotypes culturels. Dans l'optique d'améliorer la couverture du lexique de l'évaluation, nous proposons et expérimentons deux méthodes d'enrichissement automatique basées sur des hypothèses sémantiques et morphologiques.

Améliorer la couverture du vocabulaire évaluatif

« Il suffit d'un lexique pour contenir tous les mots.
Mais à la pensée, il faut l'infini. » – Alexandre Pouchkine

6.1	Une couverture insuffisante des lexiques	98
6.1.1	Un silence rarement mesuré	98
6.1.2	À l'origine du silence	99
6.1.2.1	<i>Les syntagmes évaluatifs</i>	100
6.1.2.2	<i>Les stéréotypes culturels</i>	100
6.1.2.2	<i>Les mots peu fréquents</i>	101
6.1.3	Rompre le silence en limitant le bruit	101
6.2	Apprentissage de subjectivèmes par hypothèses sémantiques	103
6.2.1	Les tests sémantiques : des indicateurs pour discriminer les subjectivèmes	103
6.2.1.1	<i>Mesurer le degré de subjectivité</i>	103
6.2.1.2	<i>Mesurer la polarité des adjectifs inconnus</i>	105
6.2.2	Extraire des candidats à partir d'un corpus et du web	106
6.2.3	Construire un modèle de classification	107
6.2.4	Expérimentation et résultats	110
6.2.5	Discussion	113
6.3	Apprentissage de subjectivèmes négatifs par hypothèses morphologiques	114
6.3.1	Les affixes : des indicateurs pour discriminer les subjectivèmes	114
6.3.2	Extraire des candidats à partir d'un dictionnaire	114
6.3.3	Construire un modèle de classification	116
6.3.4	Expérimentation et résultats	118
6.4	Conclusion	121
6.4.1	Synthèse	121
6.4.2	Perspectives	121

Les travaux anglophones du domaine ont introduit le terme « opinion words » pour désigner les mots utilisés pour exprimer une évaluation. L'extraction et la catégorisation sémantique de ces mots est l'une des problématiques les plus anciennes de la fouille d'opinions (voir chapitre 2, §. 2.3). Elle se décompose traditionnellement en deux niveaux :

- niveau **subjectif** : déterminer si un mot est un marqueur de subjectivité (*subjectif/objectif*) ;
- niveau **polarité** : déterminer l'axiologie des marqueurs subjectifs (*positif/négatif/neutre*).

À l'exception des travaux de Wilson (2008), les ressources lexicales construites ont le principal défaut de ne pas avoir été évaluées en contexte sur des analyses à granularité fine. Il est alors assez difficile de savoir si elles provoquent du bruit (évaluation qualitative) ou du silence (évaluation quantitative). En revanche, pour le français, nous avons observé dans les expérimentations du chapitre 5 (§. 5.3) que le lexique utilisé ne permet de couvrir que la moitié des phénomènes d'évaluations. Les résultats obtenus par Wilson (2008) confirme cette tendance. Dès lors, ces observations suggèrent que la couverture des lexiques de mots évaluatifs constitue la principale faiblesse des ressources actuelles (§. 6.1). Nous justifions cette idée en montrant que les ressources existantes, francophones ou anglophones, ne prennent pas en compte certains types d'expressions ou de mots évaluatifs à l'origine du silence. Pour améliorer la couverture des ressources existantes nous proposons alors deux méthodes :

- la première s'appuie sur des hypothèses sémantiques et le web comme corpus pour extraire des expressions et des mots subjectifs (par la suite, *subjectivèmes*) (§. 6.2) ;
- la seconde s'appuie sur des hypothèses morphologiques et un dictionnaire pour extraire des subjectivèmes négatifs (§. 6.3).

Nous appliquons ces deux méthodes à l'enrichissement automatique du lexique de l'évaluation. Ces deux méthodes font l'objet d'un protocole d'évaluation distinct et d'une discussion présentés dans chacune des sections concernées. Nous discutons des améliorations possibles restantes en conclusion de ce chapitre (§. 6.4).

6.1 Une couverture insuffisante des lexiques

En fouille d'opinions, la couverture des ressources lexicales qu'elles soient francophones ou anglophones, ont rarement été évaluées lors d'analyses à granularité fine. Néanmoins, deux expérimentations, l'une anglophone, l'autre francophone, montrent que ces lexiques sont perfectibles (§. 6.1.1). En particulier, le principal levier d'action consiste à augmenter la taille des lexiques en prenant en compte certaines catégories de subjectivèmes totalement absentes ou peu représentées dans les ressources actuelles (§. 6.1.2). Les méthodes actuelles pour extraire et catégoriser automatiquement les subjectivèmes d'une langue ne sont pas adaptées pour traiter ces types de subjectivèmes ou pour conserver une précision importante (§. 6.1.3). Nous motivons alors le développement de deux nouvelles méthodes s'appuyant sur des hypothèses linguistiques.

6.1.1 Un silence rarement mesuré

Il existe trois catégories d'approches pour extraire et catégoriser les mots évaluatifs d'une langue : manuelle, par dictionnaires ou par corpus (voir chapitre 2, §. 2.3). Ces différentes approches ont été utilisées pour créer plusieurs lexiques :

- pour le français, il existe principalement deux lexiques : le *lexique de l'évaluation* (voir chapitre 4, §. 4.2) et le *lexique des sentiments* (Mathieu, 2006) ;
- pour l'anglais, les lexiques les plus fréquemment cités sont : le *Subjectivity Lexicon* (Wilson et collab., 2005) et *SentiWordNet* (Baccianella et collab., 2010).

Face au manque de corpus de référence, l'évaluation de ces lexiques est complexe et rarement discutée. Typiquement, ils sont évalués soit :

- **hors-contexte**, en les comparant à des lexiques développés manuellement (eux-mêmes difficilement évaluables) ;
- **en contexte** mais de façon indirecte, en mesurant le gain obtenu par des méthodes de catégorisation d'opinion au grain document ou au grain phrase. Or, pour ces granularités, les erreurs peuvent être masquées. Dès lors, le bruit et le silence causés par les lexiques ne sont pas réellement mesurés.

À notre connaissance, seul le Subjectivity Lexicon a fait l'objet d'une évaluation en contexte sur une analyse à granularité fine des évaluations (Wilson, 2008). Dans ses travaux de thèse, Wilson (2008) a pour objectif de détecter les passages axiologisés d'un corpus anglophone. Elle utilise pour cela environ 8 000 mots évaluatifs axiologisés issus du Subjectivity Lexicon. Sa méthode s'appuie sur ce lexique et tient également compte de quelques autres caractéristiques pour désambiguïser les mots du lexique dans leurs contextes (voir chapitre 2, §. 2.2.1.1). Sa méthode est évaluée sur 494 documents du corpus MPQA annotés manuellement (voir chapitre 4, §. 4.1) et obtient :

- une précision de 71,7 % ;
- un rappel de 59,5 %.

En d'autres termes, ces résultats montrent qu'environ 28 % des passages détectés correspondent à du bruit et 40,5 % des passages axiologisés du corpus ne sont pas couverts par le Subjectivity Lexicon. Les résultats obtenus sur l'anglais par Wilson (2008) offrent un parallèle intéressant avec nos propres résultats pour le français. Lors des expérimentations décrites dans le chapitre 5 (§. 5.4), selon les corpus nous obtenons :

- une précision allant de 80,5 % à 92,8 % ;
- un rappel allant de 51,8 % à 54,3 % .

Le gain en précision ne nous semble pas surprenant : contrairement au Subjectivity Lexicon, le lexique de l'évaluation a été développé manuellement, il contient moins d'entrées que le Subjectivity Lexicon (environ 1 000 contre 8 000) mais celles-ci correspondent à des subjectivèmes plus stables. En revanche, à l'image des résultats obtenus avec le Subjectivity Lexicon, le score de rappel de nos expérimentations est critique. Nous pensons néanmoins que la couverture de ces lexiques peut être améliorée par des approches automatiques (§. 6.1.3).

À partir des expérimentations menées dans le chapitre 5 et d'observations fondées sur les lexiques francophones et anglophones, nous présentons les caractéristiques des subjectivèmes à l'origine du silence.

6.1.2 À l'origine du silence

Pour comprendre l'origine des erreurs de notre méthode de délimitation et de catégorisation des passages évaluatifs, nous avons observé les cas d'échecs dans le corpus test de Blogoscopie (voir chapitre 5, §. 5.4). Nous relevons principalement trois caractéristiques à l'origine du silence :

- la faible représentation d'expressions ou de syntagmes évaluatifs dans le lexique de l'évaluation (*donner de la nourriture aux cochons, une vie de chien, un combat de coq, etc.*) ;
- la non prise en compte de stéréotypes culturels évaluatifs relativement stables (*marquer un but, avoir des diplômes, se lever tôt, voter pour un parti extrémiste, etc.*) ;
- l'absence de certains mots peu fréquents, de formes fléchies ou de néologismes (*profanation, liberticide, facebookard, etc.*).

Précisons que ces trois caractéristiques ne sont pas nécessairement indépendantes les unes des autres. Des observations manuelles sur les ressources francophones et anglophones existantes montrent que ces caractéristiques ne sont habituellement pas prises en compte.

6.1.2.1 Les syntagmes évaluatifs

À de rares exceptions, les entrées des lexiques sont des unigrammes de mots. Les syntagmes évaluatifs y sont donc très peu représentés. Un syntagme est constitué d'un mot *noyau* et d'un ou plusieurs mots *satellite*. Ce noyau ou ses satellites peuvent être eux-mêmes des subjectivèmes, à l'image de :

- *une lueur d'espoir, un motif de fierté, une campagne de diffamation*, etc. ;
- *une erreur technique, un risque sanitaire, une pollution visuelle, gagner ses galons*, etc.

Néanmoins, dans un très grand nombre de cas, l'ensemble d'un syntagme est évaluatif sans qu'aucun des éléments du syntagme ne le soit (ni le noyau, ni les satellites) :

- *un coup de bol, une tempête dans un verre d'eau, une source de motivation, rester scotcher, un panier de crabe, un bout en train, un règlement de compte, transporter le public, une levée de bouclier, une bulle d'oxygène*, etc.

Nous notons l'importance des usages métaphoriques¹ dans ses exemples d'expressions évaluatives. Ces subjectivèmes sont particulièrement sous représentés dans les ressources par rapport à leur fréquence dans le discours évaluatif. Sur une base de 30 exemples pour le français (et leurs traductions anglais) rencontrés dans le corpus test Blogoscopie, nous n'avons trouvé aucun de ces syntagmes dans les quatre ressources francophones ou anglophones citées. Ces syntagmes sont également employés pour exprimer des stéréotypes culturels et constituent la deuxième caractéristique à l'origine du silence des lexiques.

6.1.2.2 Les stéréotypes culturels

D'après sa définition sociologique, les stéréotypes culturels sont les idées partagées par une communauté, ils sont les produits des émotions d'un milieu social. En ce sens, à la différence des syntagmes évaluatifs qui sont sémantiquement relativement stables, les stéréotypes culturels sont subjectifs en fonction de l'idéologie de l'énonciateur et donc relativement instables. Par exemple, l'expression *innovation technologique* est un stéréotype positif dans le milieu scientifique sans que les deux mots qui le compose aient un sens intrinsèquement positif. De la même façon, *marquer beaucoup de buts* est un stéréotype positif dans le milieu sportif.

Les lexiques actuels ne rendent pas compte de l'existence de cette catégorie de subjectivème. Néanmoins, de part les apprentissages statistiques sur lesquels sont basés le SentiWordNet et le Subjectivity Lexicon, on trouve quelques traces de stéréotypes culturels. Ainsi, selon le SentiWordNet : *to proof* (prouver) est positif, *little* (petit) est négatif, *cognizance* (connaissance) est positif. Pour le Subjectivity lexicon : *baby* (bébé) est subjectif, *inundated* (inondé) est négatif, etc.

Ces deux ressources anglophones sont limitées aux mots simples. La présence de ces mots dans les ressources anglophones pose problème car elle mélange deux types de subjectivèmes qui sont soit stables, soit instables. Ainsi, pour SentiWordNet, *petit* est négatif, ce qui entraîne un bruit important pour des analyses à granularité fine. Par exemple, *un petit coin de paradis* ou *une petite brune* peuvent renvoyer à des stéréotypes positifs. Nous donnons ci-dessous quelques exemples de stéréotypes culturels subjectifs axiologiques dans certains discours. Ils sont absents des ressources existantes et ne sont pas détectés lors d'analyse à granularité fine :

- *un gain de temps, une chance de médaille, un ado rebelle, un effondrement de l'économie, une avancée en matière de, un guépier juridique, un complot médiatique, la politique de l'autruche, réussir sa vie, griller un feu, fuir les responsabilités, polluer, développement durable, innovation technologique, collaborer*, etc.

1. Nous discutons de la fréquente corrélation entre métaphores et le discours évaluatifs dans (Vernier et Ferrari, 2007).

Selon le contexte idéologique (lié à énonciateur, à l'actualité, à la culture ambiante, etc.), un même subjectivème peut se charger d'une axiologie différente, voire opposée. Par exemple, *collaborer* est un subjectivème négatif dans le contexte de la seconde guerre mondiale alors qu'il est plutôt chargé d'une axiologie positive dans le contexte actuelle. L'aspect instable de ces stéréotypes est une caractéristique importante qu'il faut prévoir dans les méthodes et les ressources.

6.1.2.3 Les mots peu fréquents

La troisième caractéristique est la faible représentation des mots peu fréquents dans les lexiques (en particulier, les lexiques francophones). Par mot « peu fréquent », nous faisons référence aux mots dont le nombre d'occurrences dans des corpus de taille conséquente est faible. Par exemple, en prenant le web comme corpus, *pingre* apparaît 150 000 fois selon Yahoo!Search alors qu'*avare* apparaît à 2 500 000 reprises. *pingre* est considéré comme un mot peu fréquent. Il est absent du lexique de l'évaluation et du lexique des sentiments alors qu'*avare* y est présent par exemple. De part leur mode de création manuelle fondé sur l'observation en corpus, un grand nombre d'unigrammes de mots subjectifs relativement peu fréquents ne sont pas présents :

- *plébiscite, persécution, profanation, défouloir, connivence, piquette, baume, liberticide, matraquage, bouc émissaire, cacophonie, etc.*

Parmi les mots peu fréquents, les formes fléchies et les néologismes sont également très peu représentés dans les ressources actuelles. Ainsi, des formes fléchies comme *journaloux, crasseux, scribouillard, racontard, individualisme* ou des néologismes *facebookard, hoax, plan social, lepeniste, islamophobie, ultralibéralisme, leadership, etc.* permettent souvent d'exprimer un contenu subjectif sous couvert de noyaux « objectifs », de mots nouveaux ou empruntés à d'autres langues.

Dans la section suivante, nous présentons notre objectif consistant à améliorer la couverture des lexiques de subjectivèmes sans augmenter considérablement le bruit. Nous montrons en quoi les méthodes existantes sont inadaptées pour cet objectif.

6.1.3 Rompre le silence en limitant le bruit

Compte tenu des observations précédentes, notre objectif consiste à améliorer l'extraction automatique des subjectivèmes à l'origine du silence des lexiques. Les approches du domaine, par dictionnaire ou par corpus, sont inadaptées pour deux raisons :

- soit parce qu'elles ne permettent de traiter que les unigrammes de mots ;
- soit parce qu'elles provoquent trop de bruit pour l'analyse à granularité fine des évaluations.

Approche par dictionnaire Les méthodes par dictionnaire ou par ressource ontologique utilisent quelques mots grains dont la subjectivité ou la polarité est connue pour amorcer l'extraction et la catégorisation de mots évaluatifs. Il peut s'agir :

- **d'exploiter les relations sémantiques de synonymie et d'antonymie** présentes dans la ressource (typiquement, WordNet). Si beaucoup de synonymes ou d'antonymes d'un mot candidat sont subjectifs (respectivement positifs, respectivement négatifs), alors le mot candidat est subjectif (respectivement positif, respectivement négatif) (Kamps et collab., 2004 ; Andreevskaia et Bergler, 2006). Principale faiblesse de cette approche, elle suppose que le dictionnaire soit exhaustif. Or, les syntagmes, les stéréotypes culturels et certains mots peu fréquents n'y sont pas présents. De plus, jusqu'à présent cette méthode n'a été utilisée que pour extraire les adjectifs évaluatifs de l'anglais et n'a pas été évaluée sur des analyses à granularité fine ;

- d'exploiter la **définition des mots** candidats dans un dictionnaire. Si la définition d'un mot contient une proportion importante de mots subjectifs, alors ce mot est subjectif (Takamura, Inui et Okumura, 2007 ; Banea et collab., 2008). De la même façon, les dictionnaires utilisés ne contiennent que des mots simples et ne permettent pas d'extraire les expressions et les néologismes notamment.

Approche par corpus Les méthodes par corpus dérivent d'une unique hypothèse linguistique, introduite par les travaux de Turney (2002) (voir chapitre 1, §. 2.1.1.1) et dérivée par la suite pour s'appliquer soit à la subjectivité, soit à l'axiologie :

- Hypothèse 1 : (1-1)** Plus un mot w est utilisé dans un contexte subjectif et plus il est probable que w soit subjectif.
(1-2) Plus un mot w est utilisé dans un contexte positif (respectivement négatif) et plus il est probable que w soit positif (respectivement négatif).

Dans cette hypothèse, la notion de *contexte* est variable. Selon les travaux du domaine, il s'agit soit de considérer :

- l'**étiquette globale** associée à un document. La subjectivité et la polarité du contexte sont déterminées par l'étiquetage de chaque document du corpus (Blitzer, Dredze et Pereira, 2007 ; Torres-Moreno et collab., 2007 ; Plantié et collab., 2007). Les mots qui sont statistiquement plus saillants dans une catégorie (subjectif, objectif positif, négatif) sont alors considérés comme ayant ce trait. Principale limite, les corpus pré-étiquetés sont limités à certains types de corpus monothématiques. De plus, certains mots extraits sont spécifiques à un corpus et provoquent du bruit lors d'analyses à granularité fine sur d'autres corpus ou d'autres thématiques ;
- un calcul de similarité avec des **mots grains** présents dans une fenêtre autour d'un candidat. Dans la même idée que le point précédent, cette méthode exploite statistiquement les mots candidats qui ocurrent fréquemment dans la même fenêtre que des grains positifs (respectivement négatifs) pour considérer qu'ils sont positifs (respectivement négatifs) (Turney, 2002 ; Bestgen, 2002). Cette méthode est efficace pour catégoriser globalement des documents mais elle extrait également beaucoup de candidats bruités. Elle a jusqu'à présent été utilisée pour extraire des unigrammes ou des n-grammes de mots, mais elle n'a pas été adaptée pour extraire des syntagmes.

Dans ces deux cas, l'approche par corpus est intéressante dans le sens où elle est plus exhaustive que l'approche par dictionnaire. Néanmoins, selon nous, elle s'appuie sur une hypothèse linguistique trop faible et produit trop de bruit pour des analyses à granularité fine sur des corpus multithématiques.

Les méthodes du domaine ont été principalement développées dans l'objectif de catégoriser la subjectivité ou la polarité axiologique à des grains larges (document ou phrase). De ce fait, les listes de mots évaluatifs extraites par ces méthodes n'ont pas été évaluées sur leur précision et leur couverture du langage évaluatif, mais plutôt sur leur capacité moyenne à distinguer le subjectif de l'objectif et le positif du négatif. L'objectif de notre travail consiste alors à :

- extraire les subjectivèmes absents des lexiques actuels (syntagmes évaluatifs, stéréotypes culturels, mots peu fréquents) ;
- conserver une précision élevée pour une analyse à granularité fine.

Nous proposons pour cela deux méthodes d'apprentissages automatiques fondées :

- sur des hypothèses sémantiques et le web comme corpus ;
- sur des hypothèses morphologiques et un dictionnaire.

Nous décrivons et expérimentons ces deux méthodes dans les deux sections suivantes (§. 6.2 et §. 6.3).

6.2 Apprentissage de subjectivèmes par hypothèses sémantiques

Nous proposons une première méthode pour améliorer la couverture des lexiques de mots évaluatifs (ou *subjectivèmes*). Dans notre cadre, le lexique de l'évaluation est considéré comme le lexique de référence dont il faut améliorer la couverture (par la suite, le *lexique initial*) (voir chapitre 4, §. 4.2). Pour cela, nous attachons une attention particulière aux mots peu fréquents, aux syntagmes évaluatifs et aux stéréotypes culturels. Nous regroupons ces différents types de candidats par l'emploi du mot *terme*. La méthode a un double objectif, il s'agit d'apprendre :

- de nouveaux subjectivèmes (catégorisation « *subjectif/objectif* ») ;
- caractériser l'axiologie de ces subjectivèmes (« *positif/négatif/ambigu* »). Pour ce deuxième objectif, nous nous limitons aux adjectifs.

Cette méthode s'appuie sur le principe de tests sémantiques dont les résultats sont utilisés comme indicateurs pour mesurer le degré de subjectivité et la polarité d'un terme candidat (§. 6.2.1). La première phase de notre méthode consiste à extraire des candidats à grande échelle et à réaliser ces tests sémantiques à partir du web (§. 6.2.2). Pour filtrer et conserver les candidats les plus stables, nous utilisons une phase d'apprentissage supervisé et un processus de classification (§. 6.2.3). Les candidats classés comme étant des subjectivèmes suffisamment stables sont ajoutés au lexique de l'évaluation (par la suite, le *lexique enrichi*). Nous testons et évaluons l'enrichissement réalisé en mesurant hors-contexte et en contexte l'amélioration de la couverture du lexique de l'évaluation (§. 6.2.4).

6.2.1 Les tests sémantiques : des indicateurs pour discriminer les subjectivèmes

Les deux types de tests sémantiques présentés consistent à observer la fréquence d'un terme candidat dans des contextes qui contraignent son sens. Ces observations fournissent des indicateurs permettant d'induire :

- le degré de subjectivité d'un candidat (§. 6.2.1.1) ;
- la polarité axiologique d'un candidat (les adjectifs uniquement) (§. 6.2.1.2).

6.2.1.1 Mesurer le degré de subjectivité

L'incidence de certains adjectifs (*vrai, véritable*) ou adverbes (*littéralement*) sur l'énonciation a été étudiée en linguistique (Legallois, 2005 ; Suhamy, 2006). On considère ainsi que l'adverbe *littéralement* ne doit pas être interprété littéralement, et qu'il a au contraire, de par l'usage courant, une fonction intensive qui révèle les représentations mentales et donc la subjectivité du locuteur. De là viennent des expressions comme *le contribuable est littéralement écrasé d'impôts, le pétrole qui flambe littéralement sur les marchés*, etc. Nous développons cette idée pour formuler l'hypothèse 2.

Hypothèse 2 : Plus un mot ou une expression w est utilisé à la suite d'un marqueur d'intensité et plus il est probable que w soit un subjectivème. Inversement, si un mot ou une expression n'est jamais utilisé à la suite d'un marqueur d'intensité, il est probable qu'il ne soit pas un subjectivème.

Pour illustrer cette hypothèse, si les énoncés (1),(2), (3) et (4) ont du sens, en revanche, intuitivement, les exemples (5), (6), (7) et (8) semblent mal construits.

- | | |
|---|---|
| (1) Il est <i>particulièrement</i> dynamique | (5) C'est <i>terriblement</i> scalaire |
| (2) C'est <i>véritablement</i> une hérésie | (6) C'est <i>particulièrement</i> législatif |
| (3) Il est <i>littéralement</i> tombé sous le charme | (7) C'est <i>littéralement</i> un oiseau |
| (4) Il a <i>littéralement</i> soulevé la foule | (8) Il est <i>franchement</i> allé à l'école |

En s'appuyant sur les usages linguistiques rencontrés sur internet et leur fréquence, ces intuitions sont potentiellement vérifiables. Par exemple, à partir du moteur de recherche Yahoo!Search, les requêtes exactes (9), (10), (11) et (12) apportent un indice sur le degré de subjectivité des candidats en gras. Les candidats *hérésie* et *soulever la foule* semblent potentiellement subjectifs en considérant leur fréquence, alors que *scalaire* et *manger au restaurant* semblent objectifs.

- (9) *véritablement* **scalaire** → 0 occurrence
(**scalaire** → 650 000 occurrences)
- (10) *véritablement* **mangé au restaurant** → 0 occurrence
(**mangé au restaurant** → 25 100 occurrences)
- (11) *véritablement* **une hérésie** → 13 occurrences
(**une hérésie** → 460 000 occurrences)
- (12) *véritablement* **soulevé la foule** → 15 occurrences
(**soulevé la foule** → 9 990 occurrences)

Les valeurs numériques fournies par Yahoo! soulèvent évidemment des questions de fiabilité. À partir de certaines valeurs, Yahoo! ne retourne qu'un ordre de grandeur. Toutefois, ces valeurs, même approximatives, constituent a priori des indices intéressants pour caractériser la subjectivité de termes candidats. À partir de cette hypothèse, nous définissons un ensemble S_1 de tests sémantiques construits selon l'algorithme 2, où :

- α est un candidat donné dont le degré de subjectivité est inconnu ;
- Δ_{int} est une liste de marqueurs d'intensité.

Algorithme 2 Construire un ensemble de tests sémantiques S_1 pour mesurer la subjectivité de α

$\Delta_{int} \leftarrow$ [particulièrement, terriblement, parfaitement, véritablement, littéralement, réellement, franchement, véritable, parfait, terrible, réel]

$S_1 \leftarrow \emptyset$

Pour tout δ_{int} élément de Δ_{int} **faire**

$s \leftarrow \delta_{int} \alpha$ {ex : *Véritable coup de trafalgar*}

$S_1 \leftarrow S_1 + s$

fin pour

return S_1

Dans la liste Δ_{int} , nous conservons les adverbes étudiés par Legallois (2005) et Suhamy (2006). Nous y ajoutons, de façon empirique, les 11 adverbes ou adjectifs d'intensité les plus fréquents et les moins ambigus (exclusion de : *trop, vrai*). Les adverbes d'intensité étant beaucoup plus stables sémantiquement et d'un nombre assez réduit par rapport aux termes subjectifs, il ne nous paraît pas nécessaire de chercher à faire varier cette liste. En revanche, l'apparition au fil du temps de nouveaux termes autour de marqueurs d'intensité est intéressante à détecter. Nous précisons ultérieurement l'utilisation de ces tests dans une démarche d'apprentissage (§. 6.2.2).

6.2.1.2 Mesurer la polarité des adjectifs inconnus

De façon voisine, nous empruntons l'idée de Hatzivassiloglou et McKeown (1997) pour mesurer la subjectivité et la polarité axiologique des adjectifs selon l'hypothèse 3.

Hypothèse 3 : Plus un adjectif w inconnu est mis en opposition avec un adjectif subjectif positif (respectivement négatif), plus il est probable que w soit subjectif et qu'il soit négatif (respectivement positif).

Pour illustrer cette hypothèse, si les énoncés (13) et (14) ont du sens, en revanche les exemples (15) et (16) semblent sémantiquement mal construits.

(13) C'est **joli** mais **inutile**

(15) Elle est **jolie** mais **scalaire**

(14) C'est **atroce** mais **efficace**

(16) C'est **atroce** mais **douloureux**

De la même façon, les usages linguistiques rencontrés sur internet semblent confirmer cette intuition. Par exemple, l'adjectif négatif *introverti* est connu du lexique de l'évaluation. Le résultat de la requête (17) envoyée à Yahoo! apporte un indice sur la subjectivité et la polarité de l'adjectif *conscientieux* :

(17) **introverti** mais **conscientieux** → 26 occurrences

Selon ce principe, nous définissons un deuxième ensemble S_2 de tests sémantiques à partir de l'algorithme 2 où :

- α est un adjectif dont la subjectivité et la polarité axiologique sont inconnues ;
- Δ_{axiol} est un ensemble d'adjectifs subjectifs dont la polarité axiologique est connue. Il est composé de 100 adjectifs (50 positifs, 50 négatifs) issus du lexique de l'évaluation initial.

Algorithme 3 Construire un ensemble tests sémantiques axiologiques S_2 pour l'adjectif α

$\Delta_{axiol} \leftarrow$ [liste de 100 adjectifs connus]

$S_2 \leftarrow \emptyset$

Pour tout δ_{axiol} élément de Δ_{axiol} **faire**

$s_a \leftarrow \delta_{axiol}$ mais α {ex : *conscientieux* mais *introverti*}

$s_b \leftarrow \alpha$ mais δ_{axiol} {ex : *introverti* mais *conscientieux*}

$s_c \leftarrow$ pas δ_{axiol} mais α {ex : *pas introverti* mais *conscientieux*}

$s_d \leftarrow$ pas α mais δ_{axiol} {ex : *pas conscientieux* mais *introverti*}

$S_2 \leftarrow S_2 + s_a + s_b + s_c + s_d$

fin pour

return S_2

Les tests s_c et s_d ont été ajoutés pour tenir compte de certains cas où un adjectif est modifié par une négation. D'autres constructions devraient idéalement être considérées : lorsque l'adjectif est modifié par un adverbe (x mais *peu* conscientieux, x mais *trop* conscientieux, etc.) et lorsque le connecteur relie un adjectif neutre et un adjectif axiologique (*un court* mais *beau* poème). Nous avons ici fait le choix de ne pas multiplier les différents tests sémantiques pour des raisons de coût de calcul ultérieur au moment de la réalisation automatique des tests. Nous supposons, a priori, que ces dernières constructions ont un nombre d'occurrences moins grand et perturberont l'apprentissage et la catégorisation automatique de façon marginale.

Une fois que les ensembles de tests sémantiques S_1 et S_2 ont été définis, la première phrase de notre méthode consiste à extraire des termes candidats, puis à réaliser ces tests sémantiques à partir du web et d'un moteur de recherche (§. 6.2.2). Les résultats de ces tests sémantiques servent de descripteurs pour une phrase d'apprentissage et de catégorisation que nous expliquons ultérieurement (§. 6.2.3).

6.2.2 Extraire des candidats à partir d'un corpus et du web

Extraction de termes candidats Deux sources sont utilisées pour extraire des termes candidats : un **corpus de blogs** et le **web**. Pour extraire des adjectifs candidats, nous utilisons un corpus de blogs constitué de 6 000 billets et commentaires. Ce corpus est extrait de la plateforme de blogs OverBlog et représente les différentes thématiques internes de la plateforme. Il est constitué d'environ 3 millions de mots. Nous utilisons le TreeTagger (Schmid, 1994) pour connaître la catégorie grammaticale des mots. De ce corpus de blogs, nous collectons 13 094 adjectifs candidats dont le lemme est différent.

Pour les autres catégories de termes candidats (syntagme nominal, syntagme verbal, nom et verbe), afin d'éviter d'avoir trop de termes non pertinents, nous extrayons à partir du web uniquement ceux qui apparaissent au moins une fois après un marqueur d'intensité (voir la liste Δ_{int} , §. 6.2.1.1). Pour cela nous interrogeons automatiquement le moteur de recherche *Yahoo!Search* avec l'API *Yahoo!Search BOSS*². Cette API permet de construire et d'exécuter automatiquement des requêtes sur le moteur de recherche *Yahoo!*. En réponse à une requête, elle fournit un flux de réponses en XML. L'analyse de ce flux³ offre notamment la possibilité de récupérer le nombre de résultats d'une requête et les passages textuels concernées.

Yahoo!Search retourne environ 100 000 passages de documents contenant l'un des marqueurs d'intensité. Nous collectons ces passages textuels et nous les segmentons automatiquement en syntagmes non récursifs. Nous utilisons pour cela une adaptation de l'algorithme de chunking de Vergne et Giguët (1998) sur UIMA. Nous extrayons les mots, les syntagmes nominaux (SN) et verbaux (SV) qui suivent directement le marqueur d'intensité. Nous collectons également les bigrammes et les trigrammes de syntagmes suivants : SN-SN, SV-SN, SV-SN-SN. À partir du web, 25 432 termes candidats distincts supplémentaires sont ainsi collectés. Finalement, nous disposons de 38 526 termes candidats (Tab. 6.1). Nous souhaitons désormais obtenir, pour chaque terme candidat, les résultats aux tests sémantiques décrits précédemment (algorithmes 2 et 3).

	catégorie	nombre
	Syntagmes nominaux (<i>SN</i> ou <i>SN-SN</i>)	10 188
	syntagmes verbaux (<i>SV</i> , <i>SV-SN</i> ou <i>SV-SN-SN</i>)	5 826
	noms (<i>unigramme</i>)	6 232
	verbes (<i>unigramme</i>)	3 196
	adjectifs (<i>unigramme</i>)	13 094
	TOTAL	38 526

TABLE 6.1 : Nombre de termes candidats par catégorie grammaticale

2. <http://developer.yahoo.com/search/boss/>

3. Nous utilisons le composant *fr.univ.nantes.lina.uima.YahooSearch* développé avec UIMA. Tutoriel et sources disponibles : <http://www.uima-fr.org>.

Réalisation automatique de tests sémantiques Pour chaque terme candidat, nous effectuons 11 tests sémantiques sous la forme de 11 requêtes exactes exécutées avec l'API de Yahoo!Search (voir algorithme 2). Une dernière requête donne le nombre d'occurrences du terme seul. Par exemple, pour les termes *mine d'informations*, *perdre la tête*, *tapageur*, *transcender* et *délivrance*, nous obtenons les résultats présentés dans le tableau 6.2.

catégorie	SN	SV	ADJ	V	Nom
Test avec le candidat <i>w</i>	<i>mine d'informations</i>	<i>perdre la tête</i>	<i>tapageur</i>	<i>transcender</i>	<i>délivrance</i>
véritable <i>w</i>	56 100	0	0	31	10 000
véritablement (un(e)) <i>w</i>	29	16	0	273	8
littéralement (un(e)) <i>w</i>	0	138	1	417	0
réellement (un(e)) <i>w</i>	27	70	2	83	68
particulièrement (un(e)) <i>w</i>	0	0	19	34	1
terriblement (un(e)) <i>w</i>	0	0	3	0	0
terrible <i>w</i>	0	0	0	0	38
parfait(e) <i>w</i>	7	0	0	0	286
franchement (un(e)) <i>w</i>	3	2	12	64	1
réel(le) <i>w</i>	113	0	0	0	408
parfaitement (un(e)) <i>w</i>	0	0	0	10	2
TOTAL <i>w</i>	56 279	226	37	912	10 812
<i>w</i> seul	1 330 000	396 000	243 000	1 560 000	11 000 000

TABLE 6.2 : Résultats obtenus aux tests sémantiques

Pour chaque adjectif candidat, nous effectuons 400 requêtes supplémentaires correspondants à l'ensemble des tests sémantiques définis dans l'algorithme 3. Les résultats de ces tests sémantiques servent de descripteurs lors de la phase d'apprentissage que nous détaillons ci-dessous. Cet apprentissage a pour objectif de filtrer les termes candidats qui sont liés à la subjectivité de manière trop instable.

6.2.3 Construire un modèle de classification

Données d'apprentissage Pour les besoins de la phase d'apprentissage supervisé, cinq juges humains ont annoté des termes extraits précédemment : 500 adjectifs, 500 noms ou syntagmes nominaux et 500 verbes ou syntagmes verbaux. Pour chacun des termes extraits, les juges doivent prendre la décision suivante : *le terme est-il subjectif?* {oui, non ou ambiguïté}. Pour les adjectifs, les juges doivent également donner la polarité de l'adjectif : *positif, négatif, neutre, ambiguü*. L'accord inter-annotateur est calculé à partir de l'indice Kappa de Fleiss⁴ (Fleiss, 1971).

Pour ces tâches, l'indice moyen obtenu entre juges est de 0,70 pour la subjectivité et de 0,82 pour la polarité, ce qui correspond à des bons taux d'accord d'après l'échelle de cet indice. Le vote majoritaire des juges est conservé comme catégorie finale. La tableau 6.3 présente des exemples de termes classés par les juges. Pour ces 1 500 exemples de termes candidats, la classe est connue. Ils sont alors utilisés comme données d'apprentissages pour entraîner un modèle SVM.

4. Par contraste avec l'indice kappa de Cohen, celui-ci permet de calculer l'accord interannotateurs pour plus de deux juges humains.

adjectifs	noms et SN	verbes et SV
aborigène (O)	république (O)	tricoter (O)
téléphonique (O)	république bananière (S)	prendre la grosse tête (S)
néo-nazi (A)	vie de chien (S)	échapper des griffes (S)
populiste (S)	vie de famille (S)	voler la vedette (S)
télégénique (S)	souffle de fraîcheur (S)	glandouiller (S)

TABLE 6.3 : Termes classés par les juges humains : objectif (O), subjectif (S) ou ambigu (A)

Catégorisation de la subjectivité des termes candidats avec un modèle SVM Le résultat de chaque test sémantique s_i , pour chaque terme candidat α , est utilisé comme descripteur ($\delta_{s_i, \alpha}$). Ce descripteur est normalisé en fonction du nombre d'occurrences d' α sur le web selon Yahoo!Search (éq. 6.1). Selon l'algorithme 2, il y a 11 tests sémantiques pour chaque candidat. α est alors décrit par :

- 11 descripteurs correspondant aux résultats des tests sémantiques (éq. 6.1) ;
- 1 descripteur correspondant à la fréquence globale du candidat seul ($Occ(\alpha)$) ;
- 1 descripteur indiquant la catégorie (objectif, subjectif) du candidat. Seuls les candidats classés subjectifs ou objectifs par les juges humains possèdent ce descripteur.

$$\delta_{s_i, \alpha} = \frac{Occ(s_i, \alpha)}{Occ(\alpha)} \quad (6.1)$$

Selon l'algorithme 3, il y a 400 tests sémantiques supplémentaires pour les adjectifs. En plus des descripteurs précédents, ils possèdent alors :

- 400 descripteurs correspondant aux résultats des tests sémantiques (éq. 6.1) ;
- 1 descripteur indiquant la polarité (positif, négatif) du candidat. Seuls les candidats classés positifs ou négatifs par les juges humains possèdent ce descripteur.

La figure 6.1 donne un aperçu en deux dimensions des noms classés par les juges humains. Dans cette figure, chaque point est un nom qui peut être positif, négatif ou ambigu. L'abscisse indique le nombre total d'occurrences d'un nom α sur le web selon Yahoo!Search. L'ordonnée indique le nombre d'occurrences d' α lorsqu'il est précédé d'un marqueur d'intensité. Par exemple :

- X : *élan collectif* apparaît 42 000 fois ;
- Y : *élan collectif* précédé d'un marqueur d'intensité apparaît 650 fois.

Malgré un espace des descripteurs réduit ici à deux dimensions, à partir des termes catégorisés par les humains, il est possible de distinguer deux zones : l'une de termes objectifs et l'autre de termes majoritairement subjectifs. Selon ce graphique, le terme inconnu *élan collectif* appartient à la zone des termes subjectifs. Les graphiques obtenus pour les syntagmes verbaux et les adjectifs sont d'allures équivalentes. À partir des matrices de descripteurs, nous entraînons trois classificateurs SVM (Joachims, 1997) pour traiter de façon distincte :

- les adjectifs ;
- les noms et syntagmes nominaux ;
- et les verbes et syntagmes verbaux.

Pour chaque classificateur SVM, l'objectif est de déterminer un hyperplan optimal discriminant les termes subjectifs et obtenir ainsi une fonction de classification. Une phase de réglage par validation croisée sur les données d'entraînement permet d'établir une fonction polynomiale comme la meilleure fonction noyau pour ces classificateurs SVM. Ces modèles sont ensuite appliqués pour catégoriser les 37 026 termes candidats restants. Les résultats sont présentés dans la section §. 6.2.4.

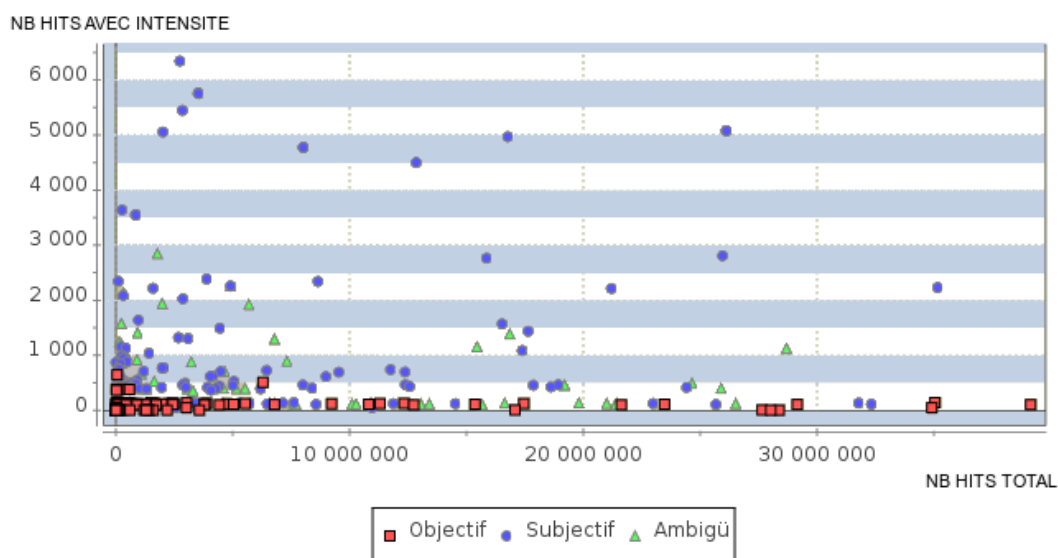


FIGURE 6.1 : Nombre d'occurrences des noms candidats en fonction de leur nombre d'occurrences précédées d'un marqueur d'intensité

Catégorisation de la polarité axiologique des adjectifs Pour chaque adjectif candidat α , le calcul de la polarité axiologique est basé sur le nombre d'occurrences d' α comme terme négatif (neg) et sur le nombre d'occurrences d' α comme terme positif (pos) :

$$\text{axiol}(\alpha) = \frac{\text{pos}}{\text{total}} - \frac{\text{neg}}{\text{total}} \quad (6.2)$$

Une visualisation graphique sous forme de nuage de mots est générée pour chaque adjectif inconnu (Fig. 6.2 et Fig. 6.3) et permet d'observer les termes connus du lexique qui s'opposent à α et ceux qui s'associent à α selon les usages linguistiques des internautes.



FIGURE 6.2 : Nuage de mots représentant les adjectifs connus du lexique qui s'opposent (polarité inverse) le plus fréquemment à l'adjectif *populiste*



FIGURE 6.3 : Nuage de mots représentant les adjectifs connus du lexique qui s'associent (polarité identique) le plus fréquemment à l'adjectif *hilarant*

La figure 6.2 montre que l'adjectif inconnu *populiste* a une axiologie qui s'oppose fréquemment à celle de *populaire* (positif). Il s'oppose également fréquemment à plusieurs autres adjectifs positifs (*honnête*, *rassurant*, *légitime*, *efficace*). Son score d'axiologie est donc négatif. Dans la figure 6.3, l'adjectif inconnu *hilarant* a été rencontré fréquemment avec une série d'adjectifs positifs dans une

construction syntaxique impliquant une négation (par exemple, *pas attachant mais hilarant*). Son score d'axiologie fait qu'il est également considéré comme positif. Dans le paragraphe suivant, nous évaluons quantitativement et qualitativement l'apprentissage effectué.

6.2.4 Expérimentation et résultats

Catégorie	Nombre	Exemples (sélectionnés aléatoirement)
Adjectifs	596	larmoyant, exorbitant, opiniâtre, lunatique, incestueux, cocace, famélique, infantile, subversif, polluant
Noms ou SN	1 390	régal, fléau, plébiscite, camouflet, marée humaine, descente aux enfers, gain de temps, cacophonie, bouffée d'air frais, capharnaüm
Verbes ou SV	488	jouer un rôle décisif, faire basculer le match, subjuguier, voler la vedette, marquer un tournant, toucher le fond, porter son équipe, trouver ses marques, inonder le marché, ovationner

TABLE 6.4 : Termes ajoutés au lexique de l'évaluation

Données en sortie La méthode exposée précédemment permet d'extraire 2 474 nouveaux termes (Tab. 6.4) et de les considérer comme des candidats suffisamment stables pour être ajoutés au lexique de l'évaluation initialement constitué de 982 entrées (lexique de l'évaluation initial). Par exemple, l'adjectif *français*, bien qu'étant très souvent associé à un stéréotype culturel subjectif (18), ce terme n'a pas été ajouté au lexique final. La fréquence très élevée de ce mot sans marqueur d'intensité est sans doute un trait qui a permis au classifieur SVM de le considérer trop instable ou ambigu. Seuls 596 adjectifs ont été à ce stade catégorisés selon leur polarité axiologique (24,1 % des nouveaux termes subjectifs). Nous cherchons à évaluer la méthode présentée sur deux aspects :

- la catégorisation de nouveaux termes *subjectifs* ;
- la catégorisation de la *polarité axiologique* des nouveaux adjectifs.

(18) la jalousie est un mal **terriblement** français !

Évaluation hors-contexte et en contexte Évaluer hors-contexte l'enrichissement du lexique nécessiterait une ressource de référence, or c'est l'absence d'une telle ressource qui fait l'objet de ce travail. Par conséquent, nous proposons une première validation de l'enrichissement effectué en considérant comme référence la liste de 1 500 mots ou expressions catégorisés par les juges-humains. Pendant la phase d'entraînement, nous avons utilisé dix validations croisées pour mesurer la précision et le rappel moyen sur ces 1 500 exemples de référence. Les résultats sont rapportés dans le tableau 6.5.

	Précision	Rappel
Objectif	75,49 % (687/910)	94,62 % (687/726)
Subjectif	77,28 % (456/590)	61,81 % (356/576)
Ambigü	0 % (0/0)	0 % (0/198)

TABLE 6.5 : Évaluation hors-contexte de l'enrichissement lexical par tests sémantiques

Les résultats montrent que même si le rappel des candidats subjectifs est plus bas que le rappel des candidats objectifs, cette méthode est capable d'extraire 456 nouveaux mots ou expressions subjectifs avec une précision de 77,28 %. Néanmoins, nous avons souligné précédemment l'importance d'une

évaluation en contexte pour déterminer si un terme est oui ou non subjectif. Ceci nous amène à proposer un protocole d'évaluation différent qui permet d'observer l'impact de l'enrichissement lexical sur un problème applicatif concret. Nous nous démarquons donc des travaux de Turney (2002) pour lesquels l'enrichissement du lexique est évalué hors-contexte à partir d'un lexique, le *General Inquirer*, développé manuellement.

Nous avons extrait 5 000 nouveaux billets et commentaires de blogs de la plateforme Over-blog sans contrainte sur leur catégorie thématique. Nous utilisons la méthode présentée dans le chapitre 5 pour délimiter et catégoriser les passages évaluatifs avec uniquement la partie enrichie du lexique (2 474 entrées). En sortie de la chaîne de traitements, deux fichiers (format CSV) sont générés pour lister : d'une part, les évaluations détectées à partir du lexique initial et d'autre part, les évaluations détectées uniquement à partir de la partie enrichie du lexique (par la suite, *partie enrichie*). La polarité axiologique du passage évaluatif et son contexte phrastique sont également renseignées selon le format CSV suivant : **passage évaluatif** ; polarité axiologique du passage ; *contexte (phrase)* (19).

- (19) **une politique énergétique ambitieuse** ; positif ; *Ségolène Royal a d'ailleurs proposé de mettre en place une politique énergétique ambitieuse [...].*

Accord inter-annotateurs Deux juges ont évalués la précision des passages évaluatifs annotés grâce à la partie enrichie du lexique. Le fichier CSV correspondant est divisé en cinq échantillons (A, B, C, D et E) répartis entre le juge humain 1 (A, B, C) et le juge humain 2 (C, D et E).

	Juge 2 : Correct	Juge 2 : Erreur	Total
Juge 1 : Correct	1 329	110	1 439
Juge 1 : Erreur	361	164	525
Total	1 690	274	1 964

TABLE 6.6 : Tableau de contingence de l'accord observé

Nous cherchons à évaluer l'accord inter-annotateur à partir de 1 964 passages évaluatifs de l'échantillon C vérifiés par les deux juges, ce qui représente un peu plus de 10 % du nombre total des passages évaluatifs. Nous constatons que les deux juges détectent 8,1 fois plus d'annotations de passages subjectifs correctes que d'annotations erronées (Tab. 6.6). De ce fait, comme l'ont montré Feinstein et Cicchetti (1990), le déséquilibre entre les deux catégories de données (prévalence) rend inadéquat le test Kappa de Cohen traditionnellement utilisé pour mesurer l'accord inter-annotateurs. D'après les données, nous pouvons uniquement conclure que les juges humains ont un taux d'accord observé de 76 % et jugent correctes, ensemble, 67 % des annotations.

Évaluation qualitative Le tableau 6.7 récapitule les évaluations des juges humains, il permet d'observer que les erreurs et les désaccords portent essentiellement sur les noms et les syntagmes nominaux (précision globale de 72,0 %). Ceux-ci sont plus enclins à déclencher des stéréotypes culturels différents chez les juges humains. Ainsi, beaucoup d'exemples contenant les noms *crise économique*, *politique écologique*, *terrorisme* ou *pandémie* entraînent des désaccords chez les juges humains (20) (21).

- (20) La **pandémie** de grippe, réelle ou inventée, permet de mettre en scène [...]

- (21) Le mot **pandémie** est d'actualité, nous l'entendons même depuis des mois.

Pour l'un des annotateurs, le choix d'utiliser le mot *pandémie* plutôt qu'*épidémie* relève d'un choix subjectif qui active un stéréotype négatif. Pour le second annotateur, ce mot n'active aucun stéréotype dans les contextes observés. Pour les autres catégories de subjectivèmes, la précision est relativement stable selon les échantillons. Les verbes ou les syntagmes verbaux ont une précision de 84,5 %. Les adjectifs ont une très bonne précision que ce soit pour déterminer leur subjectivité (93,7 %) ou leur polarité (84,5 %). Mis à part les noms et les syntagmes qui sont relativement instables, les adjectifs, les verbes et les syntagmes verbaux appris automatiquement conservent une bonne précision en contexte.

Échantillon	A	B	C	D	E	Total
<i>Verbes/SV</i>	329	600	317	270	616	2 132
erreur	90	82	29	42	87	330
correct	235	515	288	226	529	1 793
non évalué	4	3	0	2	0	9
PRÉCISION	72,3 %	86,3 %	90,8 %	83,7 %	85,8 %	84,5 %
<i>Noms/SN</i>	1 548	3 131	1 637	1 307	3 612	11 235
erreur	515	1 161	261	218	869	3 024
correct	858	1 785	1 376	1 044	2 594	7 657
non évalué	175	185	52	45	149	606
PRÉCISION	62,5 %	60,6 %	87,6 %	82,7 %	74,9 %	72,0 %
<i>Adjectifs</i>	611	1 157	610	499	1 373	4 250
erreur	18	77	28	34	73	230
correct	588	1 069	582	464	1 297	4 000
non évalué	5	11	0	1	3	20
PRÉCISION	97,0 %	93,3 %	95,4 %	93,2 %	94,7 %	94,6 %
<i>Polarité des adjectifs</i>	588	1 069	582	464	1 297	4 000
erreur	48	100	19	32	54	253
correct	540	969	563	432	1 243	3 747
PRÉCISION	91,8 %	90,7 %	96,8 %	93,1 %	95,8 %	93,7 %

TABLE 6.7 : Précision obtenue avec le partie enrichie du lexique

Évaluation quantitative Comme évoqué par les organisateurs et les participants de DEFT'09, l'absence de corpus de référence pour évaluer des tâches telles que la détection locale de la subjectivité est un problème non résolu actuellement. Il empêche la mesure de rappel puisqu'elle nécessiterait une lecture exhaustive de l'ensemble du corpus par les juges-humains. Néanmoins, nous proposons un aperçu de l'amélioration quantitative des passages annotés en comparant le nombre de passages annotés à partir de la ressource initiale, le nombre de passages annotés par la partie enrichie automatiquement et le nombre de passages annotés catégorisés comme corrects par les juges-humains (Tab. 6.8).

Nombre de passages annotés	Valeur absolue	Pourcentage
Lexique de l'évaluation initial	68 536	78,8 %
Partie enrichie uniquement	17 669	21,2 %
Partie enrichie uniquement (évaluations correctes)	13 450	15,6 %
Total	83 204	100 %

TABLE 6.8 : Évaluation quantitative de l'enrichissement du lexique de l'évaluation.

6.2.5 Discussion

D'un point de vue qualitatif, sur les 17 669 évaluations détectées grâce à l'enrichissement automatique, 13 450 ont été considérées comme correctes par les juges humains. La précision de la partie enrichie du lexique est donc de 76,12 %. Elle reste légèrement supérieure à la précision obtenue par Wilson (2008), pour l'anglais, avec le Subjectivity Lexicon (71,7 %).

D'un point de vue quantitatif, la taille du lexique de l'évaluation est passée de 982 entrées à 3 456 entrées (+ 252 %). Une première remarque concerne l'augmentation de la détection des évaluations de seulement 20 % en comparaison. Toutefois, cet enrichissement est loin de ne pas être significatif pour la raison suivante : par rapport aux termes du lexique de l'évaluation constitué manuellement (*beau, inquiétude, aimer*), les termes appris (*blasphématoire, la politique de l'autruche, faire tordre de rire*) ont une fréquence d'apparition plus faible – expliquant d'ailleurs qu'ils soient *oubliés* lors de la constitution manuelle de ressources pour la fouille d'opinions. En revanche, leur rareté rend leur usage particulièrement porteur d'informations et leur détection est de ce fait intéressante et nécessaire pour des problèmes applicatifs réels.

Dans l'exemple (22), l'enrichissement lexical permet ainsi de saisir l'opinion qui fait ici véritablement sens dans le discours (*la spoliation du peuple*) à partir de termes complexes (*spolier le peuple, payer le prix fort*), là où une ressource *classique* n'aurait cerné que le *bonheur* (avec éventuellement sa tournure négative).

- (22) La guerre actuelle du Congo a **spolié le peuple** [+lexique enrichi] congolais de ses moyens de subsistance. Pour une guerre dont la finalité est **loin de faire son bonheur** [+lexique initial], le peuple congolais **paie le prix fort** [+lexique enrichi].

Qualitativement, la précision obtenue est assez proche des résultats des méthodes proposées par Turney (2002) et Bestgen (2002). Notre approche se distingue par l'étape de pré-sélection des termes candidats par des requêtes linguistiquement motivées diminuant l'ajout de bruits dans la ressource. Cette pré-sélection nous permet d'obtenir un ensemble de mots mais également des syntagmes nominaux ou verbaux et des stéréotypes culturels qui ne sont habituellement pas traités par les approches du domaine. En se basant sur les usages linguistiques des internautes, le lexique de l'évaluation enrichi est constitué de stéréotypes linguistiques les plus socialement admis (*pantoufflard, négationniste, escroquerie*) mais aussi ceux qui sont majoritaires au moment de l'apprentissage (les termes *écologie, écologique, pollution* sont ainsi classés comme étant subjectifs de part leur intensité dans le contexte et les discours actuels). Il serait donc intéressant de reproduire cette méthode à des instants différents pour évaluer la variation de la subjectivité d'un terme au fil du temps.

En revanche, mis à part les adjectifs, nous ne mesurons pas à ce stade leur polarité axiologique. De ce point de vue, les méthodes de Turney (2002) et Bestgen (2002) sont complémentaires avec celle que nous proposons et sont une de nos perspectives pour finaliser l'enrichissement du lexique. Les adjectifs ajoutés, et à un degré moindre les groupes verbaux, permettent en contexte une détection précise des évaluations (la précision varie de 71,5 % à 97,0 %). L'ambiguïté et les sources d'erreurs proviennent essentiellement des noms et groupes nominaux en raison du caractère davantage polysémique des noms. Ainsi, les noms *farce* ou *daube* ont bien un usage subjectif fréquent (*c'est une farce cette assemblée de politiciens, front page une vrai daube*) mais occurent également dans des contextes complètement objectifs (*bien mélanger la farce, voici ma recette de daube de poisson*).

Dans la suite de ce chapitre, nous présentons une seconde méthode visant à améliorer la couverture des vocabulaires de mots évaluatifs en nous intéressant en particulier aux mots négatifs.

6.3 Apprentissage de subjectivèmes négatifs par hypothèses morphologiques

Les résultats des expérimentations menées au chapitre 5 (§ 5.4) montrent que les évaluations négatives sont moins bien détectées que les évaluations positives. L'usage de syntagmes et de mots peu fréquents pour évaluer négativement et la faible couverture du lexique de l'évaluation sont à l'origine de ce déséquilibre. Pour renforcer davantage la couverture du vocabulaire évaluatif négatif, notre deuxième méthode consiste à s'appuyer sur des hypothèses morphologiques en évaluant en particulier le rôle des affixes (§. 6.3.1). Nous extrayons des candidats à partir d'un dictionnaire (§. 6.3.2) et utilisons une phase d'apprentissage supervisé pour enrichir à nouveau le lexique de l'évaluation (§. 6.3.3). Nous expérimentons et évaluons ce second enrichissement en contexte sur une analyse à granularité fine des évaluations (§. 6.3.4).

6.3.1 Les affixes : des indicateurs pour discriminer les subjectivèmes

Le rôle des préfixes et des suffixes du français a été étudié en linguistique par Pupier (1998). Dans son travail, il propose en particulier de catégoriser les affixes du français selon l'effet qu'ils produisent sur l'axiologie d'un mot racine. Par exemple, le mot neutre *chauffeur* devient négatif par l'ajout du suffixe *-ard* dans *chauffard* (voir chapitre 2, § 3.1.3.3). Pupier observe que les préfixes *mal-* et *im-* (et ses variantes *il-* et *ir-*), ainsi que les suffixes *-ard*, *-eux*, *-asse* et *-âtre* sont des affixes à effet généralement péjoratif. Néanmoins, il note de façon non exhaustive un certain nombre d'exceptions pour chacun des affixes : *brouillard*, *crevasse*, etc.

Ces conclusions sur le rôle des affixes du français n'ont jamais été utilisées comme hypothèses pour des méthodes informatiques visant à améliorer la couverture d'un vocabulaire de mots évaluatifs. Nous proposons de mesurer l'intérêt de ces hypothèses (hypothèses 4.1 à 4.6) dans une démarche computationnelle.

Hypothèse 4.1 : Les mots ayant un suffixe en *-ard* sont probablement négatifs.

Hypothèse 4.2 : Les mots ayant un suffixe en *-eux* sont probablement négatifs.

Hypothèse 4.3 : Les mots ayant un suffixe en *-asse* sont probablement négatifs.

Hypothèse 4.4 : Les mots ayant un suffixe en *-âtre* sont probablement négatifs.

Hypothèse 4.5 : Les mots ayant un préfixe en *mal-* sont probablement négatifs.

Hypothèse 4.6 : Les mots ayant un préfixe en *im-* (ou ses variantes *il-* et *ir-*) sont probablement négatifs.

Notre objectif est double :

- améliorer la couverture du vocabulaire de mots évaluatifs du français ;
- analyser la pertinence des hypothèses 4.1 à 4.6. Si certaines de ces hypothèses sont valides dans une démarche automatique, elles pourraient être intéressantes à l'avenir pour identifier le sens axiologique de mots nouveaux ou inconnus.

Dans un premier temps, nous collectons des mots possédant l'un des huit affixes cités à partir d'un dictionnaire en ligne (§ 6.3.2). La tâche consiste ensuite à catégoriser ces mots en deux classes, *mot évaluatif négatif* ou *non*, à l'aide d'une méthode d'apprentissage supervisée (§ 6.3.3). Nous mesurons l'intérêt de chaque hypothèse en évaluant hors contexte et en contexte l'apprentissage réalisé (§ 6.3.4).

6.3.2 Extraire des candidats à partir d'un dictionnaire

En fouille d'opinions, les approches par dictionnaire ou par ontologie ont souvent été utilisées pour construire des listes de mots subjectifs (Banea et collab., 2008) ou des listes de mots positifs et

négatifs (Kamps et collab., 2004 ; Takamura et collab., 2007). Notre travail s'inspire en particulier de l'approche de Banea et collab. (2008).

Banea et collab. (2008) ont pour objectif de construire une liste de mots subjectifs roumains afin d'alimenter un classifieur de phrases subjectives à base de règles (ils n'identifient pas la polarité des mots subjectifs). À partir d'un dictionnaire en ligne et d'une liste de 60 mots grains subjectifs roumains, leur méthode consiste à extraire tous les candidats du dictionnaire dont la définition contient l'un des mots grains. Cette opération est répétée plusieurs fois en ajoutant les mots candidats extraits à la liste de mots grains. Dans le but de diminuer le bruit, une étape de filtrage consiste ensuite à mesurer en corpus une similarité entre chaque candidat extrait et les mots grains. Cette similarité est basée sur une méthode d'analyse sémantique latente (LSA) (Dumais et collab., 1988) et se rapproche de l'hypothèse employée par Turney (2002) : plus un mot w co-occure fréquemment avec un mot grain subjectif et plus w est probablement subjectif (voir hypothèse 1, §. 6.1.3).

De notre point de vue, les résultats obtenus par Banea et collab. (2008) pour la catégorisation de phrases subjectives (précision de 62 % et rappel de 69 %) montrent que le vocabulaire subjectif acquis est trop bruité pour être exploitable dans une approche à granularité fine. En intégrant des hypothèses linguistiques sur les affixes, nous espérons ainsi diminuer le bruit des approches par dictionnaire. Notre travail se distingue également de celui de Banea et collab. (2008) sur les aspects suivants :

- nous n'utilisons pas de liste de mots grains mais des hypothèses morphologiques pour extraire les candidats évaluatifs du dictionnaire ;
- nous utilisons une méthode d'apprentissage supervisée pour filtrer les candidats évaluatifs.

Wiktionnaire : un dictionnaire riche en formes fléchies Nous utilisons le *Wiktionnaire*⁵, branche francophone du projet Wiktionary. Ce dictionnaire est libre et contient environ un million d'entrées pour le français, soit environ 180 000 lemmes et 840 000 flexions. Il a la particularité de pouvoir évoluer rapidement par l'intermédiaire de contributions d'internautes et ainsi contenir des mots récents issus de néologismes. Pour comparaison, les dictionnaires papiers pour le français contiennent moins d'entrées :

- Robert Junior 1999 : 20 000 mots ;
- Petit Larousse 2009 : 59 000 mots ;
- Petit Robert 2001 : 60 000 mots ;
- Grand Robert 2001 : 100 000 mots ;
- Lexique *Morphalou* (issu de la nomenclature du Trésor de la Langue Française) : 539 413 formes fléchies, appartenant à 68 075 lemmes.

Chaque entrée du wiktionnaire est composée d'une étymologie et de sa ou ses définition(s) (23).

(23) **Entrée : Journaloux**

Étymologie : Dépréciatif dérivé de journal avec le suffixe **-eux**.

Définition : Nom commun masculin. 1. (Familier) (Péjoratif) Journaliste.

La domesticité du secrétaire particulier, la prostitution du journaloux et la négritude du romancier ou du dramaturge sont les formes diverses prises par le prolétariat de la plume. (Pierre Michel, Mirbeau & la « négritude », 2004, Éditions du Boucher)

Dans le cadre de notre travail, ce dictionnaire est intéressant car un effort particulier a été réalisé pour que les mots possédant des affixes y soient largement représentés. En revanche, l'étymologie

5. <http://fr.wiktionary.org>

donnée par le wiktionnaire ne renseigne pas toujours explicitement sur la polarité axiologie et la présence d'un affixe.

Du point de vue de sa disponibilité, la base de données de Wiktionnaire pour le français est libre d'utilisation⁶. Elle est interrogeable au moyen de requêtes SQL pour obtenir toutes les entrées du dictionnaire. Nous collectons ainsi, en première approche, un corpus brut où chaque document correspond à une définition et l'étymologie d'une entrée du dictionnaire commençant ou se terminant par l'un des huit affixes considérés dans notre travail (même s'il ne s'agit pas d'une construction étymologique). Cette première liste est constituée de 46 159 entrées du Wiktionnaire dont la répartition par affixe est présentée dans le tableau 6.9.

	-ard	-eux	-âtre	-asse	mal-	im-	ir-	in-	total
nombre d'entrées	1 297	1 432	83	17 681	3 475	3 780	16 844	1 567	46 159

TABLE 6.9 : Nombre d'entrées du Wiktionnaire selon leur affixe

Le nombre d'entrées pour chaque affixe paraît à première vue élevé. Il faut préciser que dans cette première liste, apparaissent également :

- des noms propres : *Champagnac-le-Vieux*, *Irak*, *Imola*, etc.
- des flexions : *irradierions*, *irradieriez*, *décantasse*, etc.
- des variantes orthographiques : *irakien*, *iraquien*, etc.
- des noms communs dont la racine correspond fortuitement à un affixe : *guépard*, *malaisien*, *image*, etc.
- des adjectifs correspondants aux noms d'habitants d'un lieu géographique : *savoyard*, *chamoniard*, *irakien*, etc.

À partir de cette liste de 46 159 mots, nous souhaitons désormais filtrer le bruit et les exceptions afin de conserver uniquement les mots évaluatifs négatifs. Le filtrage est effectué à l'aide d'une phase d'apprentissage supervisé.

6.3.3 Construire un modèle de classification

Données d'entraînement Pour les besoins du protocole expérimental, 5 juges humains ont catégorisés manuellement un sous-ensemble des 46 159 candidats évaluatifs extraits avec la répartition détaillée dans le tableau 6.10. Chaque juge a catégorisé 280 entrées et 100 entrées ont été catégorisées par tous les juges. Pour chacun des mots extraits, les juges humains doivent décider si le candidat est : *un mot évaluatif négatif* ou *non*. L'accord inter-annotateur kappa de Fleiss (Fleiss, 1971) est calculé à partir des 100 entrées catégorisées par chaque juge. L'indice moyen obtenu est de 0.84, ce qui correspond à un bon taux d'accord.

	-ard	-eux	-âtre	-asse	mal-	im-	ir-	in-	total
nombre d'entrées	140	140	20	140	140	140	140	140	1 000

TABLE 6.10 : Nombre de candidats évaluatifs du Wiktionnaire catégorisés manuellement

6. Téléchargeable à l'adresse <http://download.wikimedia.org/fr/wiktionary/20101014/>

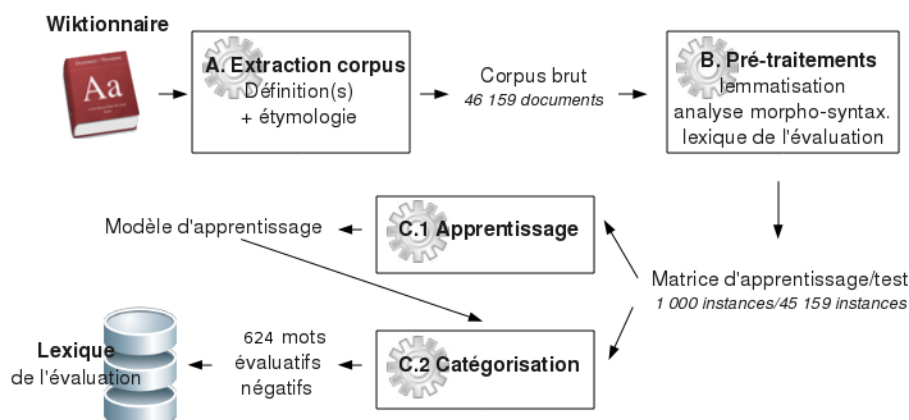


FIGURE 6.4 : Chaîne de traitement pour collecter et filtrer les candidats évaluatifs du wiktionnaire

Pré-traitements et modèle de classification En amont de la génération d'un modèle de classification, plusieurs pré-traitements sont effectués sur chaque document du corpus Wiktionnaire :

- un découpage en mot ;
- un étiquetage des rôles grammaticaux ;
- une lemmatisation (Schmid, 1994) ;
- une annotation des entrées du lexique de l'évaluation initial.

Le fonctionnement général de la chaîne de traitement réalisée est présenté dans la figure 6.4. Le modèle de classification est défini par un ensemble d'instances d'apprentissage et d'attributs :

- chaque document d_i appartenant à l'ensemble D des documents du corpus est une **instance** du modèle ;
- l'ensemble W des formes lemmatisées des mots apparaissant dans les définitions et les étymologies de D (les mots fonctionnels sont exclus) constitue les **attributs** du modèle ;
- un document possède un ensemble de **descripteurs** $w_{i,j}$, où i correspond à l'indice du document d dans la collection et j l'indice du mot dans l'ensemble W . $w_{i,j}$ a pour valeur 1 si w_j est présent dans le document d_i et 0 dans le cas contraire ;
- un document possède également deux **descripteurs** indiquant la fréquence du vocabulaire positif et la fréquence du vocabulaire négatif (éq. 6.3).

$$TF_{\text{polarité},i} = \frac{\text{NB-mots}_{\text{polarité},i}}{\text{NB-mots}_i} \quad (6.3)$$

Un modèle de classification non linéaire SVM (Joachims, 1997) est entraîné à partir des 1 000 entrées catégorisées manuellement et de leurs descripteurs. À partir d'une phase de dix validations croisées, nous effectuons une première évaluation hors-contexte (Tab 6.11).

	Précision	Rappel
Mot évaluatif négatif	89,8 % (97/108)	91,5 % (97/106)
Mot non évaluatif	98,9 % (883/892)	98,7 % (883/894)

TABLE 6.11 : Évaluation hors-contexte de l'enrichissement lexical par hypothèses morphologiques

Sur les 1 000 entrées catégorisées manuellement, uniquement 106 sont des mots évaluatifs négatifs d'après les juges (10,6%) (Tab. 6.11). Lors des différentes validations croisées, la méthode

SVM permet d'identifier 97 de ces mots (soit un silence de 8,5 %). Pour enrichir le vocabulaire sans y ajouter de bruit, la mesure de précision des mots évaluatifs négatifs est le score qui nous intéresse particulièrement. Hors-contexte, la meilleure configuration de l'apprentissage SVM obtient une précision de 89,8 %. Une évaluation hors-contexte nous semblant toujours comporter un biais, nous expérimentons en contexte l'apprentissage effectué. Nous en présentons les résultats dans la sous-section suivante. Dans cette expérimentation, le modèle SVM est appliqué à la catégorisation des 45 159 autres mots à traiter.

6.3.4 Expérimentation et résultats

Données en sortie À partir des 45 159 mots à catégoriser, le modèle de classification SVM permet d'ajouter 624 nouveaux mots négatifs non présents dans le lexique de l'évaluation initial. Le tableau 6.12 donne le nombre de mots négatifs appris par affixe et des exemples sélectionnés aléatoirement.

-ard	-eux	-âtre	-asse	mal-	im-	ir-	in-
94	70	34	46	112	90	75	103
papelard	nationaux	parâtre	paperasse	malformer	imbrûlable	irresponsable	inmettable
tocard	journalaux	bellâtre	tiédasse	maladresse	immiséricorde	irrégularité	insolent
ringard	bigleux	acariâtre	feiniasse	malodorant	immodestie	irréconciliable	inapaisé
pétochard	footeux	idolâtre	dindasse	maltraiter	immodéré	irrespect	inadéquation
charognard	prétentieux	marâtre	blondasse	malignité	immoler	irrecevable	inamical
taulard	migraineux	douceâtre	tatasse	maladif	immondice	irrespirable	inamiable
furibard	vaniteux	saumâtre	merdasse	malcommode	immortel	irrespirabilité	inapprouvé
patriotard	pouilleux	grisâtre	fadasse	malfauteur	immuable	irritant	inappréciable
zonard	niaiseux	jaunâtre	ragougnette	malhabile	impair	irrecupérable	inapte
geignard	chichiteux	verdâtre	savantasse	malicieux	impondérable	irréfléchi	inassouvi
scribouillard	chialeux	rougeâtre	chiennasse	malversation	impopulaire	irréformable	inattentif
smigard	crasseux	brunâtre	connasse	mal léché	importuner	irrépréhensible	inaudible
racontard	crapuleux	noirâtre	chaudasse	mal en point	imposer	irrésolu	indiscret
marouflard	vaporeux	violâtre	grognette	mal-être	improbant	irrévérencieuse	indisposé
pleurnichard	boutonneux	bleuâtre	couillasse	malbarré	impromptu	irriter	individualisme
binoclard	poussiéreux	beigeâtre	bouillasse	mal-en-point	impropre	irisable	indécemment

TABLE 6.12 : Nombre de mots négatifs et exemples appris à partir d'hypothèses morphologiques

Évaluation en contexte L'évaluation en contexte est réalisée à partir des mêmes 5 000 billets et commentaires de blogs utilisées lors de l'expérimentation précédente (§. 6.2.4). Le protocole expérimental est identique : nous annotons automatiquement les passages évaluatifs en différenciant les passages évaluatifs détectés :

- à partir du lexique de l'évaluation initial ;
- à partir des 624 mots appris du Wiktionnaire (par la suite, *deuxième partie enrichie*).

En sortie de la chaîne de traitements, deux fichiers (format CSV) sont générés pour lister : d'une part, les évaluations détectées à partir du lexique de l'évaluation initial et d'autre part, les évaluations détectées à partir de la deuxième partie enrichie du lexique. La polarité axiologique et le contexte phrastique sont également renseignés selon le format CSV suivant : **passage évaluatif** ; polarité ; *contexte (phrase)*. Les exemples (24) et (25) présentent des passages évaluatifs détectés grâce à l'apprentissage des mots peu fréquents *ignominieux* et *ineptie*.

(24) **nouvelle intervention ignominieuse** ; négatif ; *À la moitié de la deuxième mi-temps, nouvelle intervention ignominieuse du chihuahua SARKO* ;

(25) **des inepties** ; négatif ; *Domenech nous ressort toute l'année des inepties sur sa supposée hiérarchie des gardiens [...]* ;

Accord inter-annotateurs Comme lors de l'expérimentation précédente (§. 6.2.4), deux juges-humains se sont chargés d'évaluer la précision des passages évaluatifs annotés grâce à la deuxième partie enrichie du lexique (Tab. 6.13). Le fichier CSV correspondant est divisé en cinq échantillons (A, B, C, D et E) répartis entre le juge humain 1 (A, B, C) et le juge humain 2 (C, D et E).

	Juge 2 : Correct	Juge 2 : Erreur	Total
Juge 1 : Correct	380	10	392
Juge 1 : Erreur	16	22	36
Total	390	38	428

TABLE 6.13 : Tableau de contingence de l'accord observé

Sur l'échantillon C contenant 428 passages évaluatifs détectés à l'aide de la deuxième partie enrichie du lexique de l'évaluation, le taux d'accord observé (0.94 %) entre les deux juges est plus élevé que pour l'expérimentation précédente (0.76 %). Ce résultat laisse supposer que les termes appris contiennent ici moins de stéréotypes culturels instables. Quelques désaccords interviennent néanmoins sur des termes qu'un des juges estime positif ou ambigu (26).

- (26) « Une fille pour le diable » est un film qui demeure bien plus **malicieux** et subtil dans sa représentation [...].

Évaluation qualitative et quantitative Nous présentons les résultats qualitatifs de l'évaluation en contexte dans le tableau 6.14. En premier lieu, nous observons que la précision globale des passages évaluatifs annotés avec la deuxième partie enrichie du lexique est supérieure à la précision obtenue par la méthode automatique précédente (première partie enrichie) : 89,3 % contre 76,12 %. Elle est même équivalente à celle du lexique initial de l'évaluation construit manuellement. Les affixes semblent donc un indicateur intéressant pour détecter les mots négatifs de manière automatique. Concernant les suffixes et les préfixes, nous observons que :

- les suffixes sont globalement plus fréquents (63 %) que les préfixes (37 %) ;
 - la précision moyenne des suffixes (89,4 %) et des préfixes (89,0 %) est proche.
- Néanmoins, nous pouvons remarquer des disparités au sein des préfixes et des suffixes :
- les termes préfixés en -âtre et -asse sont rarement utilisés et leur précision est moins bonne (57,4 % et 75,0 %). Par exemple, *caillaise* est employé de façon non subjective dans le corpus ;
 - les termes préfixés en -ard et -eux sont des indicateurs précis ;
 - les termes suffixés en in- et ir- sont un peu moins fréquents que les termes suffixés en mal- et im-, mais leur précision est équivalente.

D'un point de vue quantitatif, nous notons que sur 5 000 documents, seulement 2 118 passages impliquant un affixe ont été rencontrés (soit en moyenne 0,42 terme par document). Ce résultat tend à montrer que même si les affixes sont un indicateur très précis, ils ont un impact relativement marginal au niveau du document. Le tableau 6.15 montre que par rapport au lexique de l'évaluation initial et à la première d'enrichissement, les hypothèses morphologiques ne permettent d'annoter que 2,4 % d'annotations supplémentaires. Néanmoins, dans une approche à granularité fine, nous pensons que chaque passage évaluatif, en particulier s'il active un mot évaluatif rare, est important à détecter. Par exemple, dans l'exemple (27), toute la difficulté du parcours de l'équipe de France en Coupe du Monde est résumée par un seul terme (*tortueux*) qu'il est important de ne pas rater, même si ce terme est assez peu fréquent (*tortueux* possède moins d'un million d'occurrences sur le web d'après Yahoo !Search).

- (27) [...] sur la route de l'Afrique du Sud, **un chemin tortueux** dont la première étape est fixée, déjà, début septembre.

échantillon	A	B	C	D	E	Total
TOTAL évaluations	401	371	428	462	456	2 118
PRÉCISION	89,6 %	90,5 %	88,8 %	86,8 %	91,0 %	89,3 %
-ARD évaluations	48	39	55	65	78	285
PRÉCISION	91,7 %	81,6 %	95,8 %	95,4 %	98,7 %	92,3 %
-EUX évaluations	68	63	71	79	80	361
PRÉCISION	94,1 %	95,2 %	87,3 %	91,5 %	97,5 %	93,4 %
-ÂTRE évaluations	2	0	19	24	9	54
PRÉCISION	50 %	- %	63,2 %	66,7 %	33,3 %	57,4 %
-ASSE évaluations	5	1	4	24	22	56
PRÉCISION	80,0 %	0 %	25,0 %	79,2 %	77,3 %	75,0 %
TOTAL préfixes	123	103	149	192	189	756
PRÉCISION	91,1 %	88,3 %	85,9 %	88,5 %	91,0 %	89,0 %
MAL- évaluations	77	85	84	93	74	413
PRÉCISION	87,0 %	92,9 %	90,5 %	88,2 %	83,8 %	88,6 %
IM- évaluations	65	79	87	90	97	418
PRÉCISION	86,2 %	84,8 %	88,5 %	85,6 %	92,8 %	87,8 %
IN- évaluations	70	64	49	50	62	295
PRÉCISION	87,1 %	95,3 %	93,9 %	90,0 %	96,8 %	92,5 %
IR- évaluations	66	40	59	37	34	236
PRÉCISION	93,9 %	95,0 %	93,2 %	78,4 %	79,4 %	89,4 %
TOTAL suffixes	278	268	279	270	267	1362
PRÉCISION	88,5 %	91,4 %	91,0 %	86,3 %	89,5 %	89,4 %

TABLE 6.14 : Précision obtenue avec la seconde partie enrichie du lexique

Lexique utilisé	nb entrées du lexique	nb passages annotés
Lexique de l'évaluation initial uniquement	982	68 536 (77,6 %)
1ère partie enrichie uniquement	2 474	17 669 (20,0 %)
1ère partie enrichie uniquement (correct)	2 474	13 450 (15,2 %)
2ème partie enrichie uniquement	624	2 118 (2,4 %)
2ème partie enrichie uniquement (correct)	624	1 951 (2,2 %)
<i>1ère partie enrichie</i> \cap <i>2ème partie enrichie</i>	106	196 (0,2 %)
<i>1ère partie enrichie</i> \cap <i>2ème partie enrichie (correct)</i>	106	190 (0,2 %)
Lexique total	3 974	88 318 (100 %)

TABLE 6.15 : Évaluation quantitative de l'enrichissement du lexique de l'évaluation

6.4 Conclusion

6.4.1 Synthèse

Dans ce chapitre, nous avons abordé la problématique de l'amélioration de la couverture des lexiques de mots évaluatifs (ou subjectivèmes). Trois types de subjectivèmes sont habituellement absents des ressources : les *syntagmes nominaux et verbaux*, les *stéréotypes culturels* et les *mots peu fréquents*. Les méthodes automatiques existantes d'extraction et de catégorisation de mots évaluatifs ne sont pas adaptées pour traiter ces cas. Les lexiques francophones et anglophones, qui résultent de ces méthodes ou qui sont développés manuellement, sont à l'origine de nombreux cas de silence lors d'analyses à granularité fine : de 40 % à 50 % des évaluations ne sont pas détectées selon nos expérimentations et celles de Wilson (2008).

Pour pallier ce problème, nous avons proposés deux méthodes fondées sur des hypothèses sémantiques et morphologiques pour extraire de nouveaux subjectivèmes. Une attention particulière a été portée sur la mesure en contexte de la précision de cet enrichissement automatique. Ces méthodes permettent d'améliorer le lexique de l'évaluation initial :

- dans sa première version, établie manuellement, le lexique de l'évaluation contenait 982 entrées permettant de repérer 50,1 % des passages évaluatifs d'un corpus multithématique avec une précision de 88,4 % ;
- la première méthode permet d'extraire 2 474 nouvelles entrées. Ces entrées ont une précision de 76,1 %, elles permettent de détecter 25,7 % de passages évaluatifs en plus. La nature de ces nouvelles entrées est intéressante puisqu'elles contiennent des syntagmes, des stéréotypes culturels et des termes peu fréquents ;
- La seconde méthode permet d'extraire 624 nouvelles entrées (dont 106 entrées en commun avec la première méthode d'apprentissage). Ces entrées ont une précision de 89,3 %, elles permettent de détecter environ 3 % de passages évaluatifs négatifs en plus. Cette amélioration ne nous semble pas négligeable dans le contexte des analyses à granularité fine. Elle permet également de tendre à ré-équilibrer la ressource pour mieux détecter les évaluations négatives.

Au final, le lexique de l'évaluation possède donc 3 974 entrées, ce qui lui permet de mieux rivaliser avec la taille des lexiques anglophones. De plus, contrairement à une majorité de lexiques du domaine, celui-ci a été évalué en contexte. Cela nous permet d'estimer le bruit qu'il va provoquer sur des corpus multithématique francophone de 10 % à 20 %.

6.4.2 Perspectives

Dans nos expérimentations, nous avons observé que l'instabilité sémantique de certains subjectivèmes concernent en particulier les noms et les syntagmes nominaux (précision en contexte de 72,0 % d'après le tableau 6.7). Dès lors, une des perspectives principales des travaux présentés dans ce chapitre est de désambiguïser dans leur contexte certains subjectivèmes trop instables. Un levier possible pour désambiguïser ces subjectivèmes consistent dans un premier temps à identifier quelle est la cible qu'ils évaluent. Par exemple, le mot *bijou* active un stéréotype subjectif positif lorsqu'il évalue une voiture (28) et il est objectif quand le contexte concerne une collection de bijoux (29).

(28) Aston Martin et Jeager Lecoultre collaborent pour un **bijou**. – évaluatif

(29) [...] collabore avec Topshop pour une collection de **bijoux**. – non évaluatif

Nous abordons la problématique de l'identification automatique de la cible d'une évaluation dans le chapitre suivant.

Identifier la cible d'une évaluation

7.1	Une problématique récente peu explorée	124
7.1.1	Quatre facteurs de complexité	124
7.1.1.1	<i>Différentes formes textuelles pour un unique objet</i>	125
7.1.1.2	<i>Les relations méronymiques entre objets</i>	125
7.1.1.3	<i>Présence de plusieurs objets candidats autour de l'évaluation</i>	126
7.1.1.4	<i>Proximité aléatoire entre l'évaluation et sa cible</i>	126
7.1.1.5	<i>Structurer les textes selon les objets et leurs évaluations</i>	126
7.1.2	Approches existantes pour l'anglais	127
7.1.2.1	<i>La proximité</i>	127
7.1.2.2	<i>Les dépendances syntaxiques</i>	128
7.1.2.3	<i>Une combinaison de traits lexicaux et grammaticaux</i>	128
7.1.2.4	<i>La résolution des anaphores pronominales</i>	128
7.1.3	Adapter et renforcer les méthodes pour le français	129
7.2	Un corpus multithématique de référence	129
7.2.1	Protocole d'utilisation du corpus Blogoscopie	129
7.2.1.1	<i>Segmentation du corpus</i>	129
7.2.1.2	<i>Les couples cible-évaluation</i>	130
7.2.2	Observations statistiques	130
7.2.2.1	<i>Le nombre d'objets dans un énoncé</i>	130
7.2.2.2	<i>La proximité des évaluations et de leur cible</i>	131
7.3	Méthodes d'identification de la cible évaluée	132
7.3.1	Méthodes applicables à un niveau intraphastique	133
7.3.1.1	<i>Approche par combinaison d'indices lexicaux et grammaticaux</i>	133
7.3.1.2	<i>Approche syntaxique (baseline)</i>	134
7.3.2	Méthodes applicables à un niveau intra- et interphrastique	135
7.3.2.1	<i>Approche par mesure de saillance des objets cibles</i>	135
7.3.2.2	<i>Approche par proximité (baseline)</i>	136
7.4	Expérimentation et résultats	137
7.4.1	Résultats pour les cas intraphrastiques	137
7.4.2	Résultats pour les cas interphrastiques	138
7.5	Conclusion	139

Afin d'analyser des textes dont le contenu est de plus en plus hétérogène (blogs, forums, etc.), l'identification de la cible d'une évaluation correspond à un besoin criant. Cette problématique est très récemment abordée dans la littérature anglophone du domaine. Néanmoins, les méthodes existantes nécessitent d'être renforcées et adaptées au français. Par ailleurs, ce travail implique que les évaluations aient été correctement détectées et délimitées avant de pouvoir identifier leur cible.

Dans ce chapitre, nous nous intéressons spécifiquement à cette problématique et considérons que la détection et la délimitation des évaluations ont été réalisées en amont (voir chapitres 5 et 6). En premier lieu, nous présentons la complexité de l'identification de la cible d'une évaluation et motivons le développement de méthodes évaluées sur le français (§. 7.1). Pour adapter ce travail au français, nous nous appuyons sur le corpus Blogoscopie (voir chapitre 4, §. 4.1), seul corpus francophone disponible dans lequel les passages évaluatifs et leur cible sont annotés (§. 7.2). Nous proposons quatre méthodes (dont deux méthodes *baseline*) pour identifier la cible d'une évaluation (§. 7.3). Nous les expérimentons sur une sous-partie du corpus Blogoscopie et discutons des résultats obtenus (§. 7.4).

7.1 Une problématique récente peu explorée

La cible d'une évaluation L'identification de la cible d'une évaluation a souvent été ignorée ou considérée comme un aspect de second plan. Par exemple, les travaux de catégorisation de l'opinion au niveau du document considèrent qu'un texte est monothématique et n'évalue qu'un seul objet donné (*un film, un livre, un appareil photo*, etc.). Ce n'est que récemment que quelques travaux anglophones ont directement axés leur étude sur cette problématique (Kessler et Nicolov, 2009 ; Jakob et Gurevych, 2010). Ils ont introduit le terme *target*, traduit ici par *cible*, pour désigner l'objet concerné par une évaluation. Dans la pratique, ils ont limité cette cible à quelques catégories d'objets :

- un produit particulier ou une marque (*iPhone, EOS 5D de Canon, Apple, Matrix*, etc.) ;
- une caractéristique d'un produit (*la durée de vue de la batterie, le scénario d'un film*, etc.).

En fait, de part le large éventail d'objets qu'elle peut couvrir, la notion de cible est complexe à circonscrire. Elle est d'ailleurs souvent laissée floue dans les travaux du domaine. Dans notre cadre, nous considérons que tout objet peut être la cible d'une évaluation. Nous tâchons de mieux appréhender cette notion en présentant quatre facteurs qui rendent complexe la problématique d'identification de la cible d'une évaluation (§. 7.1.1). Plusieurs approches ont été proposées mais celles-ci ne résolvent que partiellement la tâche et n'ont été appliquées qu'à l'anglais (§. 7.1.2). Nous précisons leurs limites et motivons le travail présenté dans ce chapitre (§. 7.1.3).

7.1.1 Quatre facteurs de complexité

L'identification de la cible d'une évaluation consiste à relier un passage évaluatif avec l'objet du monde directement concerné. Dans (1), il y a ainsi trois passages évaluatifs (*aller mal, honte, déception*) qui affectent un objet cible unique (*L'équipe de France*).

- (1) L'équipe de France va mal. Entre honte et déception.

Néanmoins, cette tâche s'avère particulièrement complexe. En particulier, le clivage entre *signifiant* et *signifié* introduit par Saussure (1916) est un aspect primordial du problème. L'objet évalué (la cible) est un concept signifié représenté dans le texte par des signifiants textuels arbitraires. Par exemple, dans (1), nous pouvons nous interroger sur le concept réellement signifié par la signifiant *équipe de France* : de quelle « *équipe de France* » s'agit-il ? quel est l'objet du monde correspondant ? est-ce le même objet que dans (2) ?

- (2) Cette équipe de France nous a fait plaisir. Elle avait une âme et un coeur. Je suis heureux pour le tennis français.

Dès lors, à partir d'un texte, on ne peut identifier la cible d'une évaluation qu'à travers ses signifiants. Il s'agit néanmoins de retrouver celui qui permet de discriminer l'objet du monde évalué avec le moins d'ambiguïté. Certains signifiants, comme les entités nommées, permettent en général de discriminer plus facilement le concept signifié. La cible est ainsi mieux identifiée dans (3) que dans (4).

- (3) J'adore jouer à la Wii
- (4) la console a l'air super !!

Nous précisons quatre facteurs qui rendent complexe la tâche d'identification de la cible d'une évaluation ci-dessous.

7.1.1.1 Différentes formes textuelles pour un unique objet

Premier facteur, un même objet peut être représenté dans le texte par différentes expressions nominales ou pronominales. Dans (5) et (6), on parle de l'objet du monde *équipe de France de football* par une variante nominale métonymique (*les Bleus*) et par une anaphore pronominale (*elle*). Dans un but applicatif, il apparaît nécessaire de regrouper les évaluations qui portent sur le même objet et de nommer celui-ci le plus précisément possible. On ne pourra ainsi pas considérer le pronom *elle* comme la cible de l'évaluation *séduisante*. *Les Bleus* est une réponse intermédiaire plus acceptable pour identifier l'objet cible évalué (*l'équipe de France de football*). Sur cet aspect, la résolution des anaphores nominales et pronominales est un défi majeur.

- (5) [...] lors du fiasco des Bleus en juin.
- (6) Elle a enfin été séduisante [...].

7.1.1.2 Les relations méronymiques entre objets

Les objets sont potentiellement liés à d'autres objets par des relations méronymiques¹. Dès lors, même si évaluer un méronyme A d'un mot B peut être une façon d'évaluer indirectement B, il importe de considérer le méronyme A comme la cible exacte de l'évaluation. Dans (7), (8) et (9) on évalue tout d'abord *la défense*, *les coups de pied arrêtés* et *le marquage sur les corners* de *l'équipe de France* et non *l'équipe de France* dans sa globalité.

- (7) Disons-le clairement : la défense française n'a pas été très bonne, ce soir.
- (8) Elle se révéla insuffisante sur les coups de pied arrêtés
- (9) Le marquage sur les corners reste encore approximatif.

L'identification des relations méronymiques entre objets est un autre aspect majeur du problème. Dans le domaine de la fouille d'opinion, elle a été abordée notamment par Kobayashi et collab. (2007) sur un corpus de critiques de restaurants en japonais. Néanmoins, le passage à des corpus multithématiques est actuellement non résolu. Si représenter les relations méronymiques entre objets dans des ressources ontologiques est une option envisagée, certaines relations méronymiques sont en revanche particulièrement complexes et instables. Dans (10), *la commission d'enquête* est ainsi reliée sémantiquement à l'objet *équipe de France* en raison de l'actualité.

- (10) La commission d'enquête a accouché d'une souris.

1. Un méronyme A d'un mot B est un mot dont le signifié désigne une sous-partie du signifié de B.

7.1.1.3 Présence de plusieurs objets candidats autour de l'évaluation

Le troisième facteur est la présence de plusieurs objets distincts dans le contexte d'un même passage évaluatif. Dans (11), les deux passages évaluatifs portent sur des cibles distinctes qu'il faut pouvoir déterminer parmi les quatre objets présents dans la phrase : *la dernière finale de la coupe du monde, l'équipe de France, le coursier, les pizzas*. L'objet le plus proche de l'évaluation n'est pas nécessairement sa cible.

- (11) [...] pizzas commandées lors de la dernière finale de coupe du monde avec l'équipe de France, coursier courageux, pizzas aussi banales que le match.

Cet aspect est sans doute l'un des aspects les plus importants de la problématique. Il s'agit de relier les évaluations aux bonnes cibles pour éviter les erreurs d'analyses automatiques et les contre-sens. Dans un énoncé, et en particulier sur les blogs, l'énonciateur entremêle plusieurs objets distincts dans son discours. Nous présentons des statistiques sur ce point dans la section 7.2.2.

7.1.1.4 Proximité aléatoire entre l'évaluation et sa cible

Dernier facteur de complexité, la cible évaluée ne se situe pas toujours à proximité de l'évaluation. Dans (12), le passage évaluatif *de la provocation*² porte sur *la ligne du sélectionneur* qui se trouve dans la phrase précédente. Le pronom *ce* en est une anaphore. Néanmoins, la présence de nombreux autres objets dans le contexte (*un iota, le jeu, les joueurs, la concrétisation de cette vue*) rend complexe l'identification de la cible de l'évaluation pour une approche automatique. Dans (13), l'évaluation *un ouf de soulagement* porte sur la réaction de *l'équipe de France*, mais aucune anaphore pronominale ou nominale n'est présente dans la phrase contenant l'évaluation. Nous présentons des statistiques sur la distance aléatoire entre l'évaluation et sa cible dans la section 7.2.2.

- (12) [...] force est de constater que la ligne du sélectionneur n'a jamais bougé d'un iota : le jeu appartient aux joueurs. Ce qui a souvent pu passer pour de la provocation n'est en fait que la concrétisation de cette vue.
- (13) C'est une réaction d'orgueil mais l'équipe de France devait se rassurer et rassurer son public [...]. C'était un devoir professionnel. Un ouf de soulagement.

En synthèse, nous proposons dans le paragraphe suivant une démarche pour identifier la cible d'une évaluation. L'objectif est de structurer les textes analysés selon les objets et les évaluations qu'ils contiennent.

7.1.1.5 Structurer les textes selon les objets et leurs évaluations

Pour résoudre la problématique d'identification de la cible d'une évaluation, la démarche générale doit prendre en compte les quatre facteurs présentés précédemment. Le processus consiste alors à :

- repérer les différents objets impliqués dans un énoncé. Ceux-ci sont généralement représentés textuellement par des groupes nominaux ou pronominaux (tâche T_1) ;
- identifier l'objet cible qui est directement concerné par une évaluation (tâche T_2) ;
- résoudre les relations d'anaphores nominales ou pronominales entre les objets (tâche T_3) ;
- résoudre les relations méronymiques entre les objets (tâche T_4).

À terme, l'objectif est de structurer les évaluations exprimées dans un corpus autour de plusieurs objets. Par exemple, le tableau 7.1 synthétise les évaluations présentes dans les exemples de cette section.

2. Ce passage est considéré comme une évaluation même s'il s'agit d'une forme de discours rapporté.

Cible	Formes textuelles	Évaluations
Équipe de France de foot	l'équipe de France	va mal, honte, déception, fiasco,
	les Bleus	séduisante, un ouf de soulagement
<i>Défense</i>	la défense française	n'a pas été trop bonne,
	le marquage sur les corners	reste encore approximatif
<i>Attaque</i>	les coups de pied arrêtés	se révéla insuffisante
<i>Enquête de la fédération</i>	la commission d'enquête	a accouché d'une souris
<i>Raymond Domenech</i>	le sélectionneur	de la provocation
Un coursier	coursier	courageux
Des pizzas	pizzas	banales
Équipe de France de tennis	l'équipe de France	nous fait plaisir, a une âme et un coeur
	le tennis français	je suis heureux
Wii	la Wii, la console	j'adore jouer, a l'air super !!

TABLE 7.1 : Exemple de structuration des évaluations et des cibles d'un corpus

Dans le domaine, les travaux sur cette problématique se sont principalement intéressés à repérer les signifiants textuels des objets d'un énoncé et à identifier la cible concernée par une évaluation (tâches T_1 et T_2) (Popescu et Etzioni, 2005 ; Zhuang et collab., 2006 ; Kessler et Nicolov, 2009). Un seul travail récent aborde la question des anaphores pronominales (tâche T_3) (Jakob et Gurevych, 2010). Enfin, la résolution des relations méronymiques entre objets évalués est très peu abordée de manière automatique (tâche T_4) (Kobayashi et collab., 2007). Dans nos travaux, nous nous intéressons en particulier aux trois premières tâches. Nous présentons une synthèse des approches existantes pour réaliser ces tâches dans les paragraphes suivants.

7.1.2 Approches existantes pour l'anglais

Parmi les travaux du domaine sur l'identification de la cible d'une évaluation, nous distinguons quatre approches. Les méthodes s'appuyant :

- sur la proximité d'une cible et d'une évaluation ;
- sur les chemins de dépendances syntaxiques entre une cible et une évaluation ;
- sur une combinaison de traits lexicaux et grammaticaux ;
- sur la résolution des anaphores pronominales.

7.1.2.1 La proximité

La proximité est le premier aspect considéré pour relier une cible et une évaluation. Ainsi, dans les travaux de Grefenstette, Qu, Shanahan et Evans (2004), sur les textes journalistiques politique, et Mishne et Glance (2006), sur les blogs de critiques de films, les auteurs considèrent deux types d'objets cibles :

- le nom d'une personnalité politique ;
- le titre d'un film.

Dans leurs travaux, ces cibles sont signifiées dans le texte par une entité nommée (*George Bush*, *Matrix*, etc.). Tous les mots évaluatifs d'un document qui co-occurrent avec cette entité nommée, dans une certaine fenêtre de mots, sont considérés comme des évaluations qui portent sur la cible. Cette tâche n'étant pas le coeur de leur travail, l'efficacité de cette hypothèse n'est pas évaluée directement.

7.1.2.2 Les dépendances syntaxiques

Les dépendances syntaxiques entre une évaluation et une cible sont une deuxième solution envisagée. Plusieurs travaux se sont intéressés aux chemins de dépendances entre une cible et une classe grammaticale spécifique de mot évaluatif :

- Hu et Liu (2004) ont extrait les chemins de dépendances syntaxiques entre un adjectif et une cible dans un corpus de critiques de films ;
- Bethard et collab. (2004) se sont spécifiquement intéressés à la position des verbes évaluatifs (*to believe, to argue, to accuse, etc.*) dans un corpus d'articles de presse pour identifier la source et la cible.

Bloom, Garg et Argamon (2007) et Popescu et Etzioni (2005) ont développé manuellement des listes de chemins de dépendances syntaxiques entre une évaluation et sa cible dans des corpus de produits commerciaux. La précision obtenue varie autour de 0,7, mais le rappel est faible : entre 0,3 et 0,4 selon les expérimentations. La majorité de ces approches ont été entraînées sur des corpus monothématiques ce qui tend à restreindre la richesse des cas observés. De plus, ces méthodes ne sont pas applicables dans les nombreux cas où la cible se situe dans une phrase différente de l'évaluation. Kim et Hovy (2006b) ont par ailleurs montré que prendre en compte uniquement les dépendances syntaxiques ne suffit pas à déterminer la bonne cible et qu'elle est parfois confondue avec la source de l'évaluation.

7.1.2.3 Une combinaison de traits lexicaux et grammaticaux

Pour améliorer les approches syntaxiques, Kessler et Nicolov (2009) propose une méthode qui combine des traits lexicaux et grammaticaux et qui tire profit d'une phase d'apprentissage supervisé. Ils ont annoté manuellement un corpus de critiques de produits technologiques (14) et de critiques de voitures. Celui-ci contient environ 4 900 évaluations et 25 000 objets annotés.

- (14) John recently purchased a digital camera. It had a great zoom lens, a mildly disappointing flash and was very compact.

À partir des exemples annotés, leur approche consiste à décrire chaque couple évaluation-objet du corpus par des traits lexicaux et grammaticaux (chemin de dépendances, chemin de classes grammaticales, catégorie grammaticale de l'évaluation et de la cible, nombre d'objets intercalés, cible évaluée, etc.) et à entraîner une fonction de préférence pour choisir le meilleur objet candidat. Nous réutilisons cette méthode pour le français dans la section 7.3.1.1. Cette méthode obtient les meilleurs résultats pour l'anglais³, néanmoins elle se restreint toujours au grain phrase et ne traite pas les nombreux cas où la cible n'est pas présente dans la même phrase que l'évaluation. Par ailleurs, elle n'a été évaluée que dans les cas plus simples où l'évaluation est un adjectif ou un verbe.

7.1.2.4 La résolution des anaphores pronominales

Très récemment, Jakob et Gurevych (2010) sont à notre connaissance les seuls à sortir du grain phrase. Ils se sont intéressés spécifiquement à la résolution d'anaphore pronominale dans un corpus de critiques de films pour améliorer les résultats sur cet aspect. Ils utilisent la méthode de Zhuang et collab. (2006) basée sur des chemins de dépendances syntaxiques les plus fréquents entre une cible et une évaluation. Lorsque la cible est un pronom, ils utilisent un système de résolution d'anaphores pronominales pour l'anglais. Ils s'appuient également sur une hypothèse *ad-hoc* : dans le cas où le système ne trouve pas d'antécédent pour un pronom, le *film* est considéré comme l'antécédent. Ils montrent que leur méthode permet de détecter 5 % de cibles corrects en plus que Zhuang et collab.

3. Nous présentons ces résultats en comparaison à ceux de nos expérimentations dans la section 7.4.

(2006) (61,4 % au lieu de 55,4%). De notre point de vue, le fait d'étudier l'identification de la cible d'une évaluation sur un corpus de critiques de film nous semble simplifier trop fortement le problème, les textes analysés portant en général sur un seul objet principal.

L'étude des approches existantes, essentiellement expérimentées sur l'anglais, nous amène à motiver notre travail sur trois aspects que nous formulons dans le paragraphe suivant.

7.1.3 Adapter et renforcer les méthodes pour le français

Notre travail est motivé par, d'une part, la volonté d'adapter les travaux existants vers le français, de vérifier ainsi les conclusions obtenues et, d'autre part, l'objectif d'améliorer l'existant sur les trois aspects suivants :

- notre approche est **multithématique**. En effet, l'aspect monothématique des études existantes (souvent ciblées sur un genre particulier : les critiques d'objets commerciaux) tend à simplifier les cas linguistiques rencontrés et à diminuer fortement le nombre de cibles portentielles dans un document ;
- notre approche **ne se limite pas à quelques catégories grammaticales** d'évaluation. Les évaluations peuvent ici être des groupes nominaux, adjectivaux, verbaux ou adverbiaux ;
- nous utilisons **le texte comme grain d'étude**. Les cas où la cible n'est pas présente dans la même phrase que l'évaluation sont pris en compte.

Pour les besoins de notre travail, un corpus de référence multithématique où les passages d'évaluations et les cibles sont annotées est nécessaire. Nous présentons le protocole d'utilisation du corpus Blogoscopie dans la section suivante.

7.2 Un corpus multithématique de référence

Actuellement, le seul corpus francophone disponible où les évaluations et les cibles sont annotées est le corpus de blogs *Blogoscopie* (voir chapitre 4, §. 4.1). Comme nous l'avons évoqué dans le chapitre 4, les blogs sont intéressants de part leur richesse linguistique plus grande que les textes de critiques de produits. De nombreux objets distincts sont présents dans un même texte et la syntaxe des phrases est plus variée. Ils sont de notre point de vue un support idéal pour aborder la problématique de l'identification de la cible d'une évaluation sans limiter sa complexité. Dans un premier temps, nous définissons le protocole d'utilisation du corpus Blogoscopie (§. 7.2.1), puis nous présentons des observations statistiques liées à l'identification de la cible d'une évaluation (§. 7.2.2).

7.2.1 Protocole d'utilisation du corpus Blogoscopie

7.2.1.1 Segmentation du corpus

Dans le corpus Blogoscopie, seule la partie *entraînement* (200 billets de blogs et 614 commentaires) a fait l'objet d'une annotation des évaluations et de leur cible. Pour le travail présenté dans ce chapitre, nous nous appuyons donc uniquement sur cette partie (voir chapitre 4, §. 4.1.3) :

- 4 129 évaluations y ont été annotées et possèdent une cible ;
- 6 876 objets y ont été annotés.

Pour les besoins de nos expérimentations, nous segmentons à nouveau ce corpus :

- une partie *entraînement* : constituée de 3 909 couples évaluation-cible et de 5 584 concepts. Cette partie correspond aux 160 premiers documents du corpus ;

- une partie *test* : constituée de 1 000 couples évaluation-cible et de 1 292 concepts. Elle correspond aux 40 documents restants.

Nous renvoyons au chapitre 4 (§. 4.1) pour une description plus précise du corpus.

7.2.1.2 Les couples cible-évaluation

Dans le corpus, seules les évaluations axiologiques, qui correspondent aux modalités d'appréciation et de jugement, possèdent une cible. Par la suite, nous employons le terme *évaluation* pour faire référence à ces différentes modalités. Selon le modèle d'annotation, la cible d'une évaluation peut être un concept (*la politique, l'écologie, le sport, etc.*) ou une instance (*SNCF, Harry Potter, Barack Obama, etc.*). Ces objets peuvent être dits *concernés, associés* ou *non-concernés* selon leur degré d'importance dans l'énoncé (voir chapitre 4, §. 4.1). Par la suite, nous regroupons toutes ces catégories sous une seule étiquette : les *objets*. Après simplification, l'annotation du corpus est analogue à l'exemple (15). Une évaluation porte sur une unique cible. Cette cible correspond à la forme textuelle d'un objet présent dans l'énoncé.

(15) Le sommet a été atteint avec <objet>le nucléaire</objet> : <objet>M. Sarkozy</objet> <évaluation cible="M. Sarkozy">a abusé</évaluation> <objet>l'opinion publique</objet> en annonçant qu'il n'y aurait pas de <objet>"nouveaux sites"</objet>.

Nous rappelons que dans ce corpus :

- les anaphores pronominales n'ont pas été annotées ;
- la forme textuelle de l'objet cible la plus proche de l'évaluation est considérée comme la cible. Lorsqu'une anaphore pronominale est présente, l'antécédent nominal le plus proche est considéré comme la cible.

Dans le paragraphe suivant, nous présentons plusieurs observations statistiques sur le nombre d'objets dans un texte de blogs et la position de la cible par rapport à son évaluation.

7.2.2 Observations statistiques

En s'appuyant sur la partie *entraînement* du corpus, nous donnons un aperçu de la problématique d'identification de la cible d'une évaluation à partir d'observations statistiques. Ces observations concernent deux aspects :

- le nombre d'objets distincts dans un énoncé ;
- la proximité des évaluations et de leur cible.

7.2.2.1 Le nombre d'objets dans un énoncé

Sur le corpus considéré, 54 % des billets (et leurs commentaires) parle en général d'au moins deux ou trois objets principaux différents (Tab. 7.2). Ces objets ne sont pas reliés par une relation méronymique. Contrairement aux textes issus des critiques de films ou de produits, les textes de blogs sont donc rarement monothématiques ou centrés sur un seul objet principal. Seuls 17 % des documents issus des blogs n'évoquent qu'un seul objet principal.

Le tableau 7.3 montre que le nombre d'objets associés ou non concernés est quant à lui très variable. Dans 30 % des documents, il existe entre un et cinq objets « secondaires » (*associés* ou *non concernés*). Dans 38 % des cas, il en existe 6 à 15 et dans 32 %, il en existe plus de 16. Dit autrement, si dans une majorité de cas un blogueur articule le sujet de son billet autour de deux ou trois sujets principaux, en revanche, on peut difficilement prédire le nombre de sujets associés qu'il va évoquer. Pour identifier la cible d'une évaluation dans un énoncé, il s'agit donc de la trouver parmi une liste

d'objets pouvant aller de 2 à plus de 20. Dans notre corpus, les blogs qui contiennent le plus d'objet sont des textes appartenant aux thématiques *business* ou *guides des achats* pour lesquels un grand nombre de noms de produits ou de valeurs boursières sont cités. Néanmoins, faute d'un nombre de blogs annotés suffisamment représentatif, nous n'avons pas étudié plus précisément le nombre d'objets selon la thématique des blogs.

Nb objets principaux / billet	1	2	3	4	5+
billets concernés	17 %	35 %	19 %	14 %	15 %

TABLE 7.2 : Nombre d'objets principaux par billet

Nb objets secondaires / billet	1-5	6-10	10-15	16+
billets concernés	30 %	23 %	15 %	32 %

TABLE 7.3 : Nombre d'objets secondaires par billet

En moyenne, il y a 24 évaluations par document. Dans environ 80 % des documents, ces évaluations portent sur 4 à 8 objets distincts, qu'ils soient principaux ou secondaires. Dit autrement, même si un blogueur évoque un nombre d'objets importants, son discours évaluatif porte sur un nombre réduit de cibles.

7.2.2.2 La proximité des évaluations et de leur cible

Dans un second temps, nous nous intéressons à la position d'une évaluation par rapport à sa cible pour évaluer la complexité de la tâche. En particulier, nous observons le signifiant textuelle de la cible le plus proche (Tab. 7.4). Nous ne prenons pas en compte les anaphores pronominales. Dans ses travaux, Kessler et Nicolov (2009) rapporte que 14 % des signifiants textuels cibles les plus proches de l'évaluation sont des anaphores pronominales. Il ne travail que sur les 86 % restants.

Position relative de la cible	avant l'évaluation		après l'évaluation	
évaluations concernées	61 %		39 %	
Distance - en nombre de mots -	0	1-4	5-10	11+
évaluations concernées	21 %	20 %	11 %	48 %
Distance - en nombre de phrases -	0	1	2	3+
évaluations concernées	56 %	19 %	6 %	19 %
Distance - en nombre d'objets intercallés -	0	1	2	3+
évaluations concernées	53 %	7 %	12 %	28 %

TABLE 7.4 : Étude de la position de la cible par rapport à son évaluation

Dans notre corpus d'entraînement, le signifiant le plus proche se situe avant le passage évaluatif dans 61 % des cas. Ce résultat se démarque un peu de celui obtenu par Kessler et Nicolov (2009), où la cible se situe avant le passage évaluatif dans 51 % des cas. Nous comptabilisons ensuite le nombre de mots, le nombre de phrases et le nombre d'objets qui séparent une évaluation et sa cible.

Le nombre médian de mots entre l'évaluation et sa cible est de 6,4 mots dans notre corpus, la moyenne est de 14,1. Dans le corpus de Kessler et Nicolov (2009), la médiane est de 2 et la moyenne de 6,2. Cette comparaison montre que les cibles sont plus proches de leur évaluation en anglais ou que la nature des corpus de critiques étudiés par Kessler et Nicolov (2009) tend à diminuer la distance entre cible et évaluation. La statistique suivante va dans le même sens et révèle un aspect important de cette problématique :

- Kessler et Nicolov (2009) mesure que 91 % des cibles sont dans la même phrase que leur évaluation ;
- nous mesurons qu'uniquement 56 % des cibles sont dans la même phrase que leur évaluation dans notre corpus. Ce score semble indiquer que les anaphores pronominales utilisées pour signifier une cible sont beaucoup plus nombreuses dans notre corpus que dans celui de Kessler et Nicolov (2009).

Par ailleurs, dans 47 % des cas, il y a au moins un objet intercallé entre l'évaluation et sa cible. De ce point de vue, contrairement aux travaux état de l'art anglophones sur des corpus monothématiques, nous devons également traiter davantage de cas complexes où le signifiant textuel nominal d'une cible et l'évaluation ne sont pas à proximité. Cet aspect nous amène à distinguer deux niveaux de complexité pour identifier la cible d'une évaluation :

- lorsque le signifiant textuel nominal d'une cible se situe dans la même phrase que l'évaluation (niveau *intraphrastique*) ;
- lorsque le signifiant textuel nominal le plus proche se situe dans une autre phrase (niveau *interphrastique*).

Les méthodes existantes du domaine ont principalement été adaptées pour traiter les cas intraphrastiques. Dès lors, nous souhaitons d'une part adapter ces méthodes au français et d'autre part les renforcer pour mieux traiter les cas interphrastiques. Nous présentons ces méthodes dans la section suivante.

7.3 Méthodes d'identification de la cible évaluée

Les observations statistiques présentées précédemment nous amènent à distinguer deux niveaux de complexité pour identifier la cible d'une évaluation. Nous nous fixons alors deux objectifs :

- évaluer les méthodes existantes qui traitent uniquement les cas intraphrastiques afin de les valider sur le français, sur un corpus multithématique et en prenant davantage en compte les différentes catégories grammaticales des évaluations ;
- développer une méthode qui traite à la fois les cas intra- et interphrastiques.

Pour le premier objectif, nous adaptons au français la méthode état de l'art (**MRankSVM**, §. 7.3.1.1) et nous la comparons à une méthode *baseline* (**MSyntaxe**, §. 7.3.1.2). Pour le second objectif, nous proposons une méthode basée sur la saillance d'une cible (**MSaillance**, §. 7.3.2.1). Celle-ci ré-utilise les résultats de MRankSVM comme une amorce. Nous la comparons avec une méthode *baseline* (**MProximité**, §. 7.3.2.2). Pour ces quatre méthodes :

- les passages d'évaluations sont annotés manuellement (annotations présentes dans le corpus Blogoscopie) ;
- le TreeTagger (Schmid, 1994) donne la catégorie grammaticale ;
- un algorithme de découpage en syntagmes non-récursifs, réimplémentant partiellement la méthode de (Vergne et Giguët, 1998), permet d'identifier les groupes nominaux et pronominaux d'un texte. Ces expressions sont considérées comme l'ensemble des objets d'un texte.

7.3.1 Méthodes applicables à un niveau intraphrastique

Parmi les approches existantes pour traiter les cas intraphrastiques, nous adaptons au français une méthode par combinaison d'indices lexicaux et grammaticaux (§. 7.3.1.1). Cette méthode correspond à la méthode état de l'art pour l'anglais. Nous la comparons à une méthode basée uniquement sur les dépendances syntaxiques entre une évaluation et sa cible (§. 7.3.1.2).

7.3.1.1 Approche par combinaison d'indices lexicaux et grammaticaux

La méthode état de l'art de Kessler et Nicolov (2009) a été évaluée sur un corpus monothématique anglais après filtrage sur les évaluations de type verbe et adjectif uniquement. Elle consiste à combiner des informations lexicales et grammaticales pour pallier les limites des approches purement syntaxiques. Nous adaptons leur approche pour traiter un corpus multithématique français sans restriction grammaticale (par la suite, **MRankSVM**). Cette méthode s'applique à un niveau intraphrastique. Elle prend en entrée :

- s : une phrase à analyser ;
- e : un passage d'évaluation présent dans s ;
- O : un ensemble d'objets présents dans s ;
- f : une fonction d'ordonnement permettant de classer les concepts de s .

La fonction f est extraite par apprentissage sur le corpus d'entraînement. Nous utilisons pour cela l'algorithme RankSVM issu de SVMLight (Joachims, 2002). Le corpus d'entraînement fournit :

- 3 909 exemples *positifs* : ils correspondent aux couples *évaluation-cible* présents dans le corpus. Ces exemples ont un rang 1 ;
- 4 000 exemples *negatifs* : ils correspondent à des couples *évaluation-objet* dont l'objet est choisi dans le contexte de l'évaluation mais n'est pas la cible. Ces exemples ont un rang 0 ;

Caractéristique	Exemple
Chemin lexical lemmatisé entre l'évaluation et sa cible	<i>pour</i>
Chemin grammatical entre l'évaluation et sa cible	Prep.
Catégorie(s) grammatical(es) de l'évaluation	DET NOM ADJ
Catégorie(s) grammatical(es) de la cible	DET NOM
Chemin de dépendances syntaxiques	↓ PP, ↓ DP
Nombre d'objets entre l'évaluation et la cible	0
Nombre de mots entre l'évaluation et la cible	1
Rang	1

TABLE 7.5 : Caractéristiques utilisées pour décrire chaque couple évaluation-cible

Chacun de ces exemples est décrit par les caractéristiques présentées dans le tableau 7.5. Dans ce tableau, nous reprenons les caractéristiques du couple *une défaite majeure-l'écologie* présenté dans l'exemple (16). Pour atteindre la cible à partir de l'évaluation, il existe une dépendance prépositionnelle (PP), puis une dépendance sur un déterminant (DP). Ces dépendances syntaxiques sont déterminées automatiquement à partir de l'outil FIPS (Wehrli, 2007).

(16) [...] est **une défaite majeure** pour **l'écologie**.

À partir des exemples, RankSVM entraîne un modèle ayant pour objectif de déterminer une fonction de préférence qui maximise le *rang*. À l'image de Kessler et Nicolov (2009), nous utilisons les paramètres par défaut de RankSVM. Lors de la phase de test, cette méthode effectue un

ordonnancement de tous les objets qui apparaissent dans la même phrase que le passage d'évaluation en s'appuyant sur le modèle appris lors de la phase d'entraînement. L'objet classé premier est considéré comme la cible de l'évaluation.

7.3.1.2 Approche syntaxique (*baseline*)

Nous comparons la méthode MRankSVM avec une méthode qui repose uniquement sur une analyse des dépendances syntaxiques (par la suite, **MSyntaxe**). Elle ne fonctionne également que dans un cadre intraphrastique. Elle prend en entrée :

- s : une phrase à analyser ;
- e : un passage d'évaluation présent dans s ;
- O : un ensemble d'objets présents dans s ;
- C : un ensemble de chemins de dépendances syntaxiques entre une évaluation et une cible avec leur probabilité correspondante.

L'ensemble C est déterminé à partir du corpus d'entraînement. Nous utilisons l'analyseur de dépendances syntaxiques FIPS (Wehrli, 2007) sur les exemples de couples évaluation-cible qui apparaissent dans la même phrase dans le corpus d'entraînement. Pour ces exemples, FIPS fournit un chemin de dépendances (visualisable sous forme d'arbre syntaxique (Fig. 7.1)). Nous classons par ordre de fréquence décroissante (voir tableau 7.6), les chemins de dépendances qui vont :

- de la fin d'une évaluation vers le début d'une cible, si l'évaluation est située avant la cible ;
- du début de l'évaluation vers la fin d'une cible, si l'évaluation est située après la cible.

L'analyse des dépendances⁴ de (17) est visualisable dans la figure 7.1. Dans cet exemple, l'évaluation est séparée de sa cible par la dépendance \downarrow PP, \downarrow DP.

(17) **J'ai une véritable fascination** pour **J.K Rowling** [...]

Les chemins de dépendances syntaxiques les plus fréquents (Tab. 7.6) sont aussi les plus courts. Par exemple, l'évaluation correspond souvent à un syntagme adjectival dépendant de sa cible (\uparrow AP), comme dans *un beau film*.

Chemin de dépendance(s)	Fréquence	Exemple
\uparrow AP	121	C'est un beau film
\downarrow DP	91	Cette question me taraude
\downarrow PP, \downarrow DP	65	J'ai une véritable fascination pour J.K Rowling
\uparrow TP, \uparrow CP	59	Un événement qui contribue à alourdir le débat
\uparrow VP, \downarrow DP	45	Ce jeu est accessible pour les enfants de 12 ans
\uparrow AdvP, \downarrow DP	29	Cette histoire de travailler plus [...] ne tient [...] pas la route

TABLE 7.6 : Chemins de dépendances syntaxiques les plus fréquents entre une évaluation et une cible

Lors de la phase de test, cette méthode effectue une analyse syntaxique de la phrase contenant l'évaluation. Le chemin de dépendance syntaxique le plus fréquent c ($c \in C$) qui permet de relier l'évaluation avec un objet o ($o \in O$) désigne l'objet o comme la meilleure cible potentielle. En utilisant un des chemins syntaxiques, il est possible que o soit un groupe pronominal (*il, aucun d'entre eux, etc.*).

4. Étiquettes de FIPS : AP = syntagme adjectival ; DP = syntagme déterminant ; PP = syntagme prépositionnel ; AdvP = syntagme adverbial ; CP = syntagme conjonctionnel ; NP = syntagme nominal ; TP = syntagme temporel ; VP = syntagme verbal.

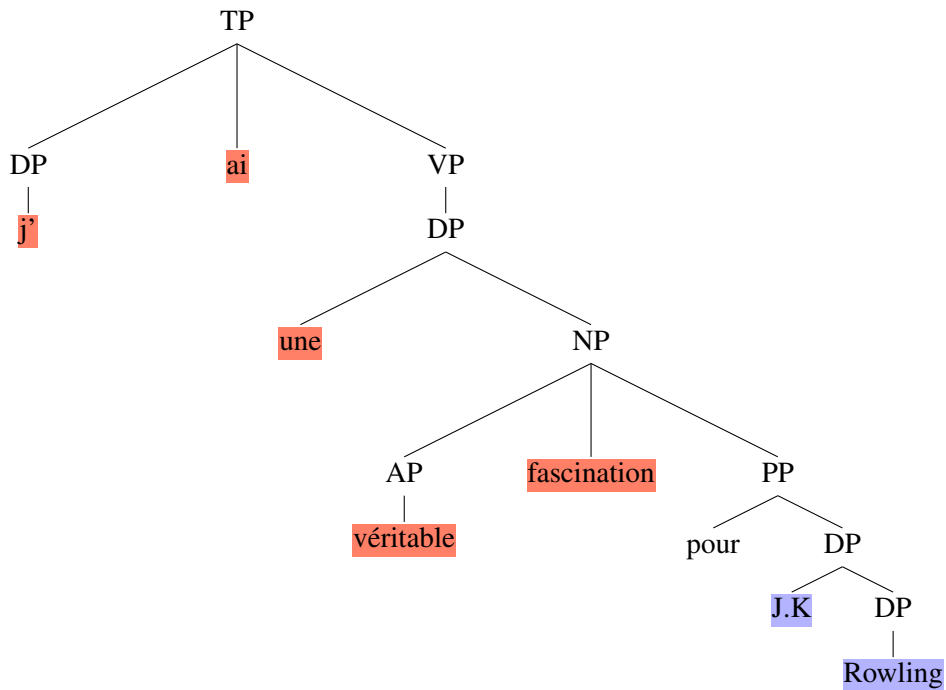


FIGURE 7.1 : Analyse des dépendances entre une évaluation et sa cible

7.3.2 Méthodes applicables à un niveau intra- et interphrastique

Notre second objectif consiste à proposer une méthode applicable à un niveau intraphrastique et interphrastique. Nous proposons une méthode (**MSaillance**) fondée sur la saillance d'une cible dans un énoncé. Nous la comparons à une méthode *baseline* (**MProximité**).

7.3.2.1 Approche par mesure de saillance des objets cibles

Couples forts et couples faibles Lors de la phase de test, en sortie de la méthode MRankSVM, pour chaque document d , nous disposons d'un ensemble de couples évaluation-cible déterminés automatiquement. Nous observons les quatre faits suivants :

- O_1 : certaines évaluations n'ont pas de cible affectée (si la phrase contenant l'évaluation ne possède aucun concept) ;
- O_2 : certaines évaluations ont pour cible un groupe pronominal ;
- O_3 : certains couples cible-évaluation ne sont pas reliés par une dépendance syntaxique couramment observée ;
- O_4 : certaines des cibles ne sont pas évaluées ailleurs dans le document.

Les couples qui vérifient l'une des observations O_1 ou O_2 , ou qui vérifient simultanément O_3 et O_4 sont considérées comme des couples *faibles*. Ils représentent des cas douteux. Les autres couples sont considérés comme des identifications *fortes* dont il est plus probable qu'elles soient correctes.

Utiliser les couples forts comme une amorce La méthode MSaillance consiste à utiliser les couples *forts* de MRankSVM comme une amorce pour corriger les couples *faibles*. Cette méthode a deux objectifs :

- identifier un signifiant nominal cible pour les évaluations qui n'en ont pas ou qui sont liées à un signifiant pronominal ;
- corriger l'affectation des autres couples faibles.

Notre approche s'appuie sur les hypothèses linguistiques suivantes :

- **hypothèse 1** : s'il n'existe pas de signifiant nominal pertinent dans la même phrase qu'une évaluation, celui-ci existe probablement dans les phrases voisines.
- **hypothèse 2** : l'énonciateur d'un discours subjectif articule ses évaluations au fil d'un document en évaluant les mêmes objets plusieurs fois. Plus un objet est la cible d'une évaluation, plus il est saillant et plus il est probable qu'il soit de nouveau la cible d'une évaluation dans le contexte.

Mesurer la saillance d'un objet cible La méthode MSaillance prend quatre paramètres en entrée :

- e : un passage d'évaluation présent dans le document d dont la relation évaluation-cible actuelle est considérée comme *faible* ;
- P : l'ensemble des phrases voisines de l'évaluation. Cet ensemble est limité empiriquement à quatre phrases avant et quatre phrases après e ;
- CF : l'ensemble des couples forts évaluation-cible contenus dans d issus de la méthode MRankSVM ;
- O : un ensemble d'objets présents dans P ;

Les objets de l'ensemble O sont préalablement lemmatisés et les mots fonctionnels sont retirés. Par exemple, *l'entraîneur de l'équipe de France* devient *entraîneur équipe France*). Nous mesurons la fréquence de chaque objet o dans le segment de phrases P et la fréquence de chaque objet o dans le document d . Ces deux fréquences nous permettent de mesurer la saillance d'un objet. Pour tout objet o ($o \in O$), nous mesurons une probabilité que o soit la cible de l'évaluation e avec la formule suivante :

$$P(o, e) = \frac{\text{NB}(o, \text{cible}, P)}{\text{NB}(o, P) \cdot \text{NB}(\text{cible}, P)} \cdot \frac{\text{NB}(o, \text{cible}, d)}{\text{NB}(o, d) \cdot \text{NB}(\text{cible}, d)} \quad (7.1)$$

$$\text{Cible}(e) = \text{MAX}(P(o, e)) \quad (7.2)$$

où :

- $\text{NB}(o, \text{cible}, P)$ est le nombre de fois où l'objet o est une cible forte d'une évaluation quelconque dans l'ensemble des phrases voisines de l'évaluation e . De la même façon, $\text{NB}(o, \text{cible}, d)$ est le nombre de fois où o est une cible forte d'évaluation dans le document d ;
- $\text{NB}(o, P)$ et $\text{NB}(o, d)$ sont les nombres de fois où l'objet o est présent dans le segment de phrases voisines P et dans le document d ;
- $\text{NB}(\text{cible}, P)$ et $\text{NB}(\text{cible}, d)$ sont les nombres de cibles d'évaluations dans le segment P et dans le document d .

Finalement, $P(o, e)$ représente la saillance de l'objet o comme cible d'évaluation dans le contexte de e . La cible de l'évaluation choisie par MSaillance est l'objet qui maximise cette probabilité. Nous comparons cette méthode à une approche fondée sur la proximité d'un objet et d'une évaluation.

7.3.2.2 Approche par proximité (*baseline*)

Indexation des mots d'un document La méthode par proximité (par la suite, **MProximité**) est applicable au niveau intraphrastique et interphrastique. Elle prend cinq paramètres en entrée :

- d : le document à analyser ;
- e : un passage d'évaluation ;
- $w_d(e)$: la position du mot de début de e dans le document d ;

- $w_f(e)$: la position du mot de fin de e dans le document d ;
- O : un ensemble d'objets présents dans d . Chaque objet est également associé à un mot de début ($w_d(o)$) et un mot de fin ($w_f(o)$).

La position dans le document est déterminée par une indexation de l'ensemble des mots du texte : $M = \{m_1, m_2, \dots, m_n\}$. Le passage d'évaluation e et l'ensemble des objets O sont délimités par un mot de début ($m_d \in M$) et un mot de fin ($m_f \in M$) dont les rangs dans le texte sont connus.

Mesure de la proximité d'un objet cible Pour tout $o \in O$, MProximité compte le nombre de mots qui séparent o de l'évaluation e dans le document d selon que l'objet o soit avant (éq. 7.3) ou après (éq. 7.4) l'évaluation. La ponctuation n'est pas prise en compte. Si e est inclu dans o , la distance en mots est égale à 0. Le choix du meilleur candidat s'effectue sur la base de la plus petite distance en mots (éq. 7.5). En cas d'égalité, la cible qui précède l'évaluation est préférée. Nous justifions ce choix par l'étude statistique réalisée précédemment (§. 7.2.2).

$$\forall o \in O \text{ se situant avant } e, \text{Proximité}(t) = w_f(o) - w_d(e) \quad (7.3)$$

$$\forall o \in O \text{ se situant après } e, \text{Proximité}(t) = w_f(e) - w_d(o) \quad (7.4)$$

$$\text{Cible}(e) = \text{MIN}(\text{Proximité}(o)) \quad (7.5)$$

Dans la section suivante, nous expérimentons les quatre méthodes présentées dans cette section sur la partie du corpus réservé aux tests. Nous comparons ces méthodes sur les cas intraphrastiques et interphrastiques et discutons des perspectives envisagées.

7.4 Expérimentation et résultats

Nous évaluons les 4 méthodes présentées précédemment sur les 40 documents du corpus réservé aux tests. Ce corpus contient 1 000 couples cible-évaluation :

- 592 de ces couples sont intraphrastiques ;
- 408 sont interphrastiques.

Pour chaque évaluation, nous considérons qu'il n'y a qu'une seule cible possible. Nous évaluons chaque méthode avec un score d'exactitude (eq. 7.6). La cible est considérée correcte si le signifiant de référence annoté dans le corpus et le signifiant déterminé automatiquement renvoient au même objet réel. Par exemple, si la référence est *la loi LRU* et que la cible trouvée est *LRU* ou *la loi sur les universités*, la réponse est acceptée.

$$\text{Exactitude} = \frac{\text{Nombre de couple cible-évaluation correct}}{\text{Nombre de couple cible-évaluation total}} \quad (7.6)$$

7.4.1 Résultats pour les cas intraphrastiques

Le tableau 7.7 présente les résultats obtenus par les quatre méthodes sur les 592 cas où la cible est dans la même phrase que l'évaluation.

Type	Intraphrastique	Interphrastique	Total
Score	Exactitude	Exactitude	Exactitude
MSyntaxe	68.8 % (407/592)	non applicable	40.7 % (407/1 000)
MRankSVM	71.5 % (423/592)	non applicable	42.3 % (423/1 000)
MProximité	53.5 % (317/592)	27.2 % (111/408)	42.8 % (428/1 000)
MSaillance	72.8 % (431/592)	60.0 % (245/408)	67.6 % (676/1 000)

TABLE 7.7 : Résultats obtenus par les quatre méthodes d'identification de la cible d'une évaluation

Un léger gain en combinant des traits lexicaux et grammaticaux Par rapport à l'approche syntaxique, la combinaison d'indices lexicaux et grammaticaux améliore légèrement l'identification de la cible (71,5 % contre 68,8 %). Cette différence est néanmoins beaucoup plus marquée dans les travaux de Kessler et Nicolov (2009) sur l'anglais. Dans leurs expérimentations, ils obtiennent un gain de 10 à 12 %. Selon nous, le gain important obtenu par Kessler et Nicolov (2009) peut s'expliquer par l'utilisation d'un corpus monothématique. Celui-ci a peut-être tendance à « figer » certaines tournures lexicales.

Adaptabilité de la méthode au français Par rapport à Kessler et Nicolov (2009) qui ont obtenu une précision de 74,8 %, notre score de précision baisse d'environ 3 points sur le français (71,5 %). S'il est difficile de dire si cette baisse est significative, elle tend néanmoins à montrer que la méthode MRankSVM est applicable pour traiter les cas intraphrastiques sur des corpus multithématiques en français.

Intérêt d'une amorce pour corriger les couples faibles intraphrastiques La méthode MRankSVM est d'autant plus intéressante qu'elle fournit de bons indices à la méthode MSaillance pour corriger certains couples intraphrastiques (72,8 % contre 71,5 %). L'hypothèse discursive consistant à reclasser les cibles d'opinions saillantes (celles qui sont fréquemment liées à une opinion) semble être une hypothèse intéressante à développer.

7.4.2 Résultats pour les cas interphrastiques

Intérêt de considérer la saillance d'un objet cible La méthode MSaillance permet d'identifier correctement 60,0 % des cibles d'évaluations en s'appuyant uniquement sur l'objet cible le plus saillant dans une fenêtre de 4 phrases autour de l'évaluation. Ce résultat est nettement supérieur à l'approche naïve consistant à prendre l'objet le plus proche (27,2 %). Ce score permet à la méthode MSaillance d'être beaucoup plus performante que la méthode état de l'art MRankSVM sur l'ensemble de la problématique d'identification de la cible d'une opinion (67.6 % contre 42,3 %). Ce score nous semble encore améliorable en introduisant d'autres types d'hypothèses sémantiques et discursives. Par exemple :

- si une cible A est évaluée fréquemment positivement dans un document, si une cible B est évaluée fréquemment négativement dans le même document, il est plus probable que A (respectivement B) soit reliée à une évaluation positive (respectivement négative) dont la cible est inconnue ou faible ;
- si une ou plusieurs cibles appartenant au même graphe d'objets (modélisant les relations méronymiques entre objets) sont évaluées dans le même passage d'un document, il est probable qu'une évaluation dont on ignore la cible porte sur un objet de ce graphe.

Comparaison avec des procédures de résolutions d'anaphores Ces hypothèses seraient intéressantes à comparer à des procédures de résolution d'anaphores pronominales et nominales plus classiques pour le français. Néanmoins, ces procédures nous paraissent à ce stade encore complexes à mettre en place sur des corpus multithématiques et en particulier sur les blogs dont le respect grammatical et orthographique est très variable. De ce point de vue, l'hypothèse de saillance d'un objet cible semble offrir une plus grande souplesse.

7.5 Conclusion

Dans ce chapitre, nous avons abordé la problématique d'identification de la cible d'une évaluation. Elle fait l'objet d'une récente attention de la communauté anglophone en fouille d'opinions. Plusieurs corpus de critiques de produits technologiques, de films ou de voitures ont été développés pour l'anglais à cette occasion. Nous avons défini la complexité de cette problématique selon quatre aspects :

- l'identification des différentes formes textuelles pour représenter un même objet ;
- la résolution des relations méronymiques entre objets ;
- le choix du meilleur objet candidat autour d'une évaluation ;
- la prise en compte de la proximité aléatoire entre l'évaluation et sa cible.

Quatre types d'approches ont été proposés jusqu'ici pour traiter en particulier les deux derniers points. Celles-ci s'appuient :

- sur la proximité entre une évaluation et sa cible ;
- sur les dépendances syntaxiques entre une évaluation et sa cible ;
- sur un apprentissage supervisé de combinaisons lexico-grammaticales ;
- sur une résolution des anaphores pronominales.

Néanmoins les travaux existants n'offrent que des solutions partielles et traitent principalement des cas où la cible est proche de son évaluation (intraphrastique). Elles ont été expérimentées et évaluées sur des corpus anglophones monothématiques.

Dans ce cadre, nous avons adapté au français trois méthodes s'inspirant de l'existant. Nous les avons évaluées sur un corpus multithématique francophone. Nous avons proposé une nouvelle méthode permettant de traiter les cas de couples évaluation-cible n'apparaissant pas dans la même phrase (interphrastique). Cette méthode s'appuie sur l'utilisation d'une méthode existante comme amorce et d'un calcul de saillance des objets cibles. Nos résultats montrent que la méthode état de l'art établie sur un corpus monothématique anglais est adaptée pour traiter un corpus français multithématique avec une légère baisse des résultats. Notre nouvelle méthode par calcul de saillance améliore l'état de l'art sur l'aspect intraphrastique de la tâche. Elle permet également de couvrir davantage de problèmes en s'intéressant aux cas interphrastiques non traités dans le domaine jusqu'à présent. De ce point de vue nous améliorons significativement les résultats pour identifier la cible d'un passage d'opinion. Ces résultats nous semblent encore améliorables en prenant mieux en compte les relations sémantiques entre les objets d'un document.

Conclusion

L'étude présentée porte sur l'analyse à granularité fine de la subjectivité. Elle s'inscrit dans un domaine émergent du TAL : la fouille d'opinions. Ce domaine est lui-même lié à un cadre applicatif porteur, celui de la veille d'opinions stratégiques sur internet. En effet, la diffusion massive de textes sur les blogs, les forums ou les sites de presses invite les industriels ou les personnalités publiques à avoir une vision précise des données d'opinions qui circulent et qui influent sur leur image de marque, leur notoriété ou leurs ventes. Néanmoins, face à ces enjeux, la complexité de l'expression d'une opinion dans la langue soulève des défis importants pour automatiser l'analyse.

Dans ce contexte, notre objectif principal était de proposer et d'évaluer des méthodes dites à granularité fine dans un cadre multithématique. Cet objectif s'oppose aux approches par catégorisation de documents souvent restreintes à un thème ou un genre. L'axe choisi correspond à un paradigme récent initié par Wilson (2008) pour l'anglais et peu exploré jusqu'ici. Il a pour ambition de mieux répondre aux défis posés par la fouille d'opinions en structurant les différents passages évaluatifs d'un document. Cette structuration consiste en particulier à associer des traits sémantiques à un passage évaluatif, à lever les ambiguïtés locales et à identifier son objet cible. Il s'agit également de dépasser la simple opposition *positif/négatif* ou *subjectif/objectif* des méthodes pionnières du domaine. Pour mener à bien cet objectif, notre démarche nous a amené à recadrer la notion d'opinion dans des théories linguistiques plus larges sur la subjectivité et l'évaluation, puis à développer des ressources francophones pour entraîner et évaluer des méthodes sur différentes problématiques. Dans la suite de ce chapitre, nous présentons un bilan des travaux et des problématiques traitées. Puis, nous discutons des perspectives envisagées.

8.1 Un bilan de l'analyse à granularité fine de la subjectivité

Ce travail de thèse s'est divisé en quatre parties : (i) le développement de ressources francophones décrivant la subjectivité dans la langue en nous intéressant en particulier aux évaluations (considérées comme les marques d'une subjectivité de second niveau). À partir de ces ressources, nous avons proposé et évalué des méthodes traitant de : (ii) la délimitation et la catégorisation des passages évaluatifs dans un énoncé, (iii) l'amélioration de la couverture des lexiques de mots évaluatifs, (iv) l'identification de l'objet cible d'une évaluation. Nous en présentons un bilan dans les sections suivantes.

8.1.1 Le développement de ressources francophones

L'état de l'art des travaux en fouille d'opinions a montré que l'analyse au grain document est l'axe le plus souvent traité dans la bibliométrie du domaine et aussi celui qui dispose du plus de ressources (voir chapitre 2, §. 2.1). D'un autre côté, les corpus et les lexiques destinés à l'analyse à granularité fine sont rares, en particulier pour le français. Dans ce contexte, nous avons développé manuellement deux ressources francophones : le corpus Blogoscopie constitué de blogs multithématiques et le lexique de l'évaluation. Le lexique est initialement constitué de 984 entrées, principalement des unigrammes de mots de classes grammaticales diverses, puis il a été enrichi automatiquement et contient finalement 3 974 entrées, dont environ 1 900 expressions. Le modèle d'annotation du corpus et la structuration du lexique ont été guidés par des théories linguistiques sur la subjectivité.

Un modèle théorique pour structurer les ressources Nous avons proposé une modélisation de la subjectivité, adaptée à un contexte informatique, en nous inspirant de théories linguistiques. Celles-ci définissent deux niveaux de subjectivité :

- le premier témoigne de la présence explicite ou implicite de l'énonciateur dans son énoncé ;
- le second rend compte d'une évaluation exprimée par l'énonciateur.

Ces deux niveaux de subjectivité sont perçus dans un énoncé par la présence de marqueurs spécifiques. Nous nous sommes intéressés en particulier à l'aspect évaluatif de la subjectivité. Le modèle proposé considère ainsi qu'une évaluation possède une modalité, une configuration d'énonciation, une polarité axiologique et évalue, dans certains cas, un objet cible. Ce modèle est une première ébauche qui nécessitera certainement d'être adaptée et confrontée à d'autres modèles. Certaines dimensions telles que la modalité ou l'axiologie méritent d'être affinées pour proposer davantage de catégories ou offrir une vision moins binaire de l'axiologie. En particulier, nous envisageons de retravailler la distinction sémantique entre les modalités de jugement et d'appréciation. Ce modèle semble néanmoins applicable en l'état pour traiter plusieurs genres et thématiques textuelles.

Un coût de création important mais nécessaire Le développement de ces ressources a représenté un coût humain important. Elles conditionnent néanmoins la faisabilité des travaux qui suivent. Pour le développement et la comparaison des méthodes, la création de corpus dédiés à l'analyse à granularité fine est un aspect qui nécessite d'être reproduit, soit sur d'autres corpus multithématiques, soit sur des corpus thématiques. La possibilité de développer des corpus par une amorce automatique est l'un des points discutés dans nos perspectives.

L'évaluation qualitative et quantitative des lexiques L'évaluation qualitative et quantitative des lexiques est un point critique que nous avons soulevé. Face à l'instabilité sémantique qui caractérisent les mots évaluatifs, seule une évaluation en contexte nous semble permettre d'estimer la qualité et la couverture des lexiques. Nous avons évalué en contexte le lexique de l'évaluation. Dans sa version initiale, lors de nos expérimentations sur les blogs et dans le cadre de DEFT'09, le lexique de l'évaluation provoque un bruit évalué de 7,2 à 8,9 % selon les corpus. Il permet de détecter environ 50 % des évaluations d'un corpus. La partie enrichie du lexique de l'évaluation détecte environ 28 % d'évaluations en plus. Elle provoque 10,7 à 23,9 % de bruits. Ce taux d'erreur varie selon les corpus et selon la catégorie grammaticale des entrées lexicales.

8.1.2 La délimitation et la catégorisation des évaluations

Articuler le lexique, la grammaire et la sémantique À partir des ressources développées, nous avons proposé une méthode pour délimiter et catégoriser les différentes évaluations contenues dans

un énoncé. Cette méthode s'appuie sur l'usage d'un lexique et sur une liste de structures évaluatives. Ces structures ont pour objectif de modéliser et d'articuler des constituants linguistiques abstraits afin de représenter les exemples d'évaluations rencontrés dans un corpus d'entraînement et ceux non rencontrés. Pour construire ces structures, nous avons défini cinq principaux types de constituants : le vocabulaire évaluatif, les marqueurs de personne, d'intensité ou de négation et les autres mots. Chaque constituant possède des traits lexicaux, grammaticaux et sémantiques permettant de modéliser des nuances de construction d'évaluations assez fines. Nous avons utilisé une méthode intermédiaire à base de règles pour acquérir et généraliser les structures évaluatives du français à partir du corpus d'entraînement Blogoscopie. Chacune de ces structures possède des méta-données indiquant la modalité, la configuration d'énonciation qu'elle exprime et l'action qu'elle provoque sur l'axiologie. À partir des 4 945 évaluations du corpus Blogoscopie, nous avons acquis 2 031 structures. La courbe d'apprentissage de ces structures montrent néanmoins que nous n'avons pas atteint de limite finie et qu'il existe d'autres structures évaluatives. De ce point de vue, la phase de généralisation des exemples rencontrés dans le corpus d'entraînement nous semble un point à développer et à améliorer.

Néanmoins, dans nos expérimentations, les structures évaluatives offrent un gain intéressant pour délimiter les passages évaluatifs d'un corpus multithématique en regard d'une approche purement lexicale. Ceci a pour conséquence d'améliorer la catégorisation sémantique des évaluations et d'effectuer des désambiguïisations locales (inversion, annulation ou activation de polarité axiologique). Par ailleurs, ces structures ont été appliquées à des corpus d'autres thématiques lors de la campagne d'évaluation DEFT'09. Les résultats obtenus semblent montrer qu'elles sont adaptées à des textes journalistiques et des débats parlementaires sans modification nécessaire.

Évolutions nécessaires La phase d'acquisition des structures évaluatives nous semblerait intéressante à ré-explore en intégrant :

- la dimension syntaxique de la langue pour mieux généraliser les structures évaluatives et traiter ainsi davantage d'exemples non rencontrés dans le corpus d'entraînement ;
- des informations sur le temps et le mode de conjugaison. En effet, comme l'indique Charaudeau (1992), les marques de temporalité et de mode de conjugaison sont des indices discriminants pour la détection et la catégorisation d'une évaluation. La conjugaison de verbes au conditionnel implique généralement une modalité d'opinion par exemple.

Enfin, il serait intéressant d'extraire des structures évaluatives à partir d'un plus grand nombre d'exemples d'évaluations pour mesurer s'il est possible d'atteindre une limite au delà de laquelle on ne rencontre plus de nouveaux exemples. En particulier, l'ajout d'exemples contenant des marques de négation nous semble une piste à développer pour renforcer le traitement des tournures inversives.

8.1.3 L'amélioration de la couverture des lexiques de mots évaluatifs

Pour améliorer les lexiques, nous avons proposé deux méthodes pour apprendre des subjectivèmes (mots ou expressions subjectifs) habituellement absents des ressources existantes : les syntagmes nominaux et verbaux, les stéréotypes culturels et les mots peu fréquents.

Apprentissage par hypothèses sémantiques Dans la première méthode, nous avons notamment introduit une nouvelle hypothèse qui considère que plus un mot ou un syntagme candidat est précédé d'un marqueur d'intensité et plus ce candidat s'imprègne de subjectivité. Si cet usage linguistique se répète fréquemment, il est probable que ce candidat soit un subjectivème suffisamment stable pour être ajouté au lexique. La méthode d'apprentissage supervisé qui découle de cette hypothèse permet d'acquérir du vocabulaire subjectif, sémantiquement stable, absent du lexique de l'évaluation initial (*ambitieux, inexcusable, misérabiliste*), mais aussi des expressions, des stéréotypes culturels et des

mots peu fréquents (*fumisterie, faire jeu égal, être brisé en deux*). Ces nouvelles entrées lexicales sont précieuses pour structurer le discours évaluatif là où les ressources existantes ne détectent que les mots évaluatifs les plus fréquents et les plus stables.

Dans nos expérimentations, nous avons mesuré que cet enrichissement permet de détecter 25,7 % d'évaluations en plus que le lexique de l'évaluation initial tout en conservant une précision allant de 72,0 % à 94,6 %. Nous avons remarqué que les ambiguïtés concernent en particulier les noms et les syntagmes nominaux alors que les verbes et les adjectifs obtiennent une bonne précision. Dès lors, une des perspectives principales est de désambiguïser dans leurs contextes certains noms trop ambigus (*bijou, trésor, tragédie, plaie, gouffre, etc.*) en s'appuyant sur la thématique de l'énoncé ou sur la catégorie sémantique des objets cibles évalués dans le contexte. Dans nos perspectives, nous envisageons également de reproduire la méthode d'apprentissage de subjectivèmes sur un corpus thématique ou orienté sur un objet cible pour extraire des expressions et des stéréotypes culturels propres à domaine (*le sport, le journalisme politique, la bourse, etc.*).

Apprentissage par hypothèses morphologiques Dans la seconde méthode, nous avons adopté une série d'hypothèses morphologiques introduites dans des travaux en linguistique, mais jamais évaluées jusqu'ici en informatique. Elles considèrent que certains affixes sont des marques qui révèlent la subjectivité et la polarité axiologique négative d'un mot candidat. Nous avons utilisé un dictionnaire pour collecter des candidats et une phase d'apprentissage supervisé pour retenir les candidats les plus stables et supprimer les exceptions. Les résultats montrent que ces affixes sont des indices très précis (mis à part le suffixe *-âtre*). Les subjectivèmes appris à partir de cette méthode ont une précision de 89,3 % en contexte et permettent de détecter 3 % d'évaluations en plus. Cette amélioration ne nous semble pas négligeable dans le cadre d'une analyse à granularité fine. Ces résultats invitent à s'intéresser à d'autres affixes pouvant produire des subjectivèmes négatifs (par exemple, *-isme* et *-iste*) et à adapter et évaluer des hypothèses morphologiques équivalentes sur d'autres langues.

8.1.4 L'identification de la cible d'une évaluation

Définition de la problématique La dernière problématique que nous avons abordée, toujours dans un cadre multithématique, est celle de l'identification de l'objet ciblé par une évaluation. Nous avons montré que la notion de cible, récemment introduite par des travaux anglophones, est complexe car elle met en jeu un clivage entre l'objet réel signifié et son signifiant textuel. Nous avons dès lors redéfini cette problématique selon quatre aspects :

- l'identification des différentes formes textuelles représentant un même objet (résolution d'anaphores pronominales et nominales) ;
- la résolution des relations méronymiques entre objets ;
- le choix de la meilleure cible autour d'une évaluation ;
- la prise en compte de la proximité aléatoire entre l'évaluation et sa cible.

Adaptation et amélioration de l'existant Dans notre travail, nous nous sommes en particulier intéressés aux deux derniers points de la problématique. Nous avons adapté au français trois méthodes s'inspirant de l'existant. Nous les avons évaluées sur un corpus multithématique francophone. Pour traiter les cas non résolus par l'existant, nous avons proposé une nouvelle méthode permettant de résoudre les couples évaluation-cible n'apparaissant pas dans la même phrase. Cette méthode s'appuie sur l'utilisation d'une méthode existante comme amorce et d'un calcul de saillance des objets cibles. L'hypothèse évaluée est de considérer que le discours évaluatif d'un énonciateur s'articule autour de quelques objets et qu'il est possible d'induire la cible d'une évaluation à partir des cibles présentes dans le cotexte.

Nos résultats montrent que la méthode état de l'art établie sur un corpus monothématique anglais est adaptée pour traiter un corpus français multithématique avec une légère baisse des résultats. Notre nouvelle méthode par calcul de saillance améliore l'état de l'art sur l'aspect intraphastique de la tâche. Elle permet également de couvrir davantage de problèmes en s'intéressant aux cas interphrastiques non traités dans le domaine jusqu'à présent. De ce point de vue nous améliorons les résultats existants pour identifier la cible d'une évaluation.

Améliorations possibles Ces résultats nous semblent néanmoins améliorables en prenant mieux en compte les relations sémantiques entre les objets d'un document (relation d'anaphore pronominale ou nominale, relation méronymique, etc.). Ceci constitue l'une des perspectives que nous présentons dans la section suivante. Enfin, l'identification de l'objet cible d'une évaluation est une problématique intéressante car elle offre des perspectives pour désambiguïser le vocabulaire évaluatif en fonction de la cible. Par exemple, l'expression *cordon bleue* est subjective lorsqu'elle porte sur une cible de type *personne*. Nous discutons également de ce point dans les perspectives de cette thèse.

8.2 Perspectives

Les travaux réalisés dans cette thèse nous orientent vers deux axes de perspectives. Elles visent à poursuivre (i) le développement de ressources dédiées à l'analyse à granularité fine de la subjectivité et (ii) la structuration du discours évaluatifs.

8.2.1 Développer des ressources

Pour faire émerger des méthodes d'analyse à granularité fine de la subjectivité, nous proposons plusieurs pistes de réflexions concernant le développement des corpus et des lexiques.

8.2.1.1 Corpus

Annoter des corpus avec une amorce automatique L'annotation manuelle de corpus représente un coût humain important : environ un homme-mois pour annoter 10 à 15 documents de taille standard¹ dans nos travaux et ceux de Wiebe et collab. (2006). Dès lors, une piste peu développée jusqu'ici consiste à amorcer automatiquement l'annotation des passages subjectifs ou évaluatifs d'un corpus. Il s'agit d'avoir recours à des annotateurs humains uniquement lors de phases de corrections et de renforcements². Néanmoins, la correction est une opération coûteuse car elle nécessite de vérifier chaque annotation posée par le système automatique. Il est donc nécessaire de minimiser le nombre de corrections. Dès lors, la précision de l'amorce automatique est un facteur clé alors qu'une couverture plus faible a moins d'effets négatifs.

Dans ce contexte, la méthode de délimitation et de catégorisation des évaluations que nous avons présentée dans ce manuscrit possède une précision intéressante (environ 90 %) qu'il est possible de ré-exploiter. Le lexique de l'évaluation initial, voire dans sa version enrichie, fournit de bons indices et permet de couvrir de 50 à 65 % des évaluations. Avec notre méthode, l'annotation automatique du corpus Blogoscopie (250 documents) prend environ 20 minutes avec une machine standard. Il est donc possible d'annoter de très gros volume de textes rapidement : environ 18 000 documents de taille standard en une journée. Pour comparaison, l'ensemble de la blogosphère francophone

1. de 500 à 1 000 mots.

2. Annoter les passages oubliés par le système automatique.

produit à l'heure actuelle environ 50 000 documents quotidiennement³. Il est donc possible d'analyser l'ensemble de la production textuelle des blogs francophones avec peu de ressources matérielles.

Dans le domaine, cette idée rejoint les expérimentations menées par Riloff et collab. (2003) pour annoter des phrases subjectives en anglais. Ils ont utilisé un classifieur à base de règles et un lexique d'indices forts pour conserver une haute précision. Le corpus obtenu a servi à entraîner un nouveau classifieur et à extraire de nouveaux indices. Ces expérimentations méritent d'être reproduites dans un cadre intraphrastique, multithématique et pour le français. L'annotation de corpus par amorce automatique est une perspective également envisagée par les organisateurs de DEFT'09 qui proposent de distribuer les corpus annotés par les différents systèmes des participants. L'observation des désaccords entre les systèmes est une première étape pour minimiser le coût des corrections humaines.

Annoter des corpus par domaine thématique et par objet cible Les particularités du discours évaluatif dans certains domaines thématiques (*médical, e-learning, cuisine, etc.*) ou sur certains objets cibles (*l'écologie, un sportif, une élection, etc.*) invitent à spécialiser les ressources pour mieux prendre en compte les évaluations implicites ou les stéréotypes culturels associés à un domaine ou à un objet. De ce point de vue, la création de corpus par thématique ou ciblé sur un objet est une perspective importante. Pour en minimiser le coût, il s'agit d'annoter manuellement uniquement le vocabulaire ou les stéréotypes spécifiques à la thématique ou à l'objet visé. Par exemple, *bloquer* est un stéréotype négatif dans le domaine du *e-learning* (1), alors qu'il est positif dans le sport (2). *engranger les points* et *prendre la tête* sont des stéréotypes positifs gravitant autour des objets de type *sportif* ou *équipe sportive* (3), mais peuvent être neutre ou négatif dans un autre contexte (4).

- (1) Je **bloque** ! Pouvez-vous m'aider à dépasser ce problème ?
- (2) Sur l'attaque qui suit, Turiaf **bloque le tir** de Kotchenko
- (3) Tourcoing **engrange les points** et **prend la tête**.
- (4) on **prend la tête** de tout le monde avec le réchauffement climatique

Ces corpus permettraient d'aborder la problématique de désambiguïisation des subjectivèmes que nous précisons dans le paragraphe suivant.

8.2.1.2 Lexique

Nous envisageons deux perspectives pour améliorer les lexiques du domaine : (i) proposer des méthodes pour désambiguïser les subjectivèmes selon la thématique ou la cible, (ii) évaluer en contexte et réunir les lexiques existants.

Désambiguïser les subjectivèmes selon la thématique ou selon la cible La perspective de désambiguïser les subjectivèmes nécessite de disposer de corpus d'entraînements thématiques ou orientés sur une cible où les couples cible-évaluation sont annotées. De tels corpus peuvent être développés par amorce automatique, puis corrigés manuellement. Dans cette perspective, il s'agit :

- d'identifier une liste de subjectivèmes instables issus des lexiques existants (*petit, énorme, surprenant, boulet, cordon bleu, gagner, inonder, faire basculer, bloquer, prendre la tête, etc.*) ;
- d'analyser statistiquement le comportement sémantique de ces subjectivèmes selon la thématique ou la cible ;
- de modéliser les connaissances lexicales acquises sur chacune de ces entrées.

3. Source : Wikio.

Pour une langue donnée, l'objectif consisterait par exemple à construire un lexique des subjectivèmes stables (pertinents dans un contexte multithématique), puis des sous-lexiques spécifiques à un thème ou un objet pour résoudre les cas plus instables.

Évaluer et réunir les lexiques existants Pour favoriser la perspective précédente, l'évaluation en contexte et la confrontation des lexiques existants est une étape nécessaire. Pour le français, la réunion du lexique de l'évaluation et du lexique des sentiments permettrait ainsi :

- d'augmenter le nombre de subjectivèmes considérés ;
- de prendre en compte la richesse des traits sémantiques du lexique des sentiments ;
- de considérer les cas de désaccords entre les lexiques comme des subjectivèmes instables.

L'évaluation en contexte de ces lexiques est dépendante du développement de corpus annotés de référence.

8.2.2 Structurer le discours évaluatif

Le deuxième axe de perspectives présente des réflexions sur la structuration du discours évaluatif. Dans les travaux en analyse du discours, un énoncé peut-être vu comme une organisation particulière d'unités linguistiques ayant chacune un rôle particulier dans le discours. Dans le cadre de la fouille d'opinions, la délimitation d'unités discursives et la modélisation de leurs relations constituent une perspective majeure pour mieux extraire le sens évaluatif général d'un énoncé (Asher et collab., 2008). Pour atteindre cet objectif ambitieux, nous envisageons de nous appuyer sur les travaux présentés dans cette thèse pour : (i) calculer les relations sémantiques entre les cibles évaluées, (ii) identifier les unités évaluatives du discours.

8.2.2.1 Calculer les relations sémantiques entre les cibles

Pour structurer le discours évaluatif, la première perspective consiste à résoudre les relations sémantiques entre les signifiants textuels des objets cibles. Nous illustrons cette problématique à partir de l'exemple (5). Dans cet exemple, les évaluations positives sont en bleu, les évaluations négatives sont en rouge. Les deux objets cibles principaux de l'énoncé et leurs objets associés sont en violet (*Fadela Amara*) et jaune (*les hommes politiques*).

- (5) Ce livre, *j'avais hâte* de lire car outre le fait de dévoiler la réalité sur *Fadela Amara*, il m'a permis de découvrir *une partie du combat de ma chouchou !!* [...] *une des personnalités préférées des Français*, [...] et qui *n'a pas froid aux yeux*. Le parcours de *Fadela Amara* *tient* à la fois *du conte de fées* et de *la construction médiatique et politique*. [...]
- Je me disais : "Tiens, c'est *vraiment un pas en avant* de mettre dans le gouvernement *une femme* [...], issue des banlieues donc qui *sait de quoi elle parle*. [...] qui n'a pas été dans *le monde politique*, qui *aura donc une autre vision*". *Quelle déception !!* [...] *Je n'avais pas vraiment une bonne image* des *hommes politiques* en général [...] qui me donnaient plus l'impression d'être là pour *leur propre image personnelle* et *leurs ambitions sur-développées*. *Une belle cause, un beau combat détournés pour les ambitions* de *certains*. Que de *mensonges*, que de *manipulation*... Comment, au nom du *pouvoir*, peut-on ainsi *briser l'élan* de *personnes* qui *se battent avec leurs tripes* pour *soutenir des gens qui souffrent* [...]. *C'est à vomir !*

Texte extrait de la plateforme Over-blog - 09/09/2009

Dans le contexte de la fouille d'opinions, nous proposons de prendre en compte trois catégories principales de relations :

- « *X est équivalent à Y* » : les signifiants textuels X et Y renvoient au même objet réel. Ils sont reliés par :
 - une anaphore pronominale (*Fadela Amaral elle*) ;
 - ou une anaphore nominale (*Fadela Amaral ma chouchou*).
- « *X est associé à Y* » : X fait référence à un objet réel associé à l'objet décrit par Y. X et Y ont soit :
 - une relation de méronyme/holonyme (*Fadela Amaral son combat*) ;
 - un holonyme commun (*le combat* et *le livre* sont des méronymes de *F. Amara*) ;
 - ou ont été associés de façon pragmatique (choix de l'énonciateur ou actualité) (*Fadela Amara et la construction médiatique*).
- « *X s'oppose à Y* » : les objets signifiés par X et Y s'opposent à un niveau sémantique ou pragmatique dans le discours de l'énonciateur. Par exemple, dans (6), *Fadela Amara femme* s'oppose à *hommes politiques*. Il importe de remarquer l'opposition entre ces deux objets pour ne pas confondre les évaluations qui portent sur l'un ou sur l'autre objet.

La résolution de ces relations permettrait de construire un graphe des objets de l'énoncé où chaque signifiant textuel d'objet est un noeud relié à un autre noeud par une relation d'équivalence, d'association ou d'opposition. Bien que cette problématique soit complexe, il serait intéressant d'expérimenter des méthodes utilisant soit des ressources externes pour décrire un domaine ou soit des techniques endogènes pour mesurer automatiquement le degré d'association entre deux objets.

8.2.2.2 Identifier les unités évaluatives du discours

À partir des passages évaluatifs délimités et des graphes d'objets de l'énoncé, l'objectif final consiste à structurer le discours sous forme d'unités évaluatives homogènes. Ainsi, la figure 8.1 représente un aperçu des unités évaluatives de l'énoncé (5). Dans une première unité, l'énoncé contient des évaluations positives autour d'objets associés à *Fadela Amara*. Cette première unité peut être affinée, en délimitant les zones textuelles propres à chaque objet associé. Dans une seconde partie, l'énoncé contient surtout des évaluations négatives autour des objets faisant référence aux *hommes politiques*. Les zones ayant la même polarité axiologique et portant sur le même objet principal sont reliées entre elles.



FIGURE 8.1 : Structuration des unités évaluatives du discours

La perspective d'une telle structuration permet selon nous de ré-introduire les problématiques

de catégorisation au grain document (voir chapitre 2, §. 2.1). En effet, il s'agirait dès lors, non pas de retrouver la polarité globale du document, mais d'identifier la polarité du document par rapport à un objet cible ou un sous-ensemble d'objets cibles. Selon la modélisation 8.1, le document (5) est globalement positif en ce qui concerne *Fadela Amara* et globalement négatif sur la *politique*. D'autres types d'unités discursives sont envisageables pour représenter les ruptures de modalités et de configurations d'énonciations dans le discours. En ce sens, la structuration du discours évaluatif d'un document constitue une passerelle importante pour unifier les problématiques du domaine à grain large et à grain fin.

Document non annoté du corpus

<page mes_blog_rank="62" mes_mediametrie="" tags_blog="actualités, dupont aignan, france, realite, republique" thematique="Sarah Palin" url="http://les3r.over-blog.com/">

<billet age="" auteur="" id_b="B1067282552" profession="" url="http://les3r.over-blog.com/article-24217300.html">

<date>2008-10-29 15 :33 :00**</date>**

<titre>Sarah Palin et ses deux neurones redécorent le bureau ovale**</titre>**

<texte> Même si personne ne veut y croire, le gouverneur de l'Alaska pourrait bien devenir la première femme de l'histoire à présider aux destinées du «monde libre». Un site internet propose de découvrir le bureau ovale version Sara Palin. Dopé à la mauvaise foi et à l'humour, bien sûr... Et si Barack Obama, poussé dans le dos par deux apprentis néo-nazis, trébuchait dans la dernière ligne droite ? Et si John McCain, tout juste arrivé au pouvoir, passait l'arme gauche (sic) ? Et si, finalement, l'inénarrable Sarah Palin devenait la nouvelle locataire de la Maison Blanche, à quoi pourrait bien ressembler son mandat ? Effectivement ça fait beaucoup de «si» mais, grâce au site Palinaspresident.us, on a un début de réponse à ces questions que personne n'ose se poser. Avec lui, on découvre le gouverneur de l'Alaska, tout sourire, trônant au milieu du bureau ovale. Et l'ex-Miss Wassilla 1984 a de quoi être ravie : elle ne devient pas seulement la première femme à présider le «monde libre», elle jouit aussi d'un espace de travail qu'elle a entièrement redécoré selon ses goûts. En cliquant au hasard dans le bureau ovale, on peut découvrir son livre préféré (Politics for dummies, La politique pour les nuls), sa fibre écolo (il est possible de dégommer un jeune faon ou un innocent lapin blanc) ou bien encore sa maîtrise parfaite de la géographie et de la géopolitique («Corée du Nordistan», «Iranistan», «Quelquepartdistan»). Aucun détail n'a été oublié, même Joe le plombier et le téléphone rouge sont là. Bref, ce site ne sert à rien, si ce n'est à s'occuper durant ses heures de boulot. Ce qui n'est déjà pas si mal ?

Mercredi 29 Octobre 2008 - 15 :29 Gérald Andrieu www.marianne2.fr

</texte>

</billet>

</page>

Document annoté du corpus

```

<page mes_blog_rank="50" mes_mediametrie="" tags_blog="" thematique="Harry Potter"
url="http://mars-studio.over-blog.com/">
  <billet age="" auteur="Servane" id_b="B1020258109" profession="" url="http://mars-studio.over-
blog.com/article-13966101.html" orthographe="standard" syntaxe="correcte">
    <date>2007-11-21 13 :08 :00</date>
    <titre>Ma <Appreciation type="AIF" forme="Harry Potter">passion</Appreciation> <IA
cc="C1, C2">Harry Potter</IA></titre>
    <texte> <partie organisation="enonciatif"> Il n'y a pas d'âge pour être fan de quelque chose,
eh bien entre autre, <Appreciation type="AEF" forme="harry">je suis fan</Appreciation> de <IA
cc="C1, C2">harry</IA>, des <CC id_c="C1">romans</CC>, des <CC id_c="C2">films</CC>, de tout !
Alors je rajoute une <CC id_c="C3">peinture</CC> de <IA cc="C3, C1">CHO</IA> dans la <CA
cc="C3">galerie</CA>. Pour anecdote, j'ai eu l'impression bizarre de me peindre, quelle sensation bizarre !
Non pas que je ressemble à CHO, ce sont ses <CA cc="C3, CHO">yeux bridés</CA> qui m'<Appreciation
type="AE_forme_exclamative_D" forme="yeux bridés">ont un peu perturbés !</Appreciation> g
Pardonnez les <CA cc="C3">photos</CA>, les <CC id_c="C3">peintures</CC> sont trop grandes
pour être scannées. j'encourage tous les <CA cc="C1, C2, Harry Potter">fans</CA> à aller
visiter ce <CC id_c="C4">site</CC> ci : La <CA cc="C4, Harry Potter">Bible</CA> Harry
Potter, <Appreciation type="AE_forme_exclamative_F" forme="site">FABULEUX !!</Appreciation>
http://www.encyclopedie-hp.org/ Le site ci dessous est un site ou j'y ai trouvé <Appreciation type="AIF"
forme="fan art">énormément de très beaux</Appreciation> <CA cc="C1, C2, Harry Potter">fan art</CA>
de Harry Potter, allez y faire un tour ! http://www.harrymedia.com à bientôt et <Appreciation type="AEF"
forme="com">merci</Appreciation> pour vos <CA cc="C4" niveau="familier">com</CA> ! </partie>
</texte>
  </billet>
  <comment auteur="Béatrice" id_co="CO1" orthographe="standard" syntaxe="correcte"
  <date>2007-11-22 08 :55 :10</date>
  <texteco> <partie organisation="enonciatif">Non seulement <Appreciation type="AEF"
forme="femme écrivain">j'aime bien</Appreciation> la <CA cc="C2, C3">femme écrivain</CA>
et son univers, j'ai lu tous les Harry Potter et vu les flms (<Appreciation type="AIF" forme="Harry
Potter">nous sommes des incondtionnels</Appreciation> at home) mais en plus <Appreciation
type="AEF" forme="femme écrivain">je lui très reconnaissante</Appreciation> car elle a permis à mon
fils Loïc de s'intéresser aux livres, à la lecture... depuis la sortie du tout 1er livre, il a trouvé du plaisir
à lire...avant c'était maman qui lisait les histoires le soir, maintenant, de temps en temps c'est lui qui
vient s'installer à mes côtés et qui me fait la lecture... Alors <Appreciation type="AEF" forme="Mme
Rowling">Merci</Appreciation> <IA cc="femme écrivain">Mme Rowling</IA></partie> </texteco>
  </comment>
</page>

```


Extrait du lexique de l'évaluation

```
<lexical_entry id = "ADJ_abominable" ambiguity="false">
  <morpho>
    <name>abominable</name>
    <category>ADJ</category>
  </morpho>
  <evaluation type = "Appreciation" subtype="defavorable">
    <occurrence>      <annotation file="religions (2).xml" subject="assassinats" target="assassinats">
<form config="implicite">abominable</form></annotation>      </occurrence>
  </evaluation>
  <attestation origin="corpus_annotate"/>
</lexical_entry>
<lexical_entry id = "ADJ_absurde" ambiguity="false">
  <morpho>
    <name>absurde</name>
    <category>ADJ</category>
  </morpho>
  <evaluation type = "Appreciation" subtype="defavorable">
    <occurrence>      <annotation file="Dev_durable-3254146-1020830306.xml" subject="compétition,
première ville">      <form config="implicite">cela est absurde</form></annotation>      </occurrence>
  </evaluation>
  <attestation origin="corpus_annotate"/>
</lexical_entry>
<lexical_entry id = "NOM_absurdité" ambiguity="false">
  <morpho>
    <name>absurdité</name>
    <category>NOM</category>
  </morpho>
  <evaluation type = "Appreciation" subtype="defavorable">
    <occurrence>      <annotation file="LRU-2073242-1020179930.xml" subject="système"
target="universités">      <form config="implicite">absurdité rare</form></annotation>      </occurrence>
  </evaluation>
  <attestation origin="corpus_annotate"/>
</lexical_entry>
```


Bibliographie

- 1995, *Proceedings of the 6th Message Understanding Conference (MUC-6)*, Morgan Kaufmann.
- Abney, S. 1991, « Parsing by chunks », dans *Principle-based parsing : Computation and Psycholinguistics*, Kluwer Academic Publishers, p. 257–278.
- Adamic, L. A. et N. Glance. 2005, « The political blogosphere and the 2004 u.s. election : divided they blog », dans *Proceedings of the 3rd international workshop on Link discovery*, LinkKDD '05, ACM, New York, NY, USA, p. 36–43.
- Andreevskaia, A. et S. Bergler. 2006, « Mining wordnet for fuzzy sentiment : Sentiment tag extraction from wordnet glosses », .
- Anscombre, J.-C. et O. Ducrot. 1976, « L'argumentation dans la langue », *Langages* 42, p. 5–27.
- Asher, N., F. Benamara et Y. Y. Mathieu. 2008, « Distilling opinion in discourse : A preliminary study », dans *Proceedings of the International Conference on Computational Linguistics (Posters)*, édité par D. Scott et H. Uszkoreit, p. 7–10.
- Attardi, G. et M. Simi. 2006, « Blog mining through opinionated words », dans *Text Retrieval Conference*.
- Aue, A. et M. Gamon. 2005, « Customizing sentiment classifiers to new domains a case study », dans *Proceedings of Recent Advances in Natural Language Processing (RANLP'05)*.
- Baccianella, S., A. Esuli et F. Sebastiani. 2010, « Sentiwordnet 3.0 : An enhanced lexical resource for sentiment analysis and opinion mining », dans *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, édité par N. C. C. Chair), K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner et D. Tapias, European Language Resources Association (ELRA), Valletta, Malta.
- Bally, C. 1932, *Linguistique générale et linguistique française*, Francke.
- Banea, C., R. Mihalcea et J. Wiebe. 2008, « A bootstrapping method for building subjectivity lexicons for languages with scarce resources », dans *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, European Language Resources Association (ELRA), Marrakech, Morocco.
- Banfield, A. 1982, *Unspeakable sentences Narration and representation in the language of fiction*, Routledge and Kegan Paul.
- Benveniste, E. 1966, *Problèmes de linguistique générale II*, Gallimard éd..
- Bestgen, Y. 2002, « Détermination de la valence affective de termes dans de grands corpus de textes », dans *Actes du Colloque International sur la Fouille de Texte*, p. 81–94.
- Bestgen, Y. et G. Lories. 2009, « Un niveau de base pour la tâche 1 (corpus français et anglais) de deft'09 », dans *Actes de l'atelier de clôture du 5^e Défi Fouille de Textes (DEFT'09)*, Paris, France, p. 63–74.

- Bethard, S., H. Yu, A. Thornton, V. Hatzivassiloglou et D. Jurafsky. 2004, « Automatic extraction of opinion propositions and their holders », dans *Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text*, p. 22–24.
- Blitzer, J., M. Dredze et F. Pereira. 2007, « Biographies, Bollywood, boom-boxes and blenders domain adaptation for sentiment classification », dans *Proceedings of the Association for Computational Linguistics (ACL)*, p. 440–447.
- Bloom, K. et S. Argamon. 2010, « Unsupervised extraction of appraisal expressions », dans A. Farzindar and V. Keselj (Eds.) : *Canadian AI 2010, LNAI 6085*, p. 290–294.
- Bloom, K., N. Garg et S. Argamon. 2007, « Extracting appraisal expressions », dans *Human Language Technologies 2007 : The Conference of the North American Chapter of the Association for Computational Linguistics*, Association for Computational Linguistics, Rochester, New York, p. 308–315.
- Boser, B. E., I. M. Guyon et V. N. Vapnik. « A training algorithm for optimal margin classifiers », dans *Proceedings of the fifth annual workshop on Computational learning theory*, ACM, New York, NY, USA, p. 144–152.
- Charaudeau, P. 1992, *Grammaire du sens et de l'expression*, Hachette Education.
- Choi, Y. et C. Cardie. 2009, « Adapting a polarity lexicon using integer linear programming for domainspecific sentiment classification », dans *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, p. 590–598.
- Choi, Y., C. Cardie, E. Riloff et S. Patwardhan. 2005, « Identifying sources of opinions with conditional random fields and extraction patterns », dans *Proceedings of the Human Language Technology Conference and the Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP)*, p. 355–362.
- comScore/the Kelsey group. 2007, « Online consumer-generated reviews have significant impact on offline purchase behavior », Press Release. <http://www.comscore.com/press/release.asp?press=1928>.
- Dang, Y., Y. Zhang et H. Chen. 2010, « A lexicon-enhanced method for sentiment classification : An experiment on online product reviews », *IEEE Intelligent Systems*, vol. 25, p. 46–53.
- Das, S. R. et M. Y. Chen. 2007, « Yahoo ! for Amazon sentiment extraction from small talk on the Web », *Management Science*, vol. 53, n° 9, p. 1375–1388.
- Dave, K., S. Lawrence et D. M. Pennock. 2003, « Mining the peanut gallery opinion extraction and semantic classification of product reviews », dans *WWW '03 : Proceedings of the twelfth international conference on World Wide Web*, p. 519–528.
- Deerwester, S., S. T. Dumais, G. W. Furnas, T. K. Landauer et R. Harshman. 1990, « Indexing by latent semantic analysis », *JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE*, vol. 41, n° 6, p. 391–407.
- Denecke, K. 2009, « Are sentiwordnet scores suited for multi-domain sentiment classification ? », dans *4th International Conference on Digital Information Management, ICDIM*.
- Ding, X., B. Liu et P. S. Yu. 2008, « A holistic lexicon-based approach to opinion mining », dans *Proceedings of the Conference on Web Search and Web Data Mining (WSDM)*, p. 231–240.
- Ducrot, O. 1973, *La Preuve et le Dire*, Mame.
- Dumais, S. T., G. W. Furnas, T. K. Landauer, S. Deerwester et R. Harshman. 1988, « Using latent semantic analysis to improve access to textual information », dans *CHI '88 : Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM Press, New York, NY, USA, p. 281–285.

- Esuli, A. et F. Sebastiani. 2006, « SentiWordNet a publicly available lexical resource for opinion mining », dans *Proceedings of Language Resources and Evaluation (LREC)*.
- Feinstein, A. et D. Cicchetti. 1990, « High agreement but low kappa : I. the problems of two paradoxes », dans *J. Clin. Epidemiol.*, p. 543–548.
- Fellbaum, C., éd.. 1998, *WordNet An Electronic Lexical Database*, MIT Press.
- Ferrari, S., Y. Mathet, T. Charnois et D. Legallois. 2008, « Analyse d'opinion discours évaluatif et classification de documents », dans *actes de l'atelier FODOP'08 (fouille de données d'opinions) (INFORSID'08)*, p. 23–36.
- Ferruci, D. et A. Lally. 2004, « Uima : an architectural approach to unstructured information processing in the corporate research environment. », dans *Natural Language Engineering*, 10(3-4), p. 327–348.
- Fievet, C. et E. Turrettini, éd.. 2004, *Blog Story*, Eyrolles.
- Fleiss, J. L. 1971, « Measuring nominal scale agreement among many raters », dans *psychological Bulletin*, Vol. 76, No.5, p. 378–382.
- Forest, D., M. Bélange, D. Létourneau et A. van Hoeydonc. 2009, « Impacts de la variation du nombre de traits discriminants sur la catégorisation des documents. », dans *Actes de l'atelier de clôture du 5^eme Défi Fouille de Textes (DEFT'09)*, Paris, France, p. 75–88.
- Fourour, N. et E. Morin. 2009, « Apport du web dans la reconnaissance des entités nommées », *Revue québécoise de linguistique*, vol. 32, n° 1, p. 41–60.
- Galatanu, O. 2000, « Signification, sens, formation », dans *Education et Formation, Biennales de l'éducation, (sous la direction de Jean-Marie Barbier, d'Olga Galatanu)*, PUF, Paris.
- Grefenstette, G., Y. Qu, J. G. Shanahan et D. A. Evans. 2004, « Coupling niche browsers and affect analysis for an opinion mining application », dans *Proceedings of Recherche d'Information Assistée par Ordinateur (RIAO)*, p. 186–194.
- Grouin, C., B. Arnulphy, J.-B. Berthelin, S. E. Ayari, A. García-Fernandez, A. Grappy, M. Hurault-Plantet, P. Paroubek, I. Robba et P. Zweigenbaum. 2009, « Présentation de l'édition 2009 du défi fouille de textes (deft'09) », dans *Actes de l'atelier de clôture du 5^eme Défi Fouille de Textes (DEFT'09)*, Paris, France, p. 35–52.
- Grouin, C., J.-B. Berthelin, S. E. Ayari, T. Heitz, M. Hurault-Plantet, M. Jardino, Z. Khalis et M. Lastes. 2007, « Présentation de l'édition 2007 du défi fouille de textes (deft'07) », dans *Actes de l'atelier de clôture du 3^eme Défi Fouille de Textes (DEFT'07)*, Grenoble, France, p. 1–8.
- Généreux, M. et T. Poibeau. 2009, « Approche mixte utilisant des outils et ressources pour l'anglais pour l'identification de fragments textuels subjectifs français », dans *Actes de l'atelier de clôture du 5^eme Défi Fouille de Textes (DEFT'09)*, Paris, France, p. 113–122.
- Généreux, M. et M. Santini. 2007, « Défi : Classification de textes français subjectifs », *Actes de l'atelier du 3^eme Défi Fouille de Textes - AFIA 2007*, p. 83–92.
- Harb, A., G. Dray, M. Plantié, P. Poncelet, M. Roche et F. Troussel. 2008, « Détection d'opinion : Apprenons les bons adjectifs ! », dans *actes de l'atelier FODOP'08 (fouille de données d'opinions) (INFORSID'08)*, p. 59–66.
- Hatzivassiloglou, V. et K. McKeown. 1997, « Predicting the semantic orientation of adjectives », dans *Proceedings of the Joint ACL/EACL Conference*, p. 174–181.
- Hatzivassiloglou, V. et J. Wiebe. 2000, « Effects of adjective orientation and gradability on sentence subjectivity », dans *Proceedings of the International Conference on Computational Linguistics (COLING)*.

- Hu, M. et B. Liu. 2004, « Mining and summarizing customer reviews », dans *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, p. 168–177.
- Jackiewicz, A. 2006, « Relations intersubjectives dans les discours rapportés », dans *Revue Internationale sur le Traitement Automatique des Langues, Volume 47/2 : Discours et document : traitements automatiques*, p. 65–87.
- Jakob, N. et I. Gurevych. 2010, « Using anaphora resolution to improve opinion target identification in movie reviews », dans *Proceedings of the ACL 2010 Conference Short Papers*, Association for Computational Linguistics, Stroudsburg, PA, USA, p. 263–268.
- Joachims, T. 1997, « Text categorization with support vector machines learning with many relevant features », Rapport Interne Ls8-Report 23, Universität Dortmund.
- Joachims, T. 2002, « Optimizing search engines using clickthrough data », dans *Proceedings of the eighth ACM SIGKDD on Knowledge discovery and data mining*, ACM, New York, NY, USA, p. 133–142.
- Kaji, N. et M. Kitsuregawa. 2007, « Building lexicon for sentiment analysis from massive collection of HTML documents », dans *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, p. 1075–1083.
- Kamps, J., M. Marx, R. J. Mokken et M. de Rijke. 2004, « Using WordNet to measure semantic orientation of adjectives », dans *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC'04)*, p. 1115–1118.
- Kanayama, H. et T. Nasukawa. 2006, « Fully automatic lexicon expansion for domain-oriented sentiment analysis », dans *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Association for Computational Linguistics, Sydney, Australia, p. 355–363.
- Karttunen, L. 1984, « Features and values », dans *Proceedings of the 10th international conference on Computational linguistics, COLING '84*, Association for Computational Linguistics, Stroudsburg, PA, USA, p. 28–33.
- Kerbrat-Orecchioni, C. 1980, *L'Énonciation, de la subjectivité dans le langage*, Armand Colin (réédition 2009).
- Kessler, J. S. et N. Nicolov. 2009, « Targeting sentiment expressions through supervised ranking of linguistic configurations », dans *3rd Int'l AAI Conference on Weblogs and Social Media (ICWSM 2009)*.
- Kim, M., B.-T. Zhang et J.-S. Lee. 2010, « Subjective document classification using network analysis », *Social Network Analysis and Mining, International Conference on Advances in*, vol. 0, p. 365–369.
- Kim, S.-M. et E. Hovy. 2004, « Determining the sentiment of opinions », dans *Proceedings of the 20th international conference on Computational Linguistics (COLING'04)*.
- Kim, S.-M. et E. Hovy. 2005a, « Automatic detection of opinion bearing words and sentences », dans *Companion Volume to the Proceedings of IJCNLP-05, the Second International Joint Conference on Natural Language Processing*, p. 61–66.
- Kim, S.-M. et E. Hovy. 2005b, « Identifying opinion holders for question answering in opinion texts », dans *Proceedings of the AAI Workshop on Question Answering in Restricted Domains*.
- Kim, S.-M. et E. Hovy. 2006a, « Automatic identification of pro and con reasons in online reviews », dans *Proceedings of the COLING/ACL Main Conference Poster Sessions*, p. 483–490.

- Kim, S.-M. et E. Hovy. 2006b, « Extracting opinions, opinion holders, and topics expressed in online news media text », dans *Proceedings of the Workshop on Sentiment and Subjectivity in Text, SST '06*, Association for Computational Linguistics, Stroudsburg, PA, USA, ISBN 1-932432-75-2, p. 1–8.
- Kobayashi, N., K. Inui et Y. Matsumoto. 2007, « Opinion mining from web documents : Extraction and structurization », p. 326–337.
- Ku, L., I. Liu, C. Lee, K. Chen et H. Chen. 2008, « Sentence-level opinion analysis by copeopi in ntcir-7 », dans *Proceedings of NTCIR-7 Workshop*, Tokyo, Japan, p. 46–51.
- Laver, M., K. Benoit et J. Garry. 2003, « Extracting policy positions from political texts using words as data », *American Political Science Review*, vol. 97, n° 2, p. 311–331.
- Lee Whorf, B. 1969, *Linguistique et anthropologie*, Denoël-Gonthier.
- Legallois, D. 2005, « Pour une définition énonciative de l'enclosure vrai », dans *Les marqueurs linguistiques de la présence de l'auteur*, David Banks, L'Harmattan.
- Lin, W.-H., T. Wilson, J. Wiebe et A. Hauptmann. 2006, « Which side are you on ? identifying perspectives at the document and sentence levels », dans *Proceedings of the 2006 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology : companion volume : doctoral consortium*, p. 227–230.
- Liu, B. 2010, « Sentiment analysis and subjectivity », dans *Handbook of Natural Language Processing, Second Edition*, édité par N. Indurkha et F. J. Damerau, CRC Press, Taylor and Francis Group, Boca Raton, FL.
- Macdonald, C., I. Ounis et I. Soboroff. 2007, « Overview of the trec 2007 blog track », dans *Proceedings of The Sixteenth Text REtrieval Conference (TREC 2007)*, NIST Special Publication : SP 500-274, p. 1–9.
- Maingueneau, D. 1991, « L'analyse du discours », .
- Maingueneau, D. 1994, *Précis de Grammaire pour les Concours*, dunod éd..
- Martin, L. W. et G. Vanberg. 2008, « A robust transformation procedure for interpreting political text », *Political Analysis*, vol. 16, n° 1, p. 93–100.
- Mathieu, Y. 2006, « A computational semantic lexicon of french verbs of emotion », dans *Computing Attitude and Affect in Text : Theory and Applications, The Information Retrieval Series*, vol. 20, édité par W. B. Croft, J. Shanahan, Y. Qu et J. Wiebe, Springer Netherlands, p. 109–124.
- Mathieu, Y. Y. 2005, « Annotation of emotions and feelings in texts », dans *Affective computing and intelligent interaction*, p. 350–357.
- Matsumoto, S., H. Takamura et M. Okumura. 2005, « Sentiment classification using word subsequences and dependency sub-trees », dans *Proceedings of PAKDD'05, the 9th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining*, p. 301–311.
- Maurel, S., P. Curtoni et L. Dini. 2007, « Classification d'opinions par méthodes symbolique, statistique et hybride », *Actes de l'atelier du 3ème Défi Fouille de Textes - AFIA 2007*, p. 109–116.
- Maurel, S., P. Curtoni et L. Dini. 2008, « L'analyse des sentiments dans les forums », dans *actes de l'atelier FODOP'08 (fouille de données d'opinions) (INFORSID'08)*, p. 9–22.
- Miller, G. A., C. Leacock, R. Teng et R. T. Bunker. 1993, « A semantic concordance », dans *HLT '93 : Proceedings of the workshop on Human Language Technology*, Association for Computational Linguistics, Morristown, NJ, USA, p. 303–308.
- Mishne, G. et N. Glance. 2006, « Predicting movie sales from blogger sentiment », dans *AAAI Symposium on Computational Approaches to Analysing Weblogs (AAAI-CAAW)*, p. 155–158.

- Morinaga, S., K. Yamanishi, K. Tateishi et T. Fukushima. 2002, « Mining product reputations on the web », dans *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, p. 341–349.
- Mullen, T. et N. Collier. 2004, « Sentiment analysis using support vector machines with diverse information sources », dans *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, p. 412–418. Poster paper.
- Mullen, T. et R. Malouf. 2006, « A preliminary investigation into sentiment analysis of informal political discourse », dans *AAAI Symposium on Computational Approaches to Analysing Weblogs (AAAI-CAAW)*, p. 159–162.
- Nicolov, N., F. Salvetti et S. Ivanova. 2008, « Sentiment analysis : Does coreference matter ? », dans *Proceedings of Convention Communication, Interaction and Social Intelligence (AISB 2008)*.
- Ounis, I., M. de Rijke, C. Macdonald, G. Mishne et I. Soboroff. 2006, « Overview of the trec 2006 blog track », dans *Proceedings of The Fifteenth Text REtrieval Conference (TREC 2006)*, NIST Special Publication : SP 500-272, p. 1–9.
- Pang, B., L. Lee et S. Vaithyanathan. 2002, « Thumbs up ? sentiment classification using machine learning techniques », dans *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, p. 79–86.
- Paroubek, P., I. Robba, A. Vilnat et C. Ayache. 2008, « Easy, evaluation of parsers of french : what are the results ? », dans *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, European Language Resources Association (ELRA), Marrakech, Morocco, ISBN 2-9517408-4-0. [Http ://www.lrec-conf.org/proceedings/lrec2008/](http://www.lrec-conf.org/proceedings/lrec2008/).
- Plantié, M., G. Dray et M. Roche. 2007, « Défi deft'07 : Comparaison d'approches pour la classification de textes d'opinions », *Actes de l'atelier du 3ème Défi Fouille de Textes - AFIA 2007*, p. 55–69.
- Popescu, A.-M. et O. Etzioni. 2005, « Extracting product features and opinions from reviews », dans *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing (HLT/EMNLP)*, p. 339–346.
- Pupier, P. 1998, « Une première systématique des évaluatifs en français », *Revue québécoise de linguistique*, vol. 26, n° 1, p. 51–78.
- Quirk, R., S. Greenbaum, G. Leech et J. Svartvik. 1985, *A comprehensive grammar of the English language*, Longman.
- Rastier, F. 2006, « Sémiotique des sites racistes », dans *Les langages du politique [en ligne]*, 80, ENS Éditions. [Http ://mots.revues.org/index17113.html](http://mots.revues.org/index17113.html).
- Riloff, E. et J. Wiebe. 2003, « Learning extraction patterns for subjective expressions », dans *Proceedings of the 2003 conference on Empirical methods in natural language processing (EMNLP)*, p. 105–112.
- Riloff, E., J. Wiebe et W. Phillips. 2005, « Exploiting subjectivity classification to improve information extraction », dans *Proceedings of the 20th national conference on Artificial intelligence - Volume 3*, p. 1106–1111.
- Riloff, E., J. Wiebe et T. Wilson. 2003, « Learning subjective nouns using extraction pattern bootstrapping », dans *Proceedings of the Conference on Natural Language Learning (CoNLL)*, p. 25–32.
- Roche, M. et V. Prince. 2007, « Acrodef : a quality measure for discriminating expansions of ambiguous acronyms », dans *CONTEXT'07 : Proceedings of the 6th international and*

interdisciplinary conference on Modeling and using context, Springer-Verlag, Berlin, Heidelberg, p. 411–424.

Ruppenhofer, J., S. Somasundaran et J. Wiebe. 2008, « Finding the sources and targets of subjective expressions », dans *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, European Language Resources Association (ELRA), Marrakech, Morocco, ISBN 2-9517408-4-0.

Saussure, F. 1916, *Cours de linguistique générale*, Payot, Paris.

Schmid, H. 1994, « Probabilistic part-of-speech tagging using decision trees », dans *Proceedings of the International Conference on New Methods in Language Processing*, p. 44–49.

Shieber, S. M. 1986, *An Introduction to Unification-Based Approaches to Grammar, CSLI Lecture Notes Series*, vol. 4, Center for the Study of Language and Information, Stanford, CA.

Shieber, S. M. 1992, *Constraint-Based Grammar Formalisms*, MIT Press.

Stone, P. J. 1966, *The General Inquirer A Computer Approach to Content Analysis*, The MIT Press.

Strapparava, C. et A. Valitutti. 2004, « Wordnet-affect an affective extension of wordnet », dans *Proceedings of LREC*, vol. 4, p. 1083–1086.

Suhamy, H. 2006, « Métaphore et dualité », dans *Bulletin de la Société de stylistique anglaise*, numéro 28, p. 9–24.

Takamura, H., T. Inui et M. Okumura. 2007, « Extracting semantic orientations of phrases from dictionary », dans *Human Language Technologies 2007 : The Conference of the North American Chapter of the Association for Computational Linguistics ; Proceedings of the Main Conference*, p. 292–299.

Thompson, G. et S. Hunston. 2000, « Evaluation : An introduction », dans *Evaluation in Text : authorial stance and the construction of discourse*, édité par S. Hunston et G. Thompson, Oxford University Press, Oxford, England, p. 1–27.

Todorov, T. 1970, « Problèmes de l'énonciation », *Langages* 17, p. 3–11.

Tong, R. M. 2001, « An operational system for detecting and tracking opinions in on-line discussion », dans *Proceedings of the SIGIR 2001 Workshop on Operational Text Classification (OTC)*.

Torres-Moreno, J. M., M. El-Bèze, F. Béchet et N. Camelin. 2007, « Comment faire pour que l'opinion forgée à la sortie des urnes soit la bonne ? application au défi deft 2007 », *Actes de l'atelier du 3ème Défi Fouille de Textes - AFIA 2007*, p. 117–132.

Turney, P. 2002, « Thumbs up or thumbs down ? semantic orientation applied to unsupervised classification of reviews », dans *Proceedings of the Association for Computational Linguistics (ACL)*, p. 417–424.

Vergne, J. et E. Giguet. 1998, « Regards théoriques sur le «tagging» », dans *Actes de Traitement Automatique des Langues Naturelles (TALN'98)*, p. 22–31.

Vernier, M. et S. Ferrari. 2007, « Tracking evaluation in discourse », dans *Proceedings of Workshop at EUROLAN 2007 on Applications of Semantics, Opinions and Sentiments (ASOS'07)*.

Vernier, M., Y. Mathet, F. Rioult, T. Charnois, S. Ferrari et D. Legallois. 2007, « Classification de textes d'opinions : une approche mixte n-grammes et sémantique », p. 93–108.

Vernier, M., L. Monceaux et B. Daille. 2009, « Def't'09 détection de la subjectivité et catégorisation de textes subjectifs par une approche mixte symbolique et statistique », dans *Actes de l'atelier de clôture du 5ème Défi Fouille de Textes (DEFT'09)*, Paris, France, p. 101–112.

- Wang, B., B. Spencer, C. X. Ling et H. Zhang. 2008, « Semi-supervised self-training for sentence subjectivity classification », dans *Proceedings of the Canadian Society for computational studies of intelligence, 21st conference on Advances in artificial intelligence*, Canadian AI'08, Springer-Verlag, Berlin, Heidelberg, p. 344–355.
- Wehrli, E. 2007, « Fips, a deep linguistic multilingual parser », dans *Proceedings of the Workshop on Deep Linguistic Processing, DeepLP '07*, Association for Computational Linguistics, Stroudsburg, PA, USA, p. 120–127.
- Weinrich, H. 1989, *Grammaire textuelle du français*, Didier/Hatier.
- Whitelaw, C., N. Garg et S. Argamon. 2005, « Using appraisal groups for sentiment analysis », dans *CIKM '05 : Proceedings of the 14th ACM international conference on Information and knowledge management*, ACM, New York, NY, USA, ISBN 1-59593-140-6, p. 625–631.
- Wiebe, J. 2000, « Learning subjective adjectives from corpora », dans *Proceedings of the 17th National Conference on Artificial Intelligence (AAAI-2000)*, p. 735–740.
- Wiebe, J. et R. Mihalcea. 2006, « Word sense and subjectivity », dans *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, p. 1065–1072.
- Wiebe, J. et T. Wilson. 2002, « Learning to disambiguate potentially subjective expressions », dans *Proceedings of the Conference on Natural Language Learning (CoNLL)*, p. 112–118.
- Wiebe, J., T. Wilson et C. Cardie. 2006, « Annotating expressions of opinions and emotions in language », *Language Resources and Evaluation*, vol. 39, n° 2-3, p. 165–210.
- Wiebe, J. M. et E. Riloff. 2005, « Creating subjective and objective sentence classifiers from unannotated texts », dans *Proceedings of the Conference on Computational Linguistics and Intelligent Text Processing (CICLing)*, n° 3406 dans *Lecture Notes in Computer Science*, p. 486–497.
- Wiebe, J. M., T. Wilson, R. Bruce, M. Bell et M. Martin. 2004, « Learning subjective language », *Computational Linguistics*, vol. 30, n° 3, p. 277–308.
- Wilson, T., J. Wiebe et P. Hoffmann. 2005, « Recognizing contextual polarity in phrase-level sentiment analysis », dans *HLT '05 Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, Morristown, NJ, USA, p. 347–354.
- Wilson, T., J. Wiebe et R. Hwa. 2006, « Recognizing strong and weak opinion clauses », *Computational Intelligence*, vol. 22, p. 73–99.
- Wilson, T. A. 2008, *Fine-grained Subjectivity and Sentiment Analysis : Recognizing the Intensity, Polarity, and Attitudes of Private States*, thèse de doctorat, University of Pittsburgh.
- Yi, J., T. Nasukawa, R. Bunescu et W. Niblack. 2003, « Sentiment analyzer extracting sentiments about a given topic using natural language processing techniques », dans *Proceedings of the IEEE International Conference on Data Mining (ICDM)*, p. 427–432.
- Yu, H. et V. Hatzivassiloglou. 2003, « Towards answering opinion questions separating facts from opinions and identifying the polarity of opinion sentences », dans *Proceedings of the 2003 conference on Empirical methods in natural language processing (EMNLP)*, p. 129–136.
- Zhuang, L., F. Jing et X.-Y. Zhu. 2006, « Movie review mining and summarization », dans *CIKM '06 : Proceedings of the 15th ACM international conference on Information and knowledge management*, ACM, New York, NY, USA, p. 43–50.

Publications

Revues internationales

- Matthieu Vernier, Laura Monceaux (2010). Enrichissement d'un lexique de termes subjectifs à partir de tests sémantiques. Dans la revue *Traitement Automatique des Langues*, volume 51 (1), p. 125–149.

Conférences internationales

- Matthieu Vernier, Laura Monceaux, Béatrice Daille (2010). Learning Subjectivity Phrases missing from Resources through a Large Set of Semantic Tests. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, p. 1335–1341, Valletta, Malta.

Ateliers internationaux

- Nicolas Hernandez, Fabien Poulard, Matthieu Vernier, Jérôme Rocheteau (2010). Building a French-speaking community around UIMA. In *Proceedings of LREC Workshop 'New Challenges for NLP Frameworks'*, p. 64, Valletta, Malta.
- Estelle Dubreil, Matthieu Vernier, Laura Monceaux, Béatrice Daille (2008). Annotating opinion - evaluation of blogs. In *Proceedings of the LREC workshop on Sentiment Analysis : Emotion, Metaphor, Ontology and Terminology (EMOT 2008)*, Marrakech, Morocco.
- Matthieu Vernier, Stéphane Ferrari (2007). Tracking evaluation in discourse. In *Proceedings of the workshop at EUROLAN 2007 on Applications of Semantics, Opinions and Sentiments (ASOS'07)*, Iasi, Romania.

Revues nationales

- Matthieu Vernier, Laura Monceaux, Béatrice Daille, Estelle Dubreil (2009). Catégorisation des évaluations dans un corpus de blogs multi-domaine. Dans la *Revue des nouvelles technologies de l'information (RNTI)*, p.45–70.

Conférences nationales

- Matthieu Vernier, Laura Monceaux, Béatrice Daille (2011). Identifier la cible d'une évaluation dans un corpus multithématique. Dans les *Actes de la 18ème Conférence sur le traitement Automatique des Langues Naturelles (TALN 2011)*, Montpellier, France.
- Charles Dejean, Manoel Fortun, Clotilde Massot, Vincent Pottier, Fabien Poulard, Matthieu Vernier (2010). Un étiqueteur de rôles grammaticaux libre pour le français intégré à Apache UIMA. Dans les *Actes de la 17ème Conférence sur le Traitement Automatique des Langues Naturelles (TALN 2010)*, Montréal, Canada.
- Matthieu Vernier, Laura Monceaux, Béatrice Daille, Estelle Dubreil (2009). Catégorisation sémantico-discursive des évaluations exprimées dans la blogosphère. Dans les *Actes de la 16ème Conférence sur le Traitement Automatique des Langues Naturelles (TALN 2009)*, Senlis, France.

- Matthieu Vernier, Stéphane Ferrari, Dominique Legallois (2007). Suivi d'opinion dans le discours. Dans les *Actes des 5èmes Journées de la linguistique de Corpus 2007 (JLC'07)*, Lorient, France.

Ateliers nationaux

- Matthieu Vernier, Laura Monceaux, Béatrice Daille (2009). DEFT'09 : détection de la subjectivité et catégorisation de textes subjectifs par une approche mixte symbolique et statistique. Dans les *Actes du 5ème Défi Fouille de Textes (DEFT'09)*, p.101–112, Paris, France.
- Matthieu Vernier, Yann Mathet, François Rioult, Thierry Charnois, Stéphane Ferrari, Dominique Legallois (2007). Classification de textes d'opinions : une approche mixte n-grammes et sémantique. Dans les *Actes du 3ème Défi Fouille de Textes (DEFT'07)*, p.93–108, Grenoble, France.

Analyse à granularité fine de la subjectivité

Avec l'essor du web social, les internautes sont davantage enclins à partager et affirmer leurs opinions. Devant l'ampleur du phénomène, les opinions exprimées sur internet sont devenues des données stratégiques qu'il faut pouvoir suivre en temps réel et caractériser finement. Notre démarche nous amène à recadrer le terme d'opinion dans les théories linguistiques sur la subjectivité et l'évaluation. Nous nous intéressons à délimiter automatiquement les passages subjectifs d'un texte et à en catégoriser plusieurs aspects sémantiques et énonciatifs : ont-ils une modalité logique ou axiologique ? quelle est leur polarité ? le locuteur dissimule-t-il sa subjectivité ? quel est l'objet ciblé ? Notre première contribution pour aborder ces questions consiste en la distribution de deux ressources : le *corpus Blogoscopie* contient 5 900 passages évaluatifs et 6 900 objets annotés. Il est le premier corpus francophone de ce type ; le *lexique de l'évaluation* représente 4 000 mots ou expressions évaluatifs du français structurés grammaticalement et sémantiquement. À partir de ces ressources, nous proposons des méthodes pour résoudre trois problématiques liées à l'analyse de la subjectivité. La première méthode combine des caractéristiques lexicales, grammaticales et sémantiques pour délimiter et désambiguïser les évaluations d'un texte. Puis, nous évaluons deux méthodes s'appuyant sur des hypothèses sémantiques et morphologiques pour apprendre du vocabulaire évaluatif habituellement absent des ressources existantes. Enfin, nous expérimentons une nouvelle méthode, fondée sur la saillance d'un objet, améliorant l'existant sur la tâche d'identification de la cible d'une évaluation.

Mots-clés : *analyse de la subjectivité, fouille d'opinions, ressources lexicales, approche à granularité fine, identification de la cible évaluée*

Fine-grained Subjectivity Analysis

With the rise of the social web, users are more likely to affirm and share their opinions. Given the magnitude of the phenomenon, opinions express on the Internet are critical data that must be followed in real time and categorised cleverly. Our approach leads us to refocus the term of opinion in the linguistic theories of subjectivity and appraisal. We are interested in detecting subjective phrases in texts and in categorizing various semantic and enunciative aspects : do they express a logical or an axiological evaluation ? what is their polarity ? does the speaker try to conceal his subjectivity ? what is the target object ? Our first contribution is the distribution of two resources : the *Blogoscopy corpus* contains about 5900 evaluations and 6900 annotated objects. It is the first francophone corpus of this type. The *evaluation lexicon* contains 4000 French words or expressions. Each lexical entry is structured grammatically and semantically. From these resources, we experiment several methods to address three issues related to subjectivity analysis. The first method combines lexical, grammatical and semantic features to delineate and to disambiguate evaluative phrases in texts. We rely on semantic and morphological assumptions to learn evaluative vocabulary usually missing from resources. Finally, we propose a new method based on the salience of an object to identify the evaluated target.

Keywords : *subjectivity analysis, opinion mining, lexical resources, fine-grained approach, opinion target identification*