



Université  
de Toulouse

# THÈSE

En vue de l'obtention du

## DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)

---

**Présentée et soutenue par :**

**Baptiste Chardon**

le mercredi 12 juin 2013

**Titre :**

Chaîne de traitement pour une approche discursive de l'analyse d'opinion

---

**École doctorale et discipline ou spécialité :**

ED MITT : Domaine STIC : Intelligence Artificielle

**Unité de recherche :**

**Directeur(s) de Thèse :**

Farah Benamara

Nicholas Asher

**Jury :**

Maite Taboada - Rapportrice - Associate Professor (Simon Fraser University, Canada)

Alexis Nasr - Rapporteur - Professeur (LIF, Université Aix-Marseille)

Farah Benamara - Directrice de thèse - Maître de conférences (IRIT, Université Paul Sabatier)

Nicholas Asher - Directeur de thèse- Directeur de recherches (IRIT, Université Paul Sabatier)

Yvette Yannick Mathieu - Examinatrice - Directeur de recherches (Université Paris 7 - CNRS)

Florence Sèdes - Examinatrice - Professeur (IRIT, Université Paul Sabatier)

Patrick Séguéla - Invité - Gérant (Synapse Développement, Toulouse)



# Remerciements

---

Une thèse est avant tout une aventure vécue en équipe. Sans le soutien d'un grand nombre de personnes, ce manuscrit n'aurait très certainement pas vu le jour.

Je souhaiterais tout d'abord remercier infiniment ma directrice de thèse, Farah Benamara, pour ses conseils, sa patience, sa disponibilité et son soutien tout au long de ces quatre années, y compris pendant les moments de doute. L'équilibre qu'elle a su trouver entre liberté dans les pistes de recherches étudiées et soutien régulier et conséquent aux moments clés ont été un réel moteur dans mes travaux. Je souhaiterais également remercier mon directeur de thèse Nicholas Asher pour sa confiance et ses conseils avisés, qui ont beaucoup apportés à mes travaux, ainsi que pour sa sympathie et son enthousiasme communicatif.

Cette thèse CIFRE n'aurait également pas pu se concrétiser sans le financement et le soutien de Synapse Développement : j'ai particulièrement apprécié ces trois années passées en tant que thésard au sein de la société. Je souhaite tout particulièrement remercier chaleureusement Patrick Séguéla, mon encadrant côté Synapse, pour sa confiance, sa disponibilité et son aide précieuse tout au long de ce travail. Ses conseils ont été particulièrement utiles dans toutes les phases de mon travail. Merci également à Dominique Laurent, gérant de Synapse Développement pendant la majeure partie de la thèse, pour m'avoir donné l'opportunité de réaliser ces travaux, ainsi que pour ses conseils de « pionnier » qui ont été très enrichissants. Un très grand merci enfin à l'ensemble des personnes avec qui j'ai travaillé à Synapse Développement : Sophie Muller, Nicolas Arbouin, Isabelle Ladas, Frédéric Rayar, Christophe Delannoy, Clément Combes, Enguerran Colson, Lolla Choucavy, Hoelintahiana Ramamonjisoa, avec qui j'ai passé de très bons moments, autant sur le plan professionnel que personnel.

Merci beaucoup également à tous les membres des équipes LILaC et MELODI de l'IRIT, avec qui j'ai pu passer de très agréables moments de travail. Je souhaiterais tout particulièrement remercier Vladimir Popescu, Philippe Muller et Stergos Afantenos, avec qui les discussions ont été extrêmement enrichissantes, et m'ont permis d'envisager de nouvelles pistes de recherches. Merci également à Anaïs Cadhillac et Emmanuel Navarro, compagnons doctorants aux thématiques proches, et à tous les doctorants de l'équipe pour leur sympathie. Enfin, merci à Nathalie Aussenac pour son accueil au sein de l'équipe MELODI.

Un grand merci également à Yvette Yannick Mathieu pour sa sympathie et ses conseils lors des réunions de travail liées au projet CASOAR, et avec qui j'ai particulièrement apprécié travailler.

Je souhaiterais aussi remercier l'ensemble des annotateurs volontaires, pour la plupart anonymes, qui se sont prêtés au jeu des questionnaires d'évaluation sur les effets des opérateurs, ainsi que les annotateurs du projet CASOAR. Les données constituées grâce à eux ont été d'une importance cruciale pour la validation des hypothèses présentées dans le présent document.

Un grand merci aux rapporteurs de cette thèse, Maite Taboada et Alexis Nasr pour avoir accepté d'examiner ces travaux, et pour leurs conseils et remarques avisés. Merci également aux membres du jury d'avoir accepté de juger ces travaux.

Une thèse est une aventure au long cours, et celle-ci n'est vraiment possible que lorsque l'on est bien entouré. Pour cela, je tiens à remercier mes amis pour les excellents moments passés ensemble durant ces quatre ans : Thibault et Andra, Lucie et Thomas, Frédéric, Aurèle, François-Henry, Benoît, et bien d'autres qui se reconnaîtront.

Je tiens également à remercier mes parents pour m'avoir soutenu et pour leur enthousiasme concernant ce projet scientifique, Sylvie pour m'avoir guidé et donné les clés du monde de l'entreprise, et toute ma famille pour leur présence et leur soutien.

Enfin, je remercie infiniment ma Lauréline, pour m'avoir épaulé tout au long de cette thèse, et pour m'avoir encouragé des premières expérimentations dans le domaine jusqu'à la soutenance, en passant bien sûr par les longs mois de rédaction.

# Résumé

---

La structure discursive d'un texte est un élément essentiel à la compréhension du contenu véhiculé par ce texte. Elle affecte, par exemple, la structure temporelle du texte, ou encore l'interprétation des expressions anaphoriques. La structure discursive a montré son utilité dans de nombreuses applications de TAL, telles que le résumé automatique, ou encore l'inférence textuelle. Dans cette thèse, nous aborderons les effets de la structure discursive sur l'analyse de sentiments.

L'analyse des sentiments est un domaine de recherche extrêmement actif en traitement automatique des langues. En effet, ces dernières années ont vu se multiplier les sources de données textuelles porteuses d'opinion disponibles sur le web : avis d'internautes, de plus en plus centralisés par les moteurs de recherche, forums, réseaux sociaux, enquêtes consommateurs effectuées par les grandes marques. Devant cette abondance de données et de sources, l'automatisation de la synthèse des multiples avis devient cruciale pour obtenir efficacement une vue d'ensemble des opinions sur un sujet donné. L'intérêt de ces données est considérable, pour les sociétés qui souhaitent obtenir un retour client sur leurs produits ou leur image de marque comme pour les personnes souhaitant se renseigner pour un achat, une sortie, ou un voyage.

La plupart des travaux actuels proposent une analyse des opinions au niveau du document ou au niveau de la phrase en ignorant la structure discursive. Dans cette thèse, nous nous plaçons dans le contexte de la théorie de la SDRT (Segmented Discourse Representation Theory) et proposons de répondre aux questions suivantes :

- Existe-t-il un lien entre la structure discursive d'un document et les opinions émises dans ce même document ?
- Quel est le rôle des relations de discours dans la détermination du caractère objectif ou subjectif d'un segment textuel ?
- Quel est le rôle des éléments linguistiques, comme la négation et la modalité, lors de la détermination de la polarité d'un segment textuel subjectif ?
- Quel est l'impact de la structure discursive lors de la détermination de l'opinion globale véhiculée dans un document ?
- Est-ce qu'une approche basée sur le discours apporte une réelle valeur ajoutée comparée à une approche classique basée sur la notion de 'sacs de mots' ?
- Cette valeur ajoutée est-elle dépendante du genre de corpus ?

Afin de répondre à ces questions, nous proposons dans un premier temps un schéma d'annotation sur deux niveaux : tout d'abord au niveau du segment de discours, où chaque segment doit être annoté selon l'opinion qu'il véhicule, et ensuite au niveau du document, où tout segment doit être relié discursivement au reste du texte. Dans le but d'étudier l'impact des contraintes de genre du texte, notre campagne d'annotation a été conduite sur deux corpus de genres différents en français : commentaires de film d'une part, et commentaires d'articles de presse d'autre part.

Nous proposons ensuite une architecture informatique permettant d'analyser les opinions dans le discours selon différents niveaux de grain :

- *Au niveau du segment* où nous proposons de sélectionner les segments discursifs porteurs d'opinion du document à l'aide de deux classifieurs parallèles. La présence de traits discursifs permet de reconnaître une part importante des segments implicites, c'est-à-dire des segments ne comportant pas d'opinion lexicalisée.
- *Au niveau des expressions d'opinion au sein d'un segment* où nous présentons une étude de l'effet linguistique des opérateurs venant altérer les caractéristiques a priori d'une expression d'opinion : négations, intensifieurs, modaux. Nous proposons ensuite un modèle de calcul, d'inspiration géométrique, permettant de déterminer finement l'impact de ces opérateurs, et ainsi de déterminer les caractéristiques de l'opinion au niveau du segment.
- *Au niveau du document* où nous comparons plusieurs approches de remontée de l'opinion depuis le niveau des segments de discours jusqu'au niveau du document, afin de calculer l'opinion globale véhiculée par celui-ci.

# Abstract

---

The discourse structure of a document is a key element to understand the content conveyed by a text. It affects, for instance, the temporal structure of a text, or the interpretation of anaphoric expressions. The discourse structure showed its usefulness in numerous NLP applications, such as automatic summary, or textual entailment. In this thesis, we will study the effects of the discourse structure on sentiment analysis.

Sentiment analysis is an extremely active research domain in natural language processing. The last years have seen the multiplication of the available textual data conveying opinion on the web: consumer reviews, forums, social networks. With this abundance in data and sources, the automation of the summary of opinion documents becomes crucial for who wants to keep an overview of the opinion on a given subject. A huge interest lies in these data, both for the companies who want to retrieve consumer opinion on their products or their company identity, and for the consumers willing to gather information prior to buying a product or while planning to travel.

Most of the current research efforts describe an opinion extraction at the document level or at the sentence level, ignoring the discourse structure. In this thesis work, we address opinion extraction through the discourse framework of the SDRT (*Segmented Discourse Representation Theory*), and try to answer to the following questions:

- Is there a link between the discourse structure of a document and the opinions contained in that document?
- What is the role of discourse relations in the determination of whether a textual segment is objective or subjective?
- What is the impact of the discourse structure in the determination of the overall opinion conveyed by a document?
- Does a discourse based approach really bring additional value compared to a classical “bag of words” approach?
- Does this additional value depend on corpus genre?

To answer these questions, we first present an annotation scheme on two levels: first at the discourse segment level, where each segment is annotated according to the opinion it conveyed, and then at the document level, where each segment must be attached and linked to the rest of the document. In order to study the impact of corpus genre, our annotation campaign has been realized on texts of two distinct corpus genre in French: movie reviews, and news article reactions. We then present the architecture of a system analyzing opinions in discourse at different granularity levels:

- *At the segment level*, we suggest to select the discourse segments of the documents which convey an opinion using two parallel classifiers. The presence of selected discourse

features enables us to recognize an important proportion of the implicit segments, i.e. of the opinionated segments which contain a lexicalized opinion.

- *At the level of opinion expressions inside a segment*, we present a study on the linguistic effects of each operator altering the prior characteristics of an opinion expression: negations, intensity modifiers, modality operators. We then propose a computing model, geometrically inspired, enabling to determine precisely the impact of these operators, and thus the characteristics of the opinion at the segment level.
- *At the document level*, we compare several approaches to determine the overall opinion of a text, given its discourse structure and the opinion conveyed by each of its segments.



# Table des matières

---

<b>TABLE DES MATIERES</b>	<b>8</b>
<b>INTRODUCTION</b>	<b>20</b>
L'ANALYSE DES SENTIMENTS, UN DOMAINE APPLICATIF EN PHASE D'INDUSTRIALISATION	20
LA STRUCTURE DISCURSIVE, INFORMATION CLEF POUR UNE BONNE COMPREHENSION D'UN TEXTE	23
CHAINE DE TRAITEMENT POUR UNE APPROCHE DISCURSIVE DE L'ANALYSE D'OPINION	24
<b>1. CONTEXTE SCIENTIFIQUE ET MOTIVATIONS</b>	<b>28</b>
1.1. OPINION ET EVALUATIONS	28
1.2. ANALYSE DE L'OPINION : UN PROBLEME A DIFFERENTS NIVEAUX DE GRANULARITE	30
1.2.1. L'OPINION AU NIVEAU DU DOCUMENT : EXTRACTION D'UN SENTIMENT GLOBAL SUR UN SUJET PRINCIPAL	30
1.2.2. L'OPINION AU NIVEAU LOCAL : EXTRACTION D'UNE OPINION UNITAIRE VEHICULEE PAR UNE PORTION DE TEXTE	36
1.2.3. LA STRUCTURE DISCURSIVE, INFORMATION CLEF POUR UNE EXTRACTION FINE AU NIVEAU DU DOCUMENT	39
1.3. BILAN ET POSITIONNEMENT DE NOS TRAVAUX	48
<b>2. MODELE DE L'OPINION ET ANNOTATIONS MANUELLES</b>	<b>52</b>
2.1. INTRODUCTION : MOTIVATIONS ET CONTEXTE SCIENTIFIQUE	52
2.1.1. PRINCIPALES CAMPAGNES D'ANNOTATIONS DE L'OPINION	52
2.1.2. DEROULEMENT DU CHAPITRE	54
2.2. MODELE LOCAL DE L'OPINION	54
2.2.1. OPINION LOCALE : UNE DEFINITION	54
2.2.2. L'EXPRESSION D'OPINION ET SES CARACTERISTIQUES	55
2.2.3. SUJET DE L'OPINION	59
2.2.4. ÉMETTEUR DE L'OPINION	61
2.2.5. OPERATEURS PORTANT SUR L'OPINION	62
2.2.6. SCHEMA RECAPITULATIF DU MODELE D'ANNOTATION AU NIVEAU LOCAL	64
2.3. L'OPINION DANS SON CONTEXTE : NIVEAU DU SEGMENT DISCURSIF	64
2.3.1. SEGMENTS DISCURSIFS ET OPINION	65
2.3.2. STRUCTURE DISCURSIVE GLOBALE	67
2.4. EXEMPLE D'ANNOTATION	72
2.4.1. OUTIL D'ANNOTATION	73
2.4.2. ANNOTATION DU DOCUMENT	74
2.5. EN PRATIQUE : PROCESSUS D'ANNOTATION MANUELLE	77

2.5.1.	DEROULEMENT PRATIQUE	77
2.5.2.	STATISTIQUES D'ANNOTATIONS	82
<b>2.6.</b>	<b>CONCLUSION</b>	<b>94</b>
<b><u>3. ARCHITECTURE DU SYSTEME : L'ANALYSE FINE DE L'OPINION LOCALE COMME SUPPORT DE REMONTEE DE L'OPINION GLOBALE</u></b>		<b>96</b>
<b>3.1.</b>	<b>INTRODUCTION</b>	<b>96</b>
3.1.1.	CONTEXTE SCIENTIFIQUE	96
3.1.2.	POSITIONNEMENT ET SPECIFICITES DE NOS TRAVAUX	97
3.1.3.	OBJECTIFS ET DEROULEMENT DU CHAPITRE	97
<b>3.2.</b>	<b>ARCHITECTURE GLOBALE DU SYSTEME</b>	<b>97</b>
3.2.1.	LEXIQUES	98
3.2.2.	MODULE DE SEGMENTATION DISCURSIVE	99
3.2.3.	DETECTION DE LA SUBJECTIVITE	102
3.2.4.	REPERAGE LOCAL	102
3.2.5.	MODULE DE CALCUL DE L'OPINION AU NIVEAU DU SEGMENT	103
3.2.6.	FUSION AU NIVEAU DU DOCUMENT : CALCUL D'UNE OPINION GLOBALE	104
3.2.7.	REMARQUES SUR LA DETECTION DU SUJET DE L'OPINION	104
<b>3.3.</b>	<b>ENVIRONNEMENT TECHNIQUE : VERS UN LANGAGE DE REGLES COMPATIBLE AVEC UNE CHAINE INDUSTRIALISEE</b>	<b>104</b>
3.3.1.	PLATE-FORME D'EXECUTION	105
3.3.2.	LANGAGE DE REGLES	109
3.3.3.	CHOIX FINAL D'ARCHITECTURE	115
<b>3.4.</b>	<b>CONCLUSION</b>	<b>116</b>
<b><u>4. GENERATION AUTOMATIQUE D'UN LEXIQUE D'OPINION</u></b>		<b>120</b>
<b>4.1.</b>	<b>INTRODUCTION</b>	<b>120</b>
<b>4.2.</b>	<b>CONTEXTE SCIENTIFIQUE</b>	<b>120</b>
<b>4.3.</b>	<b>GENERATION D'UN LEXIQUE DE POLARITE VIA UNE TAXINOMIE GENERALE DU FRANÇAIS</b>	<b>122</b>
4.3.1.	RESSOURCE UTILISEE	122
4.3.2.	ALGORITHME PROPOSE	123
4.3.3.	ÉVALUATIONS	127
4.3.4.	CONCLUSION	130
<b>4.4.</b>	<b>GENERATION D'UN LEXIQUE DE POLARITE VIA DICTIONNAIRE DE SYNONYMES</b>	<b>130</b>
4.4.1.	RESSOURCE UTILISEE	130
4.4.2.	ALGORITHME D'EXTENSION	131
4.4.3.	ÉVALUATIONS	132
<b>4.5.</b>	<b>GENERALISATION DE CES METHODES A D'AUTRES TRAITS SEMANTIQUES D'OPINION</b>	<b>133</b>
4.5.1.	INTENSITE DES EXPRESSIONS D'OPINION	133
4.5.2.	CATEGORIES SEMANTIQUES FINES	135
4.5.3.	CONCLUSION	135
<b>4.6.</b>	<b>LEXIQUE CHOISI POUR LA CHAINE DE TRAITEMENT : LEXIQUE CASOAR</b>	<b>136</b>

<b>4.7. CONCLUSION</b>	<b>139</b>
<b>5. DETECTION DE LA SUBJECTIVITE</b>	<b>142</b>
<b>5.1. INTRODUCTION</b>	<b>142</b>
5.1.1. CONTEXTE SCIENTIFIQUE	142
5.1.2. OBJECTIFS	143
<b>5.2. CATEGORIES DE SUBJECTIVITE</b>	<b>143</b>
5.2.1. ALLER AU-DELA D'UNE SIMPLE CLASSIFICATION BINAIRE	143
<b>5.3. CLASSIFICATIONS AUTOMATIQUES</b>	<b>145</b>
5.3.1. PARTITIONS BINAIRES DE L'ENSEMBLE DES SEGMENTS	145
5.3.2. MISE EN PLACE DU CLASSIFIEUR	146
<b>5.4. RESULTATS</b>	<b>153</b>
5.4.1. DONNEES ET STATISTIQUES D'ANNOTATIONS	153
5.4.2. ANNOTATION AUTOMATIQUE	154
<b>5.5. CONCLUSION ET PERSPECTIVES</b>	<b>157</b>
<b>6. ANALYSE DE POLARITE AU NIVEAU DU SEGMENT</b>	<b>160</b>
<b>6.1. INTRODUCTION</b>	<b>160</b>
6.1.1. OBJECTIFS DU CHAPITRE	160
<b>6.2. ANALYSE LINGUISTIQUE DES EFFETS DES OPERATEURS</b>	<b>161</b>
6.2.1. CLASSIFICATIONS RETENUES	161
6.2.2. EFFETS QUALITATIFS DE CHAQUE CATEGORIE D'OPERATEUR	163
6.2.1. EVALUATION DES HYPOTHESES LINGUISTIQUES	168
<b>6.3. DE L'EXPRESSION D'OPINION AU SEGMENT SUBJECTIF</b>	<b>173</b>
6.3.1. LIEN ENTRE EXPRESSION D'OPINION ET OPERATEUR : PORTEE D'UN OPERATEUR	173
6.3.2. RAPPEL : ATTRIBUTS DE L'EXPRESSION D'OPINION EN CONTEXTE, ET DE L'OPINION AU NIVEAU DU SEGMENT	176
6.3.3. GENERALISATION : REMONTER L'OPINION AU SEGMENT	177
6.3.4. OPERATIONS ELEMENTAIRES SUR L'OPINION	179
6.3.5. DE L'OPERATEUR A L'OPERATION ELEMENTAIRE	179
<b>6.4. MODELE PARABOLIQUE POUR LE CALCUL LOCAL</b>	<b>180</b>
6.4.1. ADAPTATION PAR RAPPORT A UN LEXIQUE DONNE	182
6.4.2. MODELE PARABOLIQUE RESTREINT	184
6.4.3. MODELE PARABOLIQUE AVEC CERTITUDE	187
6.4.4. ÉVALUATION DU MODELE DE CALCUL	190
<b>6.5. CONCLUSION</b>	<b>194</b>
<b>7. CALCUL DE L'OPINION AU NIVEAU D'UN DOCUMENT</b>	<b>196</b>
<b>7.1. INTRODUCTION</b>	<b>196</b>
7.1.1. OBJECTIFS ET PRINCIPE DE FUSION : DEFINITION DE LA TACHE	196
7.1.2. NOTRE APPROCHE	197
<b>7.2. CAS D'ETUDE</b>	<b>198</b>

<b>7.3. STRATEGIES DE FILTRAGES</b>	<b>199</b>
7.3.1. FILTRES PAR TYPE DE SUJET DE L'OPINION	200
7.3.2. FILTRES PAR SUBJECTIVITE	200
7.3.3. FILTRES PAR DISCOURS	201
7.3.4. PROCESSUS D'APPLICATION DES FILTRES	204
<b>7.4. REMONTEE DE L'OPINION GUIDEE PAR LA STRUCTURE DISCURSIVE</b>	<b>207</b>
7.4.1. TACHE EFFECTUEE : FORMALISATION	208
7.4.2. ALGORITHME DE PARCOURS DU GRAPHE	208
7.4.3. REGLES DISCURSIVES DE REMONTEE DE L'OPINION	210
7.4.4. CONCLUSION : QUELLES VALEURS RETENIR POUR LE SEGMENT RESULTANT ?	229
7.4.5. EXEMPLES COMPLETS	230
<b>7.5. HEURISTIQUES DE CALCUL DE L'OPINION GLOBALE</b>	<b>242</b>
7.5.1. HEURISTIQUES NUMERIQUES	242
7.5.2. HEURISTIQUES MACRO-DISCURSIVES ET DISCURSIVES	245
<b>7.6. RESULTATS ET ANALYSE</b>	<b>246</b>
7.6.1. DONNEES EXPERIMENTALES	246
7.6.2. METRIQUES D'EVALUATION	248
7.6.3. RESULTATS OBSERVES	248
<b>7.7. CONCLUSION ET PERSPECTIVES</b>	<b>258</b>
7.7.1. CONCLUSION	258
7.7.2. PERSPECTIVES	258
<b>8. CONCLUSION ET PERSPECTIVES</b>	<b>262</b>
8.1. TRAVAUX EFFECTUES ET APPORTS	262
8.2. VALORISATION INDUSTRIELLE	263
8.3. PERSPECTIVES	267
<b>ANNEXES</b>	<b>280</b>
A.1. REGLES DE DECOUPAGE COMPLEMENTAIRE POUR LA SEGMENTATION DISCURSIVE	280
A.2. EXEMPLE D'UTILISATION DE MYELINE : PATRONS DE RECONNAISSANCE DES COMPARATIFS ET DE SUPERLATIFS RELATIFS	280
A.2.1. RECONNAISSANCE DE COMPARATIFS	280
A.2.2. RECONNAISSANCE DE SUPERLATIFS RELATIFS	284

# Liste des figures

---

Figure 0.1 : Vision humoristique des échelles d'évaluation des sites de commentaires web.....	20
Figure 0.2 : Interface web tweetfeel, comptant les tweets positifs et négatifs sur un sujet .....	21
Figure 0.3 : Interface web Twitter Sentiment .....	22
Figure 1.1 : Catégories principales du framework Appraisal (Martin 2005; Whitelaw, Garg, and Argamon 2005a).....	28
Figure 1.2 : Représentation RST d'un court document (exemple issu de (Mann and Thompson 1988)). .....	40
Figure 1.3 : Structure de traits associée au segment (A) (exemple issu de (Asher, Benamara, and Mathieu 2009)).....	44
Figure 1.4 : Framework Appraisal (Martin 2005), catégories étudiées par (Trnavac and Taboada 2011).....	48
Figure 2.1 : Schéma récapitulatif de la structure d'une opinion au niveau local.....	64
Figure 2.2 : Exemple de relation entrante – interdite – le segment a est relié par une relation subordonnante avec le segment b, alors que celui-ci est inclut dans le segment complexe (b,c)	72
Figure 2.3 : Interface générale de Glozz.....	73
Figure 2.4 : Annotation de l'exemple sous Glozz .....	74
Figure 2.5 : Graphe de discours correspondant à l'annotation discursive de l'exemple .....	77
Figure 2.6 : Exemple de commentaires extraits du site Allociné .....	78
Figure 2.7 : Exemple de réactions à un article du Monde.....	79
Figure 2.8 : Corrélacion entre note globale et opinions explicites/implicites pour le corpus de commentaires web.....	85
Figure 2.9 : Corrélacion entre note globale et opinions explicites / implicites pour le corpus de réactions aux articles de presse .....	86
Figure 2.10 : Corrélacion entre note globale et opinions explicites/implicites + intensités, corpus de commentaires web.....	87
Figure 2.11 : Corrélacion entre note globale et opinions explicites/implicites + intensités, corpus de réactions aux articles de presse .....	88
Figure 2.12 : Corrélacion entre note global et nombre de segments pondéré par leurs intensités - Reviews.....	89
Figure 2.13 : Corrélacion entre note global et nombre de segments pondéré par leurs intensités - LeMonde.....	90
Figure 3.1 : Architecture globale du système.....	98
Figure 4.1 : Exemple de représentation des deux sens de l'adjectif « cher ».....	123
Figure 4.2 : Extrait du graphe après initialisation .....	124
Figure 4.3 : Sous-graphe / exemple de concept non affectable directement .....	125
Figures 4.4 (gauche) & 4.5(droite) : Détail du graphe à l'issue de la première itération (à gauche), et au début de la seconde itération (à droite) .....	126
Figure 4.6 : Arbre de décision issu de la résolution du graphe 4.2 .....	126

Figure 4.7 : Entrée XML pour le terme agaçant .....	138
Figure 4.8 : Entrée XML pour le terme noble.....	138
Figure 6.1 : Taxinomie retenue pour les opérateurs s'appliquant sur l'opinion.....	162
Figure 6.2 : Histogramme des résultats pour la validation empirique des hypothèses linguistiques N1.a à N4 .....	170
Figure 6.3 : Histogramme des résultats pour la validation empirique des hypothèses linguistiques M1 à M3.2 .....	172
Figure 6.4 : Placement du point E sur la parabole .....	181
Figure 6.5 : Modèle parabolique étendu.....	182
Figure 6.6 : Application de l'opération élémentaire de négation .....	184
Figure 6.7 : Application de l'opération de négation sur l'exemple .....	185
Figure 6.8 : application de l'opération élémentaire d'intensification.....	186
Figure 6.9 : Les trois paraboles du modèle parabolique avec certitude .....	188
Figure 6.10 : Exemple d'application d'un opérateur renforçant la certitude .....	189
Figure 6.11 : Exemple d'application d'une opération de négation sur une expression de certitude altérée .....	190
Figure 7.1 : Déroulement général du processus de fusion .....	197
Figure 7.2 : Graphe discursif correspondant à l'exemple cas d'étude .....	199
Figure 7.3 : Exemple simplifié de graphe de discours.....	202
Figure 7.4 : Graphe après application du filtre keep_top .....	202
Figure 7.5 : Graphe après application du filtre keep_top_recurif.....	203
Figure 7.6 : Graphe après application du filtre keep_sub_src .....	203
Figure 7.7 : Graphe après application du filtre keep_sub_src_not_dest.....	204
Figure 7.8 : Graphe de discours de l'exemple cas d'étude avec élagage par règles, option ilots.....	205
Figure 7.9 : Graphe de discours de l'exemple cas d'étude avec élagage par règles, option objectivation.....	205
Figure 7.10 : Graphe de discours de l'exemple cas d'étude avec élagage par règles, option élagage_physique.....	205
Figure 7.11 : Règles de relaxation des filtres : la flèche s'interprète comme "se relaxe en".....	207
Figure 7.12 : Processus d'application des filtres .....	207
Figure 7.13 : Arbre discursif associé à l'exemple 1 .....	210
Figure 7.14 : Graphe discursif associé à l'exemple 2.....	211
Figure 7.15 : Graphe discursif associé à l'exemple 3.....	212
Figure 7.16 : Graphe discursif associé à l'exemple 4.....	213
Figure 7.17 : Graphe discursif associé à l'exemple 5.....	214
Figure 7.18 : Arbre discursif associé aux exemples 6 et 7.....	215
Figure 7.19 : Arbre discursif associé à l'exemple 10 .....	216
Figure 7.20 : Graphe discursif associé à l'exemple 11.....	217
Figure 7.21 : Graphe discursif associé à l'exemple 12.....	218
Figure 7.22 : Graphe discursif associé à l'exemple 14.....	219
Figure 7.23 : Graphe discursif associé à l'exemple 15.....	220
Figure 7.24 : Arbre discursif associé à l'exemple 16 .....	221
Figure 7.25 : Graphe discursif associé à l'exemple 17.....	222

Figure 7.26 : Graphe discursif associé à l'exemple 18.....	223
Figure 7.27 : Graphe discursif associé à l'exemple 19.....	224
Figure 7.28 : Graphe discursif associé à l'exemple 23.....	226
Figure 7.29 : Graphe discursif associé à l'exemple 24.....	227
Figure 7.30 : Graphe discursif associé à l'exemple 26.....	228
Figure 7.31 : Graphe discursif associé à l'exemple 27.....	228
Figure 7.32 : Graphe interne à la CDU 13.....	230
Figure 7.33 : Graphe interne à la CDU 14.....	231
Figure 7.34 : Graphe interne à la CDU 16.....	232
Figure 7.35 : Graphe interne à la CDU 15.....	232
Figure 7.36 : Graphe résultant réduit.....	233
Figure 7.37 : Graphe discursif correspondant à l'exemple .....	236
Figure 7.38 : Graphe interne à la CDU 19.....	236
Figure 7.39 : Graphe interne à la CDU 20.....	237
Figure 7.40 : Graphe interne à la CDU 21.....	237
Figure 7.41 : Graphe interne à la CDU 22.....	238
Figure 7.42 : Graphe interne à la CDU 23.....	239
Figure 7.43 : Graphe interne à la CDU 24.....	240
Figure 7.44 : Graphe réduit résultant.....	241
Figure 7.45 : Fréquence d'apparition de chaque valeur d'opinion dans l'ensemble des segments .....	244
Figure 7.46 : Histogramme des fréquences des valeurs d'opinion du document, et valeurs proches cumulées .....	245
Figure 8.1 : L'interface d'affichage de résultats pour les acteurs du prototype OpenSem .....	264
Figure 8.2 : Extrait de l'interface de l'application Marmiton, dans le cadre du projet Casoar ....	267

# Liste des tableaux

---

Tableau 1.1 : Exemple : opinions repérées et sujets associés .....	42
Tableau 1.2 : Exemple : relations entre sujets .....	42
Tableau 1.3 : Catégories des opinions détaillées par (Asher, Benamara, and Mathieu 2008) .....	43
Tableau 1.4 : Résultats obtenus par (Taboada, Voll, and Brooke 2008) .....	45
Tableau 1.5 : Poids associés aux nuclei et, par relation, aux satellites .....	46
Tableau 2.1 : Classes sémantiques d'opinion.....	58
Tableau 2.2 : Exemples illustrant les valeurs de polarité .....	58
Tableau 2.3 : Valeurs d'intensité pour une expression d'opinion.....	59
Tableau 2.4 : Sujets associés aux différentes opinions présentes dans l'exemple (14).....	61
Tableau 2.5 : Sujets associés aux différentes opinions présentes dans l'exemple (15).....	61
Tableau 2.6 : Émetteurs associés aux différentes opinions présentes dans l'exemple (20).....	62
Tableau 2.7 : Émetteurs associés aux différentes opinions présentes dans l'exemple (21).....	62
Tableau 2.8 : Opinions du segment (28.a).....	67
Tableau 2.9 : Valeurs de subjectivité et d'opinion pour chacun des segments .....	76
Tableau 2.10 : Annotation opérationnelle : statistiques générales .....	83
Tableau 2.11 : Accords inter-annotateurs pour le corpus "Reviews" .....	84
Tableau 2.12 : Accords inter-annotateurs pour le corpus "Le Monde" .....	84
Tableau 2.13 : Fréquence de chaque relation de discours annoté .....	91
Tableau 2.14 : Répartition fréquentielle des types de segments reliés par type de relation.....	92
Tableau 2.15 : Préservation de la subjectivité par type de relation .....	93
Tableau 2.16 : Préservation de la subjectivité par type de relation .....	93
Tableau 3.1 : Résultats pour la classe R (borne ouvrante d'un ou plusieurs segments).....	101
Tableau 3.2 : Résultats pour la classe L (borne fermante d'un ou plusieurs segments).....	101
Tableau 3.3 : Résultats pour la classe N (Token n'étant pas une borne de segment) .....	101
Tableau 3.4 : Résultats pour la sur-segmentation (information de subjectivité donnée par oracle) .....	102
Tableau 3.5 : Résultats pour la sur-segmentation (information de subjectivité déterminée par lexique) .....	102
Tableau 3.6 : Résultats pour la sur-segmentation après correction orthographique automatique (information de subjectivité déterminée par lexique).....	102
Tableau 3.7 : Analyse des avantages et inconvénients de GATE .....	106
Tableau 3.8 : Analyse des avantages et inconvénients de Apache UIMA.....	107
Tableau 3.9 : Analyse des avantages et inconvénients de l'environnement interne Synapse ....	108
Tableau 3.10 : Avantages et inconvénients du RegExp Annotator d'UIMA .....	110
Tableau 3.11 : Avantages et inconvénients de Jape .....	111
Tableau 3.12 : Quelques primitives usuelles du Token.....	113
Tableau 3.13 : Types Myélines et leurs équivalents pour la plate-forme cible Synapse Étiquette .....	114



Tableau 3.14 : Avantages et inconvénients de Myéline.....	115
Tableau 3.15 : Récapitulatif des avantages et inconvénients de l'environnement d'exécution final .....	116
Tableau 4.1 : Couverture du lexique .....	128
Tableau 4.2 : Résultats sur tâche extrinsèque .....	129
Tableau 4.3 : Extrait de la liste de synonymes et antonymes de mauvais (sens 1) .....	131
Tableau 4.4 : $\beta$ prédits et accuracy mesurées .....	133
Tableau 4.5 : Génération de lexique d'intensités via la méthode d'extension par taxinomie....	134
Tableau 4.6 : Résultats sur les catégories .....	135
Tableau 4.7 : Répartition des entrées lexicales selon la catégorie sémantique .....	139
Tableau 4.8 : Répartition des entrées lexicales selon leur catégorie syntaxique .....	139
Tableau 5.1 : Traits utilisés pour la classification test.....	152
Tableau 5.2 : Résultats pour la classification S.....	155
Tableau 5.3 : Résultats pour la classification Op.....	155
Tableau 5.4 : Résultats pour la classification finale (Accuracy) .....	157
Tableau 6.1 : Répartition des résultats pour l'hypothèse N1.b .....	171
Tableau 6.2 : Évaluation de l'algorithme de portée de l'opérateur sur le corpus de commentaires web CASOAR.....	175
Tableau 6.3 : Règles de passage des opérateurs vers les opérations élémentaires .....	180
Tableau 6.4 : Intensité normalisée pour le lexique de l'exemple A .....	183
Tableau 6.5 : Intensité normalisée pour le lexique de l'exemple B .....	183
Tableau 6.6 : Tableau de passage de l'échelle $[-5; +5] \cap \mathbb{Z}$ .....	191
Tableau 6.7 : Tableau de passage de l'échelle $-3; +3 \cap \mathbb{Z}$ vers l'échelle des données $[-5; +5] \cap \mathbb{Z}$ .....	191
Tableau 6.8 : Résultats (f-mesure) du modèle Taboada et al. sur nos données pour les différents paramètres d'intensification .....	191
Tableau 6.9 : Résultats des différents modèles sur les annotations CASOAR .....	192
Tableau 6.10 : Résultats des différents modèles sur les annotations CASOAR .....	193
Tableau 6.11 : résultats de l'évaluation de la prise en compte des opérateurs de certitude .....	193
Tableau 7.1 : Valeurs de subjectivité et d'opinion pour chacun des segments.....	198
Tableau 7.2 : Application de chaque filtre par sujet de l'opinion sur l'exemple cas d'étude (cf. section 7.2) .....	200
Tableau 7.3 : Application de chaque filtre par sujet de l'opinion sur l'exemple cas d'étude .....	201
Tableau 7.4 : Priorités des relations coordonnantes .....	209
Tableau 7.5 : Triplets de valeurs associés aux segments de l'exemple 1.....	211
Tableau 7.6 : Triplets de valeurs associés aux segments de l'exemple 2.....	211
Tableau 7.7 : Triplets de valeurs associés aux segments de l'exemple 3.....	212
Tableau 7.8 : Triplets de valeurs associés aux segments de l'exemple 4.....	213
Tableau 7.9 : Triplets de valeurs associés aux segments de l'exemple 5.....	214
Tableau 7.10 : Triplets de valeurs associés aux segments des exemples 6 et 7 .....	215
Tableau 7.11 : Triplets de valeurs associés aux segments de l'exemple 10.....	216
Tableau 7.12 : Triplets de valeurs associés aux segments de l'exemple 11.....	217
Tableau 7.13 : Triplets de valeurs associés aux segments de l'exemple 12.....	218

Tableau 7.14 : Triplets de valeurs associés aux segments de l'exemple 14.....	219
Tableau 7.15 : Triplets de valeurs associés aux segments de l'exemple 15.....	220
Tableau 7.16 : Triplets de valeurs associés aux segments de l'exemple 16.....	221
Tableau 7.17 : Triplets de valeurs associés aux segments de l'exemple 17.....	223
Tableau 7.18 : Triplets de valeurs associés aux segments de l'exemple 18.....	223
Tableau 7.19 : Triplets de valeurs associés aux segments de l'exemple 19.....	224
Tableau 7.20 : Triplets de valeurs associés aux segments de l'exemple 23.....	226
Tableau 7.21 : Triplets de valeurs associés aux segments de l'exemple 24.....	227
Tableau 7.22 : Triplets de valeurs associés aux segments de l'exemple 26.....	228
Tableau 7.23 : Triplets de valeurs associés aux segments de l'exemple 27.....	228
Tableau 7.24 : Constituants de la CDU 13 .....	230
Tableau 7.25 : Constituants de la CDU 14 .....	231
Tableau 7.26 : Constituants de la CDU 16 .....	231
Tableau 7.27 : Constituants de la CDU 15 .....	232
Tableau 7.28 : Constituants du graphe .....	233
Tableau 7.29 : Résultat final par heuristique .....	234
Tableau 7.30 : Valeurs de subjectivité et d'opinion pour chacun des segments.....	235
Tableau 7.31 : Constituants de la CDU 19 .....	236
Tableau 7.32 : Constituants de la CDU 20 .....	237
Tableau 7.33 : Constituants de la CDU 21 .....	237
Tableau 7.34 : Constituants de la CDU 22 .....	238
Tableau 7.35 : Constituants de la CDU 23 .....	238
Tableau 7.36 : Constituants de la CDU 24 .....	240
Tableau 7.37 : Constituants du graphe .....	241
Tableau 7.38 : Valeurs d'opinion considérées comme proches d'une valeur annotée donnée..	244
Tableau 7.39 : Exemple complémentaire : valeurs d'opinion des segments.....	245
Tableau 7.40 : Meilleurs résultats (accuracy), polarité seule, pour chaque heuristique de calcul, filtrages effectués sans relaxation.....	249
Tableau 7.41 : Meilleurs résultats (accuracy), polarité seule, pour chaque heuristique de calcul, filtrages effectués avec relaxation .....	249
Tableau 7.42 : Meilleurs résultats (Pearson), polarité seule, pour chaque heuristique de calcul, filtrages effectués avec relaxation .....	250
Tableau 7.43 : Apport des différents modules de fusion, corpus de commentaires web, polarité seule .....	250
Tableau 7.44 : Meilleurs résultats (accuracy), polarité seule, pour chaque heuristique de calcul, filtrages effectués sans relaxation.....	251
Tableau 7.45 : Meilleurs résultats (accuracy), polarité seule, pour chaque heuristique de calcul, filtrages effectués avec relaxation .....	251
Tableau 7.46 : Meilleurs résultats (Pearson), polarité seule, pour chaque heuristique de calcul, filtrages effectués avec relaxation .....	252
Tableau 7.47 : Apport des différents modules de fusion, corpus de réactions aux articles de presse, polarité seule.....	252

Tableau 7.48 : Meilleurs résultats (accuracy) pour chaque heuristique de calcul, filtrages effectués sans relaxation.....	253
Tableau 7.49 : Meilleurs résultats (accuracy) pour chaque heuristique de calcul, filtrages effectués avec relaxation .....	253
Tableau 7.50 : Meilleurs résultats (corrélation de Pearson) pour chaque heuristique de calcul, filtrages effectués avec relaxation .....	254
Tableau 7.51 : Apport des différents modules de fusion, corpus de commentaires web, polarité et intensité.....	254
Tableau 7.52 : Meilleurs résultats (accuracy) pour chaque heuristique de calcul, filtrages effectués sans relaxation.....	255
Tableau 7.53 : Meilleurs résultats (accuracy) pour chaque heuristique de calcul, filtrages effectués avec relaxation .....	255
Tableau 7.54 : Meilleurs résultats (corrélation de Pearson) pour chaque heuristique de calcul, filtrages effectués avec relaxation .....	256
Tableau 7.55 : Apport des différents modules de fusion, corpus de réactions aux articles de presse, polarité et intensité.....	256
Tableau 8.1 : Fonction de conversion des catégories sémantiques CASOAR vers les catégories DoXa .....	265



# Introduction

---

## L'analyse des sentiments, un domaine applicatif en phase d'industrialisation

L'analyse des sentiments est un domaine de recherche extrêmement actif en traitement automatique des langues. En effet, ces dernières années ont vu se multiplier les sources de données textuelles porteuses d'opinion disponibles sur le web : avis d'internautes, de plus en plus centralisés par les moteurs de recherche, forums, réseaux sociaux, enquêtes consommateurs effectuées par les grandes marques. Devant cette abondance de données et de sources, l'automatisation de la synthèse des multiples avis devient cruciale pour obtenir efficacement une vue d'ensemble des opinions sur un sujet donné. L'intérêt de ces données est considérable, pour les sociétés qui souhaitent obtenir un retour client sur leurs produits ou leur image de marque comme pour les personnes souhaitant se renseigner pour un achat, une sortie, ou un voyage.

Au-delà d'une simple moyenne des opinions relevées, cette synthèse doit s'effectuer de manière fine si l'on souhaite ne pas perdre d'informations. Ainsi, la présentation d'une moyenne d'étoiles, très largement adoptée par les sites de commentaires web, ne permet pas d'exprimer clairement les nuances possibles de l'opinion. La figure suivante, tirée du webcomic XKCD, illustre ce fait de manière humoristique :



Figure 0.1 : Vision humoristique des échelles d'évaluation des sites de commentaires web<sup>1</sup>

Les premiers travaux de recherche en extraction automatique d'opinion remontent à la fin des années 1990 avec en particulier les travaux de (Hatzivassiloglou and McKeown 1997) traitant de

---

<sup>1</sup> Source : XKCD webcomic : <http://xkcd.com/1098/>. Licence : Creative Commons 2.5.

la détermination de la polarité d'adjectifs, et ceux de (Pang, Lee, and Vaithyanathan 2002; Turney 2002) sur la classification de documents suivant leur polarité. Depuis les années 2000, un grand nombre de travaux ont été publiés sur le sujet, faisant de l'extraction d'opinion un domaine très actif dans la recherche en TAL (*Traitement Automatique du Langage*). Citons, entre autres travaux, les travaux de (Wiebe, Wilson, and Cardie 2005) autour de l'annotation du corpus d'opinion MPQA (*Multi-Perspective Question Answering*), les travaux de (Taboada, Voll, and Brooke 2008; Taboada et al. 2011) concernant la prise en compte des effets des opérateurs sur l'opinion, ainsi que les campagnes d'évaluation : la campagne TREC (*Text REtrieval Conference*) (Ounis, Macdonald, and Soboroff 2008), et la campagne DEFT (*Défi Fouille de Textes*) pour le français (Grouin et al. 2007; Grouin et al. 2009).

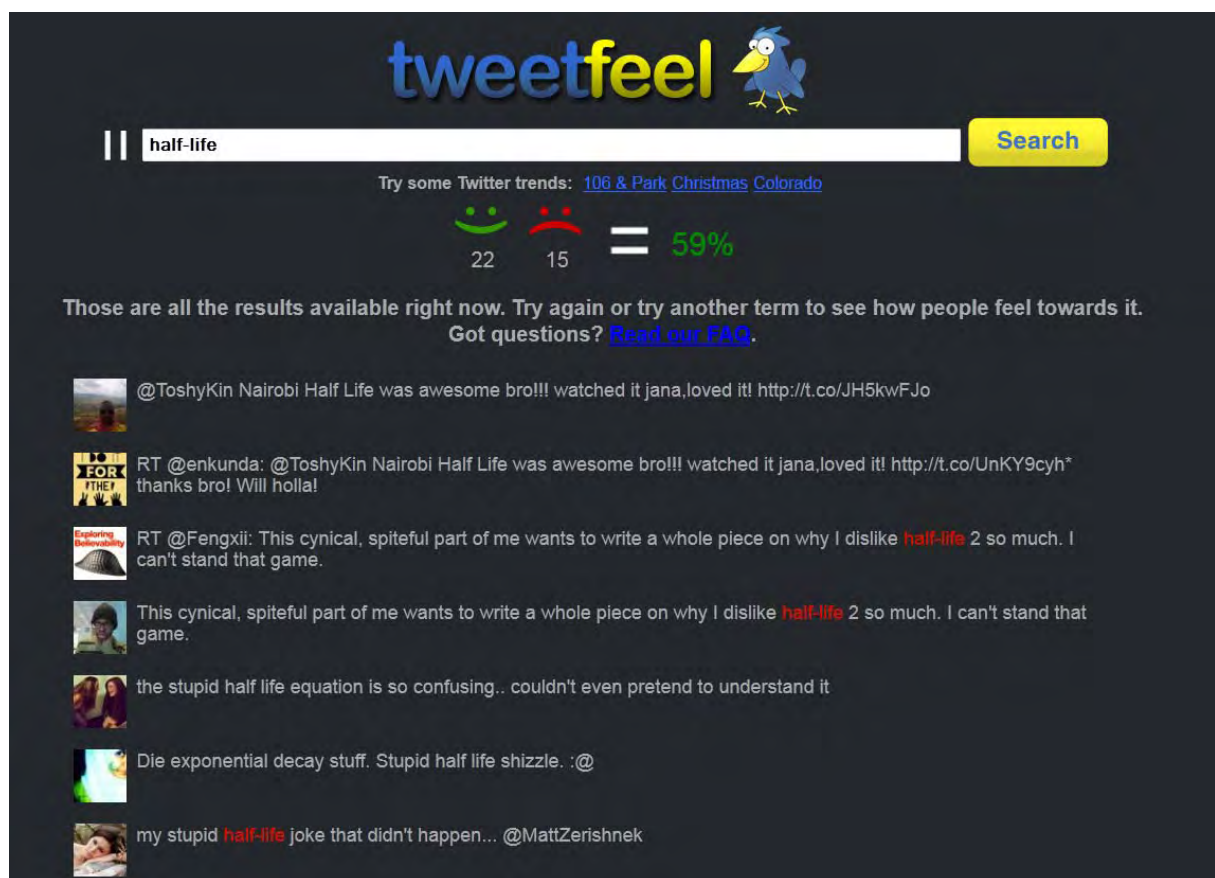


Figure 0.2 : Interface web tweetfeel, comptant les tweets positifs et négatifs sur un sujet

Plus récemment, des adaptations industrielles de l'analyse de sentiment ont vu le jour, proposant notamment aux internautes une extraction de l'opinion à partir de données notamment issues des réseaux sociaux. Citons par exemple Tweetfeel<sup>2</sup> (Figure 0.2 ci-dessus), qui propose une analyse des tweets des derniers jours, en comptant les positifs et les négatifs, ou encore Twitter Sentiment<sup>3</sup> (Figure 0.3 ci-dessous), qui présente le nombre de positifs et négatifs sous forme de camemberts et d'histogrammes. Ces applications sont en général relativement basiques, et présentent des documents classés simplement via une série de mots-clés de forte intensité : par

<sup>2</sup> <http://www.tweetfeel.com/>

<sup>3</sup> Récemment renommé Sentiment 140 : <http://www.sentiment140.com/>

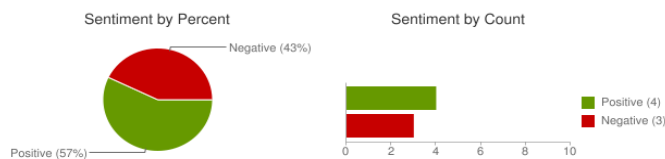
exemple, dans la Figure 0.2, nous pouvons reconnaître les mots-clés relativement intenses et tranchés « *awesome* », « *stupid* », et « *dislike* ». Ceci permet de réduire les faux positifs, mais pas de les supprimer complètement : des subtilités de sujet de l'opinion ou des opérateurs tels que la négation peuvent venir ; ainsi, les exemples positifs et négatifs de la Figure 0.3, sensés être sur Brad Pitt, semblent étranges. D'autre part, ceci n'est possible que dans le cas où l'on dispose d'une masse importante de données, par exemple un flux Twitter.

## Twitter Sentiment

Type in a word and we'll highlight the good and the bad

"brad pitt"  [Save this search](#)

### Sentiment analysis for "brad pitt"



### Tweets about: "brad pitt"

[peridot817](#): **Brad Pitt** is one of us #LFC he will never walk alone cc @Sam1jere  
Posted 38 seconds ago

[606v2rugby](#): @Sourdust #shitlit The **Brad Pitt** and the Pendulum  
Posted 3 minutes ago

[bradleyyboy](#): **Brad Pitt** supports Liverpool? So it seems only good looking/handsome people support them, no wonder we don't have that many supporters... ^^  
Posted 4 minutes ago

[Kevin\\_Quigley](#): @XelaUda @jjaquigley @radavis88 @juniorimoto hahahaha have you seen **Brad Pitt**, Tom Sellick? man up a good cause. #testicular Cancer  
Posted 6 minutes ago

Figure 0.3 : Interface web Twitter Sentiment

Afin de proposer des applications industrielles exploitant au mieux les avancées de la recherche dans le domaine, certains travaux proposent des chaînes de traitements permettant une extraction efficace de l'opinion. Citons notamment pour l'anglais (Taboada, Voll, and Brooke 2008; Taboada et al. 2011) qui proposent la chaîne d'extraction de l'opinion SO-CAL, ou encore pour le français le projet DoXa (Paroubek, Pak, and Mostefa 2010), visant à développer une chaîne de traitement pour les partenaires industriels du projet.

Les travaux de cette thèse, réalisés dans le cadre d'un contrat CIFRE entre le laboratoire IRIT de Toulouse d'une part, et la société Synapse Développement d'autre part, se placent dans cette optique. Nous proposons ainsi plusieurs modules pour une extraction de l'opinion, de l'expression d'opinion au niveau local à l'opinion globale au niveau du document. Ces modules sont par la suite implémentés dans le contexte d'une chaîne de traitement, proposée par la société Synapse Développement. Afin de tirer partie au mieux de l'extraction d'opinion fine au niveau local, nos travaux présentent l'originalité de se baser sur la structure discursive du document.

## La structure discursive, information clef pour une bonne compréhension d'un texte

La structure discursive d'un texte est un élément essentiel à la compréhension du contenu véhiculé par ce texte. Elle affecte, par exemple, la structure temporelle du texte, ou encore l'interprétation des expressions anaphoriques. La structure discursive a montré son utilité dans de nombreuses applications de TAL, telles que le résumé automatique (Marcu 2000), ou encore l'inférence textuelle (Mirkin, Dagan, and Padó 2010). Dans cette thèse, nous aborderons les effets de la structure discursive sur l'analyse de sentiments.

Le discours est un élément essentiel à la bonne compréhension d'un texte. Les textes d'opinion ne font pas exception : considérons par exemple le document suivant, commentaire de série télévisée :

*Les personnages sont antipathiques au possible. Le scénario est complètement absurde. Le décor est visiblement en carton-pâte. Mais c'est tous ces éléments qui font le charme improbable de cette série.*

Sur les quatre opinions de ce texte, les trois premières sont a priori très négatives. Néanmoins, la dernière phrase, en relation de contraste avec les trois précédentes, nous permet de déterminer la véritable polarité du document, qui est positive. Une simple moyenne des opinions aurait ici conduit à un contre-sens, et seule la prise en compte de la structure discursive permet de désambigüiser la polarité globale du document.

La structure discursive est également un bon indicateur de la subjectivité et / ou de la polarité au niveau d'un segment de discours, indépendamment de la présence d'opinion explicite (i.e. lexicalement marquées). Les exemples suivant, extrait d'un corpus de commentaires web, illustrent cela :

- (1) [Le film n'est pas mauvais,]<sub>a</sub> [bien que certaines personnes aient quitté la salle.]<sub>b</sub>
- (2) [Marie a aimé le film.]<sub>a</sub> [Son mari aussi.]<sub>b</sub>
- (3) [Laborieux]<sub>a</sub> [et copier-coller du premier opus.]<sub>b</sub>
- (4) [Très mauvais film.]<sub>a</sub> [Je ne le recommande pas.]<sub>b</sub>
- (5) [Le film était excellent.]<sub>a</sub> [Les acteurs étaient géniaux.]<sub>b</sub>
- (6) [Si vous voulez perdre votre temps,]<sub>a</sub> [allez-voir ce film.]<sub>b</sub>
- (7) [Je suppose]<sub>a</sub> [que votre politique d'embauche sera un désastre.]<sub>b</sub>

Les segments de ces exemples, délimités par des crochets, sont reliés entre eux par les relations discursives suivantes :

- Contraste (1.a, 1.b)
- Parallèle (2.a, 2.b)
- Continuation (3.a, 3.b)
- Résultat (4.a, 4.b)
- Elaboration (5.a, 5.b)
- Conditionnelle (6.a, 6.b)



- Attribution (7.a, 7.b)

Statistiquement<sup>4</sup>, les relations de *Contraste*, *Parallèle*, et *Continuation* (exemples 1, 2, 3) relient le plus souvent des segments tous deux subjectifs ou objectifs. Ceci donne un indice permettant le repérage d'opinions implicites. Par exemple, ceci peut permettre de repérer les opinions portées par les segments 1.b, 2.b, et 3.b, qui ne contiennent pas d'expression d'opinion (i.e. qui ne contiennent aucun terme que l'on peut considérer comme porteur d'opinion hors contexte). De même, ces relations relient le plus souvent soit des segments de même polarité (*Continuation*, *Parallèle*) ou de polarité opposée (*Contraste*), ce qui permet de déduire l'orientation sémantique de ces segments. De même, la relation d'*Elaboration* préserve en général la polarité entre la source et la cible : annoncer [Le film est excellent :]<sub>a</sub> [les acteurs sont mauvais.]<sub>b</sub> n'est pas cohérent discursivement parlant. Enfin, la relation d'*Attribution* permet de lier l'opinion à son émetteur et au degré d'engagement de celui-ci.

La plupart des travaux actuels proposent une analyse des opinions au niveau du document ou au niveau de la phrase en ignorant la structure discursive. Dans cette thèse, nous nous plaçons dans le contexte de la théorie de la SDRT (*Segmented Discourse Representation Theory* – (Asher and Lascarides 2003)) et proposons de répondre aux questions suivantes :

- Existe-t-il un lien entre la structure discursive d'un document et les opinions émises dans ce même document ?
- Quel est le rôle des relations de discours dans la détermination du caractère objectif ou subjectif d'un segment textuel ?
- Quel est le rôle des éléments linguistiques, comme la négation et la modalité, lors de la détermination du score – polarité, intensité – d'un segment textuel subjectif ?
- Quel est l'impact de la structure discursive lors de la détermination de l'opinion globale véhiculée dans un document ?
- Est-ce qu'une approche basée sur le discours apporte une réelle valeur ajoutée comparée à une approche classique basée sur la notion de 'sacs de mots' ?
- Cette valeur ajoutée est-elle dépendante du genre de corpus ?

## Chaîne de traitement pour une approche discursive de l'analyse d'opinion

Afin de répondre à ces questions, nous proposons dans un premier temps un schéma d'annotation sur trois niveaux (chapitre 2) : tout d'abord au niveau de l'expression d'opinion, véhiculant hors contexte une opinion, puis au niveau du segment de discours, où chaque segment doit être annoté selon l'opinion qu'il véhicule, et enfin au niveau du document, où tout segment doit être relié discursivement au reste du texte. Dans le but d'étudier l'impact des contraintes de genre du texte, notre campagne d'annotation a été conduite sur deux corpus de genres différents en français : commentaires de film d'une part, et commentaires d'articles de presse d'autre part. Nous

---

<sup>4</sup> Le chapitre 2 détaille ces statistiques sur nos corpus.

proposons ensuite une architecture informatique (cf. chapitre 3) permettant d'analyser les opinions dans le discours selon différents niveaux de grain :

- *Au niveau du segment* où nous proposons de sélectionner les segments discursifs porteurs d'opinion du document à l'aide de deux classifieurs parallèles. La présence de traits discursifs permet de reconnaître une part importante des segments implicites, c'est-à-dire des segments ne comportant pas d'opinion lexicalisée. (cf. chapitre 5)
- *Au niveau des expressions d'opinion au sein d'un segment* où nous présentons d'une part deux méthodes de génération automatique de lexique d'opinion (cf. chapitre 4), et d'autre part une étude de l'effet linguistique des opérateurs venant altérer les caractéristiques a priori d'une expression d'opinion, tels que les négations, les intensifieurs et les modaux. Nous proposons ensuite un modèle de calcul, d'inspiration géométrique, permettant de déterminer finement l'impact de ces opérateurs, et ainsi de déterminer les caractéristiques de l'opinion au niveau du segment. (cf. chapitre 6)
- *Au niveau du document* où nous comparons plusieurs approches de remontée de l'opinion depuis le niveau des segments de discours jusqu'au niveau du document, afin de calculer l'opinion globale véhiculée par celui-ci. (cf. chapitre 7)

Avant de détailler les différents éléments de notre chaîne de traitement, nous proposons dans le chapitre 1 une présentation du contexte scientifique de cette thèse, et de la littérature du domaine de l'analyse d'opinion.





# 1. Contexte scientifique et motivations

## 1.1. Opinion et évaluations

Le terme *Opinion* peut recouvrir de multiples facettes. En effet, considérons les 3 phrases suivantes :

- (1) Ce restaurant propose des plats délicieux.
- (2) Je suis jaloux du talent du cuisinier.
- (3) Je ne pourrai pas y aller ce soir, je pense qu'il va pleuvoir.

Dans le cadre de nos travaux, nous nous sommes intéressés à l'opinion évaluative, c'est-à-dire à l'évaluation, sur une échelle de valeurs, d'une entité ou d'un concept. La phrase (1) correspond à cette définition. Cette évaluation peut s'exprimer via des expressions de jugement (e.g. « *Ce plat est bon.* »), mais aussi via certains sentiments (e.g. « *J'aime beaucoup ce plat.* »), ou encore via l'émission de conseils ou recommandations (e.g. « *Je vous recommande ce plat.* »). D'autres travaux ont étudié les opinions évaluatives. La théorie Appraisal (Martin 2005) propose dans sa classification les catégories *Judgment* et *Appreciation* (Jugement et Évaluation, cf. figure 1.1).

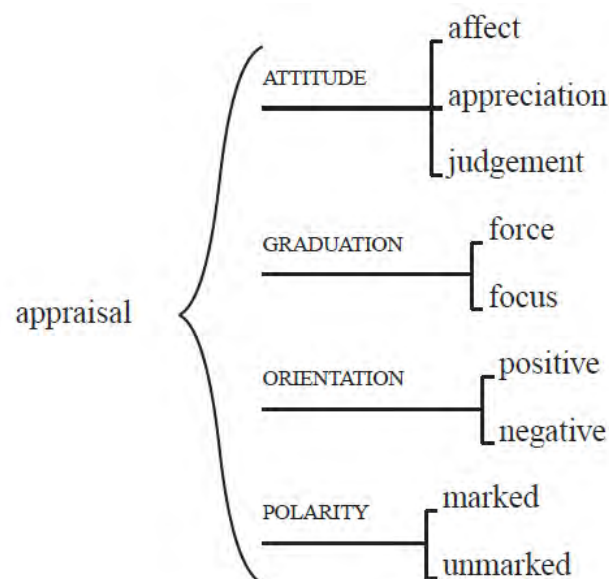


Figure 1.1 : Catégories principales du framework Appraisal (Martin 2005; Whitelaw, Garg, and Argamon 2005a)

Toutefois, tous les sentiments ne rentrent pas dans le cadre de l'évaluation. Par exemple, le sentiment de jalousie, comme dans la phrase (2), peut apparaître indépendamment de l'opinion évaluative portée à une entité. De même, certaines phrases de prédictions, relevant de l'opinion dans le langage courant, ne constituent pas des évaluations. Ainsi, dans la phrase (3), le locuteur donne son avis sur la météo du soir, sans que cela constitue une évaluation du temps en question.

La classification de Quirk (Quirk et al. 1985) regroupe quant à elle les états privés (« private states »), c'est-à-dire tout état de l'individu qui ne peut être observé ou vérifié objectivement. Rentrent dans cette catégorie les sentiments et les évaluations, mais aussi les croyances. Basée sur cette classification, les annotations du corpus MPQA (Wiebe, Wilson, and Cardie 2005) distinguent trois types d'annotations d'états privés :

- Les mentions explicites d'états privés (ex : "The U.S. fears a spill-over," said Xirao-Nima. - (Wiebe, Wilson, and Cardie 2005))
- Les événement de discours exprimant un état privé (ex : "The report is full of absurdities," Xirao-Nima said. - (Wiebe, Wilson, and Cardie 2005))
- Les éléments subjectifs explicites (ex : "We foresaw electoral fraud but not daylight robbery," Tsvan-girai said. - (Wiebe, Wilson, and Cardie 2005))

Pour le français, le projet d'annotations de blogs Blogoscopie (Dubreil et al. 2008; Vernier et al. 2009; Daille et al. 2011) donne un schéma d'annotations, inspiré des travaux de (Charaudeau 1992), distinguant cinq catégories :

- *Opinion* : cette catégorie contient les « faits ou informations présupposées que le rédacteur replace dans son univers de croyance » (Daille et al. 2011). Cette catégorie se divise en deux sous-types : les *convictions* (« Je suis persuadé que ... »), et les *opinions* (« Je pense que ... »), ces dernières pouvant être de quatre niveaux de certitude (*haute, moyenne, faible, ou prémonition*)
- *Appréciation* : cette catégorie comprend les évaluations positives ou négatives émises par le rédacteur. Cette catégorie se décline en six sous-catégories, suivant la polarité et le type d'évaluation : *appréciations explicites positives et négatives, exclamations positives et négatives, et appréciations implicites positives et négatives*.
- *Acceptation-Refus* : cette catégorie inclut les acceptations et refus par l'émetteur de l'opinion d'un acte qui lui serait demandé.
- *Accord-Désaccord* : cette catégorie comprend les mentions par l'émetteur du fait qu'il approuve ou désapprouve une opinion, une position, ou une action. Cet accord/désaccord peut se décliner en quatre sous-catégories : *accord complet, accord vague* (« Je suis globalement d'accord... »), *accord avec correction* (« Oui, mais... »), *désaccord complet*.
- *Jugement* : cette catégorie contient les jugements de valeurs, positifs ou négatifs, sur les actes d'une personne. Ces jugements peuvent être *explicites positifs, explicites négatifs, ou implicites* (dans le sens où l'auteur du jugement est *implicite*) : ce dernier cas recouvre les catégories *félicitations, pardon, accusation, reproche*.

Dans cette taxinomie, les catégories Appréciation et Jugement sont des catégories d'opinions évaluatives sur un sujet.

Les opinions non évaluatives ont également été étudiées dans la littérature : citons notamment la catégorie *Affect* du framework *Appraisal* (Martin 2005). (Chen 2008) propose une catégorisation transverse basé sur six catégories de subjectivité : incertitude (*uncertainty*), imprécision (*imprecision*), caractère vague de l'assertion (*vagueness*), ambiguïté (*ambiguity*), gradation sur une échelle non mesurable (*non-objective measurability*), et non-objectivité (*non-objectivity*).

Parmi ces catégories, l'incertitude peut être associée à des opinions non évaluatives lorsqu'elle n'est pas associée à une autre catégorie : elle joue dans ce cas le rôle d'opérateur de certitude. La catégorisation de (Chen 2008) ne permet par contre pas de distinguer entre un sentiment évaluatif et un sentiment non évaluatif, tous deux pouvant être classés dans la catégorie non-objectivité.

Dans cette thèse, nous proposons d'étudier les opinions évaluatives selon la catégorisation sémantique décrite dans (Asher, Benamara, and Mathieu 2008), dans le cadre d'une analyse discursive du document : cette analyse permettra de faire le lien entre l'opinion au niveau d'une expression subjective, et l'opinion au niveau du document. Cette classification distingue quatre types principaux d'opinions : les jugements, les sentiments, les conseils/recommandations, et les reportages. La section 2.2.2 donne plus de détail sur cette classification.

## **1.2. Analyse de l'opinion : un problème à différents niveaux de granularité**

L'opinion évaluative est une notion présente à différents niveaux de granularité : en effet, nous pouvons la retrouver à un niveau local, interne à la phrase, où une portion de texte véhicule une opinion unitaire, c'est-à-dire une opinion portée par une expression d'opinion unique. Dans l'exemple suivant, « excellent » répond à cette définition :

*(4) Ce repas était excellent.*

Nous pouvons également retrouver la notion d'opinion évaluative à un niveau global, à l'échelle du document. Par exemple, nous pouvons reprendre le commentaire de série télévisée cité précédemment :

*Les personnages sont antipathiques au possible. Le scénario est complètement absurde. Le décor est visiblement en carton-pâte. Mais c'est tout ces éléments qui font le charme improbable de cette série.*

Ce document véhicule dans son ensemble une opinion sur la série commentée. Il s'agit de l'opinion globale portée par la totalité du document, résumé des informations portées au niveau local, mais également de l'interaction entre elles.

La section 1.2.1 traite des avancées de la littérature en regard de la recherche de l'opinion au niveau du document. La section 1.2.2 traite quant à elle de l'extraction de l'opinion au niveau local. Enfin, la section 1.2.3 présente les travaux introduisant la structure discursive du document dans l'analyse de l'opinion, permettant de lier ces deux niveaux.

### **1.2.1. L'opinion au niveau du document : extraction d'un sentiment global sur un sujet principal**

L'extraction d'une opinion au niveau d'un document est un problème pouvant s'apparenter à un problème de classification : le document porte-t-il une opinion ? Est-il positif ? Est-il négatif ? On peut subdiviser ce problème de classification en deux problèmes distincts :

- Déterminer si un document véhicule une opinion (i.e. est subjectif) ou ne fait que présenter des faits (i.e. est objectif). Cette tâche est communément appelée l'analyse de subjectivité.
- Déterminer l'opinion effectivement véhiculée par un document que l'on sait subjectif. Cette tâche est connue sous le nom de l'analyse de polarité.

Nous présentons en section 1.2.1.1 la tâche d'analyse de la subjectivité, puis en section 1.2.1.2 la tâche d'analyse de polarité. En préambule, nous présentons rapidement les méthodes d'apprentissage automatique, notamment utilisée pour ces tâches.

Pour traiter ces problèmes de très nombreuses approches emploient une méthode d'apprentissage automatique.

Appliqué au TAL, le principe de l'apprentissage automatique est de déduire d'un corpus d'apprentissage un module permettant de discriminer des textes ou portions de textes en plusieurs classes (par exemple, classer des documents en positif ou négatif). Cet apprentissage peut-être supervisé – le corpus utilisé pour l'apprentissage est préalablement annoté par un juge humain afin d'identifier les classes à repérer – ou non supervisé. Dans cette section, nous verrons trois exemples de classifieurs supervisés, couramment utilisées pour la classification de documents : le classifieur Bayésien naïf (*Naive Bayes*), le classifieur Maximum Entropie (*MaxEnt*), et le classifieur SVM (*Support Vector Machine*).

Le classifieur Bayésien naïf est inspiré de la formule de calcul de probabilité conditionnelle de Bayes :

$$P(c|d) = \frac{P(c) * P(d|c)}{P(d)}$$

L'idée est d'attribuer au document  $d$  la classe  $c_0$  telle que  $c_0 = \operatorname{argmax}_c(P(c|d))$ . Pour cela et suivant la loi de Bayes, il suffit d'estimer les probabilités  $P(c)$  et  $P(d|c)$  pour chacune des classes  $c$  –  $P(d)$  est constant quel que soit la classe considérée. Pour cela, le document  $d$  est décomposé en une série de traits  $\{f_1, \dots, f_n\}$ , avec  $k_i(d)$  la fréquence de  $f_i$  dans le document  $d$ . Le classifieur Bayésien naïf est basé sur l'hypothèse que ces traits sont indépendants : nous avons donc  $P(d|c) = \prod_i P(f_i|c)^{k_i(d)}$ . L'apprentissage consiste donc à estimer la valeur de  $P(c)$  et de  $P(f_i|c)$  pour chaque trait  $f_i$ . De nombreux travaux utilisent ce type de classifieur, citons notamment (Pang, Lee, and Vaithyanathan 2002; Yu and Hatzivassiloglou 2003; Dave, Lawrence, and Pennock 2003). De nombreuses implémentations de ce classifieur sont disponibles : citons en particulier son implémentation au sein du package Weka<sup>5</sup>.

Les classifieurs Maximum Entropie (MaxEnt) sont basés quant à eux sur le calcul de l'entropie de Shannon. Le principe général est identique – attribuer au document  $d$  la classe  $c_0$  telle que  $c_0 = \operatorname{argmax}_c(P(c|d))$  – mais la formule d'estimation prend la forme suivante :

---

<sup>5</sup> <http://sourceforge.net/projects/weka/>



$$P(c|d) = \frac{1}{Z(d)} * \exp\left(\sum_i \mu_{i,c} * F_{i,c}(d, c)\right)$$

avec  $Z(d)$  une fonction de normalisation, et  $F_{i,c}(d, c')$  valant 1 si le trait  $f_i$  est présent dans le document  $d$  et si  $c' = c$ , et 0 sinon. L'entraînement consiste à évaluer les paramètres  $\mu_{i,c}$  afin de maximiser l'entropie induite. L'algorithme MaxEnt, contrairement au classificateurs Bayésien naïfs, a l'avantage de ne pas effectuer d'hypothèse sur l'indépendance des traits de classification.

Enfin, l'idée derrière l'algorithme SVM (Boser, Guyon, and Vapnik 1992) est, pour deux classes, de calculer un hyperplan permettant de séparer les exemples suivant leur classe. Cet hyperplan est déterminé à partir des exemples du corpus d'apprentissage, en maximisant la marge séparant l'hyperplan de ces exemples. Les SVM se sont régulièrement montré efficace dans le cadre de classification d'opinion : citons entre autres (Pang, Lee, and Vaithyanathan 2002; Whitelaw, Garg, and Argamon 2005a; Plantié et al. 2007).

### 1.2.1.1. Déterminer le caractère subjectif d'un document

Si pour certains genres de corpus le caractère subjectif et porteur d'opinion d'un document ne fait pas de doute (par exemple, les commentaires web sur des restaurants ou des films sont par définition porteurs d'opinion), cela n'est pas le cas pour tous les types de textes. On peut par exemple se demander si un article de presse est purement objectif (i.e. factuel : par exemple, une dépêche d'agence de presse présentant des faits d'actualité sans les commenter), ou au contraire subjectif (par exemple, un article éditorial).

(Yu and Hatzivassiloglou 2003) proposent une classification d'articles de presse suivant quatre catégories du *Wall Street Journal* : *Editorial*, *Letter to Editor*, *News*, et *Business*. Ces quatre catégories sont théoriquement associées à des articles tous subjectifs ou tous objectifs. Le classifieur utilisé est un classifieur Bayésien naïf utilisant des unigrammes comme traits, sans stemmatisation. (Wiebe et al. 2004) utilisent un algorithme de classification basé sur les  $k$  plus proches voisins (Cover and Hart 1967) pour une classification de subjectivité au niveau du document. Celle-ci s'appuie sur la détection au niveau local d'*éléments potentiellement subjectifs* (*Potential Subjective Elements – PSE*), c'est-à-dire d'éléments pouvant être subjectifs étant donné un certain contexte. Par exemple, en français, le terme « navet » peut être subjectif dans le domaine du cinéma, où il désigne un mauvais film et non un légume.

La campagne d'évaluations TREC (Text REtrieval Conference) comprenait jusqu'en 2008 un parcours d'extractions d'informations dans les blogs (Blog Track – (Ounis, Macdonald, and Soboroff 2008)). Ce parcours incluait une section de tâche de détection de l'opinion, consistant à détecter la présence d'une opinion dans des articles de blogs. Parmi les participants, (He et al. 2008) proposent un modèle de détection basé sur une classification Maximum Entropy sur chaque phrase du document : les résultats de ces classifications servent ensuite de traits pour un autre classifieur Maximum Entropy au niveau du document.

La campagne DEFT'09 (Défi Fouille de Textes) comprend, pour le Français, l'Anglais, et l'Italien, une tâche de reconnaissance du caractère subjectif au niveau d'un document. Le corpus utilisé était un corpus d'articles de presse. Parmi les participants, les meilleurs résultats (Grouin et al. 2009)

ont été obtenus pour le Français et l'Anglais par le système de (Bestgen and Lories 2009), qui propose une classification SVM standard, basée sur des unigrammes, bigrammes, et trigrammes lemmatisés et filtrés par seuil de fréquence ; différents essais d'optimisation des paramètres n'ont pas montré d'améliorations par rapport à ceux par défaut. Les autres participants ont proposé des approches basées sur l'algorithme des k plus proches voisins (Forest et al. 2009) et sur l'utilisation de lexiques spécialisés comme traits d'apprentissage pour SVM (Toprak and Gurevych 2009), sans toutefois atteindre les résultats de (Bestgen and Lories 2009).

### 1.2.1.2. Déterminer la polarité au niveau d'un document

Les travaux sur la classification de document par polarité ont pris leur essor au début des années 2000. Différentes approches ont été étudiées pour résoudre cette tâche : nous distinguons ici les approches par classification supervisée au niveau du document (sous-section a) et les approches basées sur une analyse préalable de l'opinion au niveau local de l'expression d'opinion (sous-section b). D'autres travaux ont également proposés des voies d'amélioration des résultats : nous regroupons ici les approches visant à adapter la classification au domaine (sous-section c) et les approches par méta-classification – i.e. se basant sur la combinaison de plusieurs classifieurs (sous-section d).

#### a. Approches par classification supervisée au niveau du document

(Pang, Lee, and Vaithyanathan 2002) ont testé trois types de classifieurs automatiques supervisés – *Machine à Support Vectoriel* (SVM - (Boser, Guyon, and Vapnik 1992)), *Bayésien naïf* (NB – (Mitchell 1997)), et *Maximum Entropie* (MaxEnt) – afin de déterminer la polarité d'un document. Leurs expérimentations, réalisées sur un corpus de commentaires web de films, testent également plusieurs intuitions sur les traits : unigrammes ou bigrammes, fréquences de traits ou présence de traits, nature grammaticale du mot. Les résultats montrent en général un avantage à la classification SVM sur les autres classifieurs utilisés. Le corpus utilisé pour les évaluations est disponible en ligne<sup>6</sup>. Dans cette lignée, (Dave, Lawrence, and Pennock 2003) ont utilisé un classifieur Bayésien et se sont focalisés sur la sélection et l'optimisation des traits de classification, en proposant notamment des remplacements systématiques de méta-données (nombres, noms de produit), et l'inclusion de chaînes de mots de longueurs variables (i.e. la détection automatique d'expressions pertinentes pour la classification). (Matsumoto, Takamura, and Okumura 2005) proposent quant à eux d'inclure au classifieur SVM des traits issus de la structure syntaxique du texte : ordre des mots dans le texte, arbre de dépendance syntaxique, et patrons syntaxiques fréquents. Les résultats obtenus sur le corpus de (Pang, Lee, and Vaithyanathan 2002) montrent un gain en pertinence (accuracy) important – de l'ordre de +5 %. Pour le Français et dans le contexte de l'évaluation DEFT'07, (Plantié et al. 2007) ont testé plusieurs classifieurs supervisés : un classifieur Naive Bayes, un classifieur SVM, un classifieur basé sur les réseaux de neurones RBF, et une approche basée sur les arbres de décision. Le classifieur SVM donne globalement les meilleurs résultats ; l'approche par réseaux RBF donne également de bons résultats sur certains

---

<sup>6</sup> <http://www.cs.cornell.edu/people/pabo/movie-review-data/>

corpus, en particulier ceux de développement, mais les auteurs observent une dégradation importante sur certains des corpus de test.

### ***b. Approches basées sur une analyse préalable au niveau local***

(Turney 2002) propose quant à lui une méthode non supervisée de classification basée sur les informations d'opinion relevée au niveau du mot. À partir du calcul du point d'information mutuel (*Pointwise Mutual Information* – PMI) au niveau du mot, il détermine la polarité au niveau du document via une heuristique de moyenne des scores des adjectifs et adverbes repérés. Le PMI suit, pour  $x$  et  $y$  indépendants, la formule suivante :

$$pmi(x, y) = \log\left(\frac{p(x, y)}{p(x)p(y)}\right)$$

L'idée de la classification à partir d'opinions ponctuelles relevées au niveau du mot a également été exploitée par plusieurs autres travaux. (Mullen and Collier 2004) ont quant à eux utilisé un classifieur SVM, et ont introduit dans leur jeu de traits de classification des informations issues du calcul de l'orientation sémantique au niveau du mot, suivant les méthodes présentées par (Turney 2002) (calcul du PMI, cf. paragraphe suivant) et (Kamps and Marx 2002) (proximité synonymique calculée avec WordNet, suivant les trois axes définis par Osgood (Osgood, Suci, and Tannenbaum 1967) : *potency/force* – faible ou forte, *activity/activité* – actif ou passif, et *evaluative/évaluatif* – bon ou mauvais). Les résultats obtenus montrent un gain significatif en pertinence (*accuracy*) (de l'ordre de +2 %). (Whitelaw, Garg, and Argamon 2005a) ont utilisé un classifieur SVM à partir d'informations lexicales issues du modèle *Appraisal* de (Martin 2005). (Denecke 2009) comparent une classification automatique supervisée (*Simple Logistic Classifier*, implémenté au sein de la distribution *Weka*<sup>7</sup>) à une approche à base de règles, toutes deux basées sur les scores de sentiments de la ressource lexicale *SentiWordNet* (Esuli and Sebastiani 2006a)<sup>8</sup>, sur une sélection de corpus de commentaires portant sur des domaines distincts. Les résultats montrent une nette prédominance de la première solution sur l'ensemble des domaines (gain moyen de l'ordre de 15 % en pertinence/*accuracy*).

### ***c. Approches visant à traiter le problème d'adaptation au domaine***

(Aue and Gamon 2005) observant que le passage d'un domaine d'application à un autre nécessitait en général l'annotation de nouvelles données afin de ré-entraîner les classifieurs, se sont penchés sur la problématique d'un classifieur multi-domaine. Pour cela, ils ont étudié quatre pistes pour la création d'un classifieur optimal en minimisant la réannotation spécifique sur le domaine ciblé :

- L'entraînement sur un corpus mixte, constitué de données provenant de plusieurs autres domaines.
- L'entraînement sur corpus mixte, avec limitations des traits de classification à ceux observés dans le domaine cible.

<sup>7</sup> <http://www.cs.waikato.ac.nz/ml/weka/>

<sup>8</sup> <http://sentiwordnet.isti.cnr.it/>

- L'utilisation d'un ensemble de classifieurs, chacun entraînés sur un domaine distinct où des données sont disponibles. Le choix s'effectue suivant le résultat d'un meta-classifieur, entraîné sur les sorties des classifieurs de chaque domaine.
- L'entraînement d'un classifieur sur la combinaison d'une petite quantité de données annotées et d'une grande quantité de données non annotées sur le domaine ciblé. Pour cela, un classifieur intermédiaire est utilisé pour évaluer la classe des données non annotées à partir de celles annotées.

Parmi ces méthodes, la dernière montre de meilleurs résultats, pourvu que la quantité d'exemple annoté dans le domaine cible ne descende pas au dessous d'un certain seuil (entre 50 et 100 documents suivant les tests présentés). Sur le même sujet, (Blitzer, Dredze, and Pereira 2007) testent l'adaptation de leur algorithme d'apprentissage SCL (*Structure Correspondence Learning* (Blitzer, McDonald, and Pereira 2006)) au domaine de l'analyse de sentiments, et évaluent une mesure de distance entre domaines (*A-distance*, (Ben-David et al. 2006)).

#### **d. Approches par méta-classification (classifications sur la base de plusieurs sous-classifieurs)**

(Torres-Moreno et al. 2007; Torres-Moreno et al. 2009), pour le Français et dans le cadre du Défi Fouille de Textes (DEFT'07 et DEFT'09) utilisent quant à eux une approche basée sur plusieurs classifieurs automatiques :

- le classifieur LIA-SCT (Béchet, Nasr, and Genet 2000), basé sur les arbres de décisions sémantiques (Semantic Classification Tree – (Kuhn and De Mori 1995)),
- le classifieur BoosTexter (Schapire and Singer 2000), basé sur l'algorithme AdaBoost (Freund and Schapire 1996),
- le classifieur SVM Torch (Collobert, Bengio, and Mariéthoz 2002), basé sur SVM,
- le classifieur Timble (Daelemans et al. 2007), inspiré de l'algorithme des k plus proches voisins,
- une modélisation probabiliste simplifiée à base de n-grammes lemmatisés

Ces classifieurs sont ensuite utilisés en parallèle sur les textes à classer : la classification finale est déterminée par vote. Cette méthode de vote montre des résultats supérieurs à ceux des méthodes de classification utilisées seules, et a mené aux meilleurs résultats sur l'évaluation DEFT'07 (Grouin et al. 2007; Paroubek et al. 2007).

Dans le cadre de cette même évaluation, (Vernier et al. 2007) ont également proposé une approche mixte, basée sur deux classifieurs. Le premier classifieur se base sur l'apprentissage de n-grammes caractéristiques d'une catégorie d'opinion (négative comme dans "*une vraie catastrophe*", et positive comme dans "*très belle œuvre*"), et inclut une série de traitements linguistiques (lemmatisation, détection de la négation, représentation sémantique commune à une classe de synonymes). Le second classifieur se base sur une analyse linguistique fine de chaque phrase, prenant en compte des éléments lexicaux (expressions d'opinion, opérateurs linguistiques – négations, modalités, intensités), ainsi que certains éléments surfaciques ayant trait au discours (marques de concession, position de la phrase en début ou fin de texte) ; les indices ainsi relevés permettent de calculer deux scores, positifs et négatifs, s'additionnant pour donner un score au

niveau du document. Les résultats de ces deux classifieurs sont ensuite utilisés par un méta-classifieur supervisé afin de produire la classification finale.

Enfin, (Wang and Manning 2012), remarquent que les classifieurs SVM ont tendance à mieux se comporter sur des documents de taille classique, alors que les classifieurs Bayésiens se comportent mieux sur de courts extraits. Partant de cette observation, ils proposent d'inclure à un classifieur SVM des traits issus d'une classification Bayésienne. Les résultats obtenus dépassent les classifieurs SVM et Bayésiens seuls, quelque soit la longueur du document considéré.

### 1.2.2. L'opinion au niveau local : extraction d'une opinion unitaire véhiculée par une portion de texte

Si la problématique d'identification de l'opinion au niveau du document pouvait facilement s'apparenter à une problématique de classification de textes, l'extraction d'opinions au niveau local – i.e. interne à la phrase – relève plus d'un problème d'extraction d'information. Plusieurs questions se posent pour cette tâche :

- Quelle unité de texte est pertinente pour isoler une opinion unitaire ?
- Comment caractériser les opinions ainsi extraites ?
- Comment identifier les liens entre l'opinion et son contexte immédiat, i.e. comment identifier le sujet, l'émetteur d'une opinion unitaire, et quels opérateurs viennent impacter sur une opinion unitaire ?

#### 1.2.2.1. Catégoriser l'opinion au grain local

Parmi les traits de l'opinion, au-delà de la simple polarité, certains travaux proposent une échelle d'intensité pour les opinions. Dans le cadre du projet MPQA (Multi Perspective Question Answering) et de la constitution du corpus éponyme, (Wiebe, Wilson, and Cardie 2005; Wilson, Wiebe, and Hwa 2006) distinguent quatre niveaux d'intensité, allant de faible (*low*) à extrême (*extreme*). (Mathieu 2008) relève la présence d'intensifieurs inclus dans certains verbes (par exemple, "Les fantômes terrifient Luke" est équivalent à "Les fantômes effraient beaucoup Luke"), leur donnant une intensité altérée même en l'absence d'opérateur. (Asher, Benamara, and Mathieu 2009) définissent une notion similaire de force, sur une échelle numérique allant de 1 à 3.

Au-delà d'une classification en termes de positivité ou de négativité de l'opinion, certains travaux se sont intéressés à une analyse linguistique, voire psychologique de cette notion. (Mathieu 2000; Mathieu and Felbaum 2010) proposent une taxinomie des sentiments décrivant 27 catégories pour le français, certaines étant associées à une polarité (par exemple, *Amuse/Amuser* de polarité positive, ou *Worry/Ennui* de polarité négative), d'autres ne l'étant pas a priori (par exemple, *Astonish/Étonner*). (Taboada et Grieve 2004), (Whitelaw, Garg, et Argamon 2005a), et (Whitelaw, Garg, et Argamon 2005b) utilisent le modèle Appraisal (Martin 2005), introduit précédemment (cf. section 1.1), afin de situer chaque expression dans une large taxinomie. (Vernier et al. 2009) divisent l'évaluation en plusieurs catégories (modalités) : l'Opinion, l'Accord ou le Désaccord, l'Acceptation ou le Refus, le Jugement, l'Appréciation. (Paroubek, Pak, et Mostefa 2010) pour le projet DoXa proposent quant à eux 17 catégories, inspirées des catégories de sentiment de (Mathieu 2000; Mathieu and Felbaum 2010). (Nicholas Asher, Benamara, et Mathieu 2009)

distinguent quant à eux les quatre catégories spécifiques à l'opinion suivantes : Sentiment, Jugement, Conseil, et Reportage. Par rapport aux catégories de Blogoscopie, nous traitons certains Accord/Désaccord (notamment ceux de la forme "[oui], mais [non]" au niveau du discours via une relation de Contraste.

#### 1.2.2.2. Extraction de passages porteurs d'opinion

Si le grain d'un document n'est pas toujours le plus adapté suivant les tâches, il n'est pas toujours évident de déterminer quel niveau de granularité est pertinent pour extraire une opinion.

Plusieurs travaux se sont ainsi attachés à reconnaître l'opinion au niveau de la phrase ou d'une partie définie de la phrase. (Hu and Liu 2004) proposent une méthode d'identification de la polarité d'une opinion au niveau de la phrase en se basant sur le repérage et le comptage d'adjectifs positifs ou négatifs portant sur des caractéristiques du sujet principal préalablement identifiées. L'information ainsi extraite est ensuite utilisée afin de générer un résumé au niveau d'un document. Cette approche est par la suite reprise au sein des travaux de (Ding, Liu, and Yu 2008), qui traitent également les négations, et résolvent certaines ambiguïtés de sens par le contexte local. (Bethard et al. 2004) s'intéressent quant à eux à la reconnaissance de l'opinion au niveau de la proposition, en se basant sur la structure syntaxique et sur la reconnaissance de mots d'opinion, utilisés au sein d'un classifieur SVM. (Kim and Hovy 2006a), pour la tâche d'identification des arguments pour ou contre un sujet – tâche connexe à la problématique d'identification de l'opinion au niveau local – se sont appuyés sur une segmentation au niveau de la phrase. Ces phrases sont ensuite classées en pour ou contre le sujet considéré au moyen d'une classification automatique Maximum Entropie. Citons enfin (Zirn et al. 2011), qui ont raisonné au niveau du segment discursif, et se sont basés sur l'absence ou la présence d'une relation RST (*Rhetorical Structure Theory*) de contraste entre deux segments pour détecter la polarité des segments

D'autres travaux ont considérés le niveau du mot ou de l'expression pour l'extraction de l'opinion unitaire. Citons entre autres (Wiebe, Wilson, and Cardie 2005), (Wilson, Wiebe, and Hwa 2006) pour le projet MPQA<sup>9</sup>, (Kim and Hovy 2005), (Dubreil et al. 2008; Daille et al. 2011) pour le projet Blogoscopie, (Asher, Benamara, and Mathieu 2009), et (Paroubek, Pak, and Mostefa 2010) pour le projet DoXa.

#### 1.2.2.3. Associer l'opinion à son contexte local : sujet, émetteur, opérateurs

##### *Sujet*

L'annotation du sujet de l'opinion est une notion qui a été étudiée dans plusieurs travaux. Citons notamment (Stoyanov and Cardie 2008) qui ont proposé une couche complémentaire d'annotations sur le corpus MPQA identifiant les sujets des opinions, (Kim and Hovy 2006b), le projet Blogoscopie (Dubreil et al. 2008), et le projet DoXa (Paroubek, Pak, and Mostefa 2010).

---

<sup>9</sup> Multi-Perspective Question Answering Opinion Corpus – <http://www.cs.pitt.edu/mpqa/>

La problématique de classification des sujets et de leur hiérarchisation par rapport à un sujet principal (i.e. définition de sujets partiels, représentant une partie, un aspect du sujet principal) a été étudiée entre autres par (Hu and Liu 2004) et (Popescu and Etzioni 2007), qui se sont penchés sur l'extraction automatique de caractéristiques techniques de produits dans les avis donnés par les consommateurs (*web reviews*), à l'aide de la fréquence relative des groupes nominaux présents dans le texte. (Whitelaw, Garg, and Argamon 2005a; Whitelaw, Garg, and Argamon 2005b) introduisent quant à eux la notion proche de focus, c'est-à-dire si l'opinion émise est focalisée sur le sujet principal du document ou non. Un focus extrêmement haut indique une opinion portant sur le sujet principal, alors qu'un focus extrêmement faible indique une opinion portant sur un sujet sans rapport avec celui-ci ; une valeur intermédiaire indique une opinion portant sur une partie du sujet ou sur un sujet proche. (Dubreil et al. 2008) distinguent trois types de sujet de l'opinion, suivant l'écart sémantique entre le sujet de l'opinion considérée et le sujet principal du document : les *Concepts Concernés* (*Concerned Concepts – CC*), les *Concepts Associés* (*Associated Concepts – AC*), et les *Concepts Non-associés* (*Non associated Concepts – NC*).

Notre modélisation du sujet de l'opinion se rapproche de celle de (Dubreil et al. 2008) et de la notion de focus de (Whitelaw, Garg, and Argamon 2005a; Whitelaw, Garg, and Argamon 2005b), avec trois types de sujets de l'opinion suivant l'écart entre le sujet et le sujet principal du document. Ceci sera détaillé en section 2.2.3.

### *Emetteur*

La notion d'émetteur de l'opinion a été étudiée par de nombreux travaux. Citons notamment (Bethard et al. 2004) qui ont associé l'émetteur à l'opinion au niveau de la proposition. Les annotations de MPQA comportent également de manière indirecte la position de l'émetteur par rapport au rédacteur, en précisant la chaîne complète de reportage (« nested source », i.e. la liste des locuteurs du discours rapporté). L'exemple suivant est extrait de (Wiebe, Wilson, and Cardie 2005) :

*“The U.S. fears a spill-over,” said Xirao-Nima. “fears”* est ici émise et rapportée par la chaîne <writer, Xirao-Nima, U.S.>.

Les annotations du projet Blogoscopie se concentrant sur des articles de blogs, les documents sont par nature associés à leur auteur. Ce lien implicite peut être généralisé aux opinions énoncées par le rédacteur.

La reconnaissance automatique de l'émetteur de l'opinion a été étudiée par (Choi et al. 2005), qui ont proposé une méthode de détection de la source de l'opinion via un classifieur Conditional Random Fields (CRF – (Lafferty, McCallum, and Pereira 2001)), entraîné à partir de traits lexicaux, syntaxiques, et d'une annotation sémantique. (Kim and Hovy 2006b) ont quant à eux utilisé un étiqueteur de rôle sémantique, basé sur un classifieur Maximum Entropy (MEGAM<sup>10</sup>) pour la reconnaissance de l'émetteur. (Johansson and Moschitti 2010) ont proposé une classification SVM

---

<sup>10</sup> <http://www.cs.utah.edu/~hal/megam/>

pour l'étiquetage de rôles sémantiques, associée à un modèle de réordonnement pour affiner les résultats.

Notre modélisation, détaillée en section 2.2.4, distingue le locuteur et les autres émetteurs de l'opinion. Cette distinction s'inscrit dans la lignée des travaux de (Wiebe, Wilson, and Cardie 2005), où les auteurs spécifient la source implicite "*writer*".

### **Opérateurs**

Les opérateurs portant sur l'opinion sont des éléments linguistiques ne portant pas intrinsèquement d'opinion, mais altérant les caractéristiques des opinions dans leur portée : par exemple, les négations venant influencer sur la polarité d'une opinion ; les intensifieurs viennent (ex : très, peu...) quant à eux altérer l'intensité de celle-ci. Les opérateurs ont été pris en compte et traités de diverses façons dans l'état de l'art, partiellement ou dans leur globalité.

De nombreux travaux proposent un traitement de la négation pour l'analyse de l'opinion. La plupart de ces travaux adoptent une approche de simple renversement de polarité : citons notamment (Wilson, Wiebe, and Hoffmann 2005; Polanyi and Zaenen 2006; Moilanen and Pulman 2007; Choi and Cardie 2008, 2).

Dans le cadre de la théorie Appraisal, (Whitelaw, Garg, and Argamon 2005b) modélisent les opérateurs suivant les modifications qu'ils apportent aux traits de l'opinion, en particulier la polarité et l'intensité. (J. Liu and Seneff 2009) traitent de manière plus générale les adverbes modifiant l'opinion – i.e. certains opérateurs, incluant les négations – via un modèle linéaire additif. (Wilson, Wiebe, and Hoffmann 2009) ont traité les négations et inversions de polarité en ajoutant simplement un attribut de polarité contextuelle. (Taboada et al. 2011) ont isolé les opérateurs de négation et d'intensité, et ont proposé un modèle calculatoire de leur impact : décalage fixe d'intensité pour la négation et facteur multiplicatif pour les opérateurs d'intensité.

Le chapitre 6 présente un état de l'art plus détaillé sur l'utilisation et la prise en compte des opérateurs. Citons également (Wiegand et al. 2010), qui décrit en détails le rôle des négations dans le contexte des opinions.

### **1.2.3. La structure discursive, information clef pour une extraction fine au niveau du document**

Comme nous avons pu le voir, ces deux niveaux d'extraction – le document d'une part et l'expression d'opinion d'autre part – recouvrent des problématiques et des méthodes distinctes. Ces deux niveaux ne sont pour autant pas indépendants l'un de l'autre, et la classification au niveau du document peut par exemple tirer partie de celle au niveau local. Nous allons voir dans la section suivante que certains travaux se sont intéressés à la structure discursive comme lien entre ces deux tâches.

Considérons les deux phrases suivantes :

- *Ce restaurant est très bon, mais cher.*
- *Ce restaurant est cher, mais très bon.*



Ces deux phrases sont construites avec exactement le même ensemble de mots (ou sac de mots). De même, ces deux phrases comprennent le même ensemble d'opinion (mêmes expressions d'opinion avec les mêmes opérateurs s'appliquant sur celles-ci : { *très(bon), cher* }).

Néanmoins, ces deux phrases ne véhiculent pas la même opinion. La première insiste sur le coût du restaurant, tandis que la seconde met en avant la qualité de la nourriture. Cela est dû à la présence d'une relation de contraste entre les deux parties de la phrase, marquée par le terme "*mais*".

Cet exemple illustre bien les limites d'une approche de type "sac d'opinions" : les opinions d'un texte, même portant sur un sujet commun, et émises par le même émetteur ne sont pas toutes au même niveau. Pour détecter cela, nous avons besoin de connaître la structure discursive du texte.

### 1.2.3.1. Cadre théoriques pour l'étude des phénomènes discursifs

Plusieurs théories ont été développées pour l'étude du discours. Parmi ceux-ci, nous retenons la RST (*Rhetorical Structure Theory* – (Mann and Thompson 1988)), cadre le plus utilisé dans les travaux étudiant l'effet du discours sur l'opinion, et la SDRT (*Segmented Discourse Representation Theory* – (Asher and Lascarides 2003)), l'approche que nous avons choisi d'utiliser.

En RST, les unités de discours manipulées sont des propositions, considérées comme porteuses de l'intention de communication du locuteur. La RST distingue deux catégories de relations de discours entre ces segments, les relations *hypotactic* et *paratactic*. Les relations *hypotactic* relient deux constituants n'étant pas de même importance pour le discours : un *nucleus* et un *satellite*. Les relations *paratactic* relient quant à elles deux segments d'importance égale pour le discours – deux satellites. La RST distingue de nombreux types sémantiques de relations suivant leur effet : on peut grouper celles-ci en deux catégories : les relations de cohérence, liées au contenu sémantique des unités de discours, et les relations intentionnelles, qui sont elles liées à l'intention du locuteur, à ce qu'il souhaite communiquer. La figure 1.2 présente un exemple d'arbre RST, issu de (Mann and Thompson 1988).

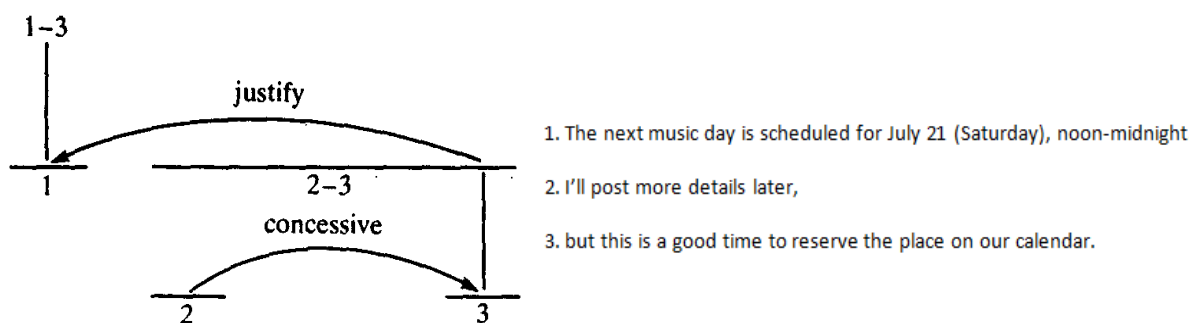


Figure 1.2 : Représentation RST d'un court document (exemple issu de (Mann and Thompson 1988)).

La SDRT (Asher and Lascarides 2003) est quant à elle une théorie discursive basée sur la DRT (Kamp and Reyle 1993). Comme en RST, l'unité minimale du discours définie en SDRT est interne à la phrase : le segment est une clause ou une proposition formant une unité sémantique. Les relations

SDRT sont par contre de nature uniquement sémantiques, et sont donc des relations de cohérence discursive. Ces relations sont également structurées hiérarchiquement, suivant qu'elles sont subordonnantes ou coordonnantes ; ces relations sont ainsi représentables sous la forme d'un graphe de discours. Ce graphe de discours obéit à la règle de la frontière droite : tout nouveau segment devant être ajouté à la structure ne peut être rattaché qu'au dernier élément ajouté à la structure, ou à un de ses subordonnants, direct ou indirect. La RST et la SDRT présente plusieurs différences au niveau de la structure discursive générée : la structure discursive RST est un arbre, tandis que la structure SDRT est un graphe. La SDRT accepte également plusieurs relations entre deux segments.

Une description plus détaillée de la SDRT est présentée au chapitre 2.

### 1.2.3.2. Principales approches discursives pour l'extraction de l'opinion

L'utilisation d'informations discursives pour l'analyse de l'opinion est une piste envisagée depuis plusieurs années, mais étudiée par un nombre restreint de travaux. Nous décrivons dans cette section les travaux de (Pang and Lee 2004; Thomas, Pang, and Lee 2006; Somasundaran, Wiebe, and Ruppenhofer 2008; Somasundaran 2010; Asher, Benamara, and Mathieu 2008; Asher, Benamara, and Mathieu 2009; Taboada, Voll, and Brooke 2008; Heerschop et al. 2011; Polanyi and van den Berg 2011; Trnavac and Taboada 2011; Zhou et al. 2011).

À notre connaissance, les premiers travaux prenant en compte des informations issues de la structure du document pour l'analyse automatique d'opinion sont ceux de (Pang and Lee 2004). Ils partent de l'hypothèse structurelle que les phrases subjectives (et respectivement objectives) ont tendance à apparaître groupées dans un texte (Wiebe 1994; Riloff, Wiebe, and Wilson 2003). Cette information est utilisée dans le cadre d'un détecteur de subjectivité, au niveau de la phrase. Les phrases ainsi déterminées comme objectives sont ensuite éliminées, afin de ne fournir à un classifieur automatique qu'un extrait de chaque document, représentatif de la subjectivité du document. Cette approche permet une légère amélioration des résultats (environ 1 % d'accuracy) pour les classifieurs Bayésien Naïf et SVM.

(Thomas, Pang, and Lee 2006) se sont penchés sur une tâche connexe à l'analyse d'opinion : la classification de discours politiques pour les débats de lois au congrès américain. Les auteurs partent de la constatation que les segments (dans ce contexte, un segment est une portion de discours, prononcé par la même personne et ininterrompue) sont fortement intégrés au sein du débat, et donc en relation discursive avec les interventions précédentes et suivantes. Par conséquent, ils proposent d'enrichir les traits de classification en ajoutant la présence de deux relations macro-discursives : "*Same-speaker constraint*" relie deux segments de discours prononcés par la même personne, et "*Different-speaker agreements*", qui relie deux segments de discours où l'un fait référence explicitement à l'autre, et annonce un accord avec sa position. La première de ces relations est simplement dérivée des données, qui contiennent le nom des intervenants. Les instances de la seconde sont repérées en deux étapes :

- Détection de la référence à un autre intervenant : seules les références explicites avec mention du nom de l'intervenant sont prises en compte

- Décision si en accord ou en désaccord avec l'autre intervenant : pour cela un classifieur SVM est entraîné. Les données d'entraînement sont automatiquement étiquetées en accord/désaccord suivant les votes de chaque intervenant.

L'inclusion de ces informations dans un classifieur SVM donne une amélioration importante des résultats par rapport à la classification des segments considérés isolés : 19 % d'accuracy en plus sur le corpus de développement, et 5 % d'accuracy en plus sur le corpus de test.

(Somasundaran, Wiebe, and Ruppenhofer 2008; Somasundaran 2010) proposent un cadre d'analyse de l'opinion via le repérage de relations discursives entre opinions unitaires. Les auteurs distinguent au niveau local deux types d'opinion, hérités de leurs travaux précédents (Wilson and Wiebe 2005) : les sentiments (*Sentiment*) – évaluations, émotions, et jugements – et les argumentations (*Arguing*) – pour ou contre une entité ou une action. Ces deux catégories sont toutes deux associées à une polarité – positive ou négative – et à un sujet. Dans leur framework, deux opinions peuvent être liées au sein d'une structure discursive, appelée *Opinion Frame*, représentant une relation typée suivant les sujets respectifs des deux opinions : l'*Opinion Frame* peut représenter soit une relation *Same*, soit une relation *Alternative*. La première correspond à des opinions portant sur des sujets référant à la même entité ou propriété. La seconde indique que les sujets sont mutuellement exclusifs (par exemple, deux modèles de télécommandes pour une télévision). L'exemple suivant, extrait de l'article (Somasundaran, Wiebe, and Ruppenhofer 2008), illustre ces deux types de relations ; le premier tableau présente les opinions annotées, le second les relations entre sujets.

C:: ... shapes **should be** *curved*, so round shapes. **Nothing** square-like.

C:: ... So we **shouldn't have too** *square corners* and that kind of things.

B:: Yeah okay. Not *the old box look*.

Opinion repérée (Opinion span)	Cible de l'Opinion (Target span)	Type
should be [O1]	curved [t1]	Arguing / Positive (AP)
Nothing [O2]	square-like [t2]	Arguing / Negative (AN)
shouldn't have [O3]	square corners [t3]	Arguing / Negative (AN)
<b>too</b> [O4]	square corners [t3]	Sentiment / Negative (SN)
<b>Not</b> [O5]	the old box look [t4]	Arguing / Negative (AN)
the old box look [O6]	the old box look [t4]	Sentiment / Negative (SN)

Tableau 1.1 : Exemple : opinions repérées et sujets associés

Couple de sujets	Relation associée
t1 – t2	Alternative
t2 – t3	Same (specification)
t3 – t4	Same (epithet)

Tableau 1.2 : Exemple : relations entre sujets

Ces *frames* sont ensuite classées suivant deux types suivant l'effet produit sur l'opinion : effet de renforcement (*reinforcing frames*), ou de non-renforcement (*non-reinforcing frames*). L'objectif de la reconnaissance et de la classification des *Opinion Frame* est double : l'ajout d'information contextuelle aux structures d'opinion repérées, et la résolution d'opinion ambiguës. Les expérimentations de reconnaissance automatique des *Opinion Frames* ont montré des résultats encourageants, avec près de 50 % de F-score pour un classifieur SVM. Les relations associées aux *Opinion Frames* ne sont pas directement déductibles des relations discursives des théories traditionnelles : par exemple, une relation de Contraste peut correspondre à une relation de renforcement ou de non-renforcement. Ce *framework* n'est donc pas directement compatible avec un *parser* de discours traditionnel.

(Asher, Benamara, and Mathieu 2008; Asher, Benamara, and Mathieu 2009) ont quant à eux proposé, à partir d'une analyse au niveau du segment de discours, une représentation sémantique de l'opinion à l'aide d'une structure de traits et de cinq types de relations rhétoriques de la SDRT. La structure de traits proposée distingue quatre catégories principales d'opinion : Reportage (*Reporting*), Jugement (*Judgment*), Conseil (*Advise*), et Sentiment (*Sentiment*). Ces catégories sont ensuite subdiviser en plusieurs sous-groupes, comprenant également plusieurs sous-catégories. Le tableau 1.3 suivant, issu de (Asher, Benamara, and Mathieu 2008) détaille ces catégories : les séparations pleines distinguent des sous-groupes, les séparations en pointillés distinguent des sous-catégories appartenant au même sous-groupe.

Catégories	Sous-catégories	Exemples
REPORTING	Inform	inform, notify, explain
	Assert	assert, claim, insist
	Tell	say, announce, report
	Remark	comment, observe, remark
	Think	think, reckon, consider
	Guess	presume, suspect, wonder
JUDGMENT	Blame	blame, criticize, condemn
	Praise	praise, agree, approve
	Appreciation	good, shameful, brilliant
ADVISE	Recommend	advise, argue for
	Suggest	suggest, propose
	Hope	wish, hope
SENTIMENT	Anger / Calm Down	irritation, anger
	Astonishment	astound, daze
	Love, Fascinate	fascinate, captivate
	Hate / Disappoint	demoralize, disgust
	Fear	fear, frighten, alarm
	Offense	hurt, chock
	Sadness / Joy	happy, sad
	Bore / Entertain	bore, distraction

Tableau 1.3 : Catégories des opinions détaillées par (Asher, Benamara, and Mathieu 2008)

Les structures de traits proposées comportent une représentation sémantique associée à la catégorie. Ainsi, les expressions de reportage sont associées à une force et à un degré d'engagement ; les autres catégories sont associées à une intensité et à une polarité. Ces structures de traits sont associées à un segment de discours. La Figure 1.3 présente un exemple pour le segment (A) [*Ce film est époustouflant*].

[Catégorie : [ <i>Sentiment</i> : Étonnement]
Modalité : [Polarité : neutre, Force : 2]
Émetteur de l'opinion : rédacteur
Topique : <i>film</i>
Lexicalisation : <i>époustouflant</i> ]

Figure 1.3 : Structure de traits associée au segment (A) (exemple issu de (Asher, Benamara, and Mathieu 2009))

Ces segments de discours sont reliés entre eux par des relations rhétoriques issues de la SDRT ; cinq types de relations sont considérées : *Contraste*, *Correction*, *Support* (groupant les relations d'*Elaboration* et d'*Explication*), *Résultat*, *Continuation*. À ces relations sont ensuite associées certaines règles de remontée de l'opinion au niveau du document. Enfin, afin de valider ce modèle de représentation de l'opinion, les auteurs ont mené une annotation sur 150 documents francophones, et 216 documents anglophones (incluant le corpus MUC6, dont les annotations de discours étaient déjà présentes). Les accords inter-annotateurs obtenus sont bons, entre 73 % et 95 % de kappa suivant le type de document.

(Taboada, Voll, and Brooke 2008), pour leur chaîne de traitement SO-CAL (Semantic Orientation Calculator), ont utilisé l'analyseur RST SPADE (Soricut and Marcu 2003). SO-CAL permet de déterminer, au niveau d'un document, une orientation sémantique incluant polarité et intensité dans une échelle sur 11 points (entre -5 et +5). Afin de déterminer ce score, SO-CAL se base sur un lexique d'adjectifs, mais aussi de noms, verbes et adverbes d'opinion. L'ambiguïté sémantique de certains mots est résolue en prenant une orientation sémantique moyenne sur les sens courants. La chaîne de traitement prend également en compte les intensifieurs, dont l'effet est modélisé comme un facteur multiplicatif, et la négation, dont l'effet est modélisé par un décalage d'une valeur fixe. À cela s'ajoute deux éléments visant à réduire le bruit et ne garder que les éléments pertinents du texte :

- D'une part, une prise en compte du sujet de chaque phrase – si une phrase ne porte pas sur le sujet principal, elle n'est pas prise en compte. La décision de pertinence du sujet d'une phrase est effectuée via classification automatique (arbre de décision, implémenté au sein de la plate-forme Weka)
- D'autre part, les auteurs prennent également en compte le contexte discursif visant à réduire le bruit, et à ne garder du texte que les segments pertinents. Pour cela, après avoir étiqueté chaque segment du texte en *nucleus* ou *satellite*, les auteurs proposent de ne considérer pertinent pour le calcul de l'orientation sémantique que les éléments appartenant à un *nucleus*.

Afin d'expérimenter ces hypothèses, les auteurs proposent plusieurs pondérations pour les phrases acceptées / rejetées suivant chacune des deux heuristiques :

- 1 pour les phrases acceptées par l'une ou l'autre des méthodes, 0 pour les phrases rejetées (pondération (1,0)) ;
- 1,5 pour les phrases acceptées par l'une ou l'autre des méthodes, 0,5 pour les phrases rejetées (pondération (1.5, 0.5)) ;
- 2 pour les phrases acceptées par les deux heuristiques, 1,5 pour les phrases acceptées par l'une des deux heuristiques seulement, et 0,5 pour les phrases rejetées par les deux (pondération (2, 1.5, 0.5)).

Les résultats obtenus sur le corpus de test sont présentés dans le tableau 1.4. La seconde colonne de résultats ajoute à ces éléments un *offset* de décalage, décalant le point de décision entre positif et négatif, permettant de contrebalancer un biais du corpus.

	Sans offset de normalisation	Avec offset de normalisation (0.62)
<b>SO-CAL baseline</b>	70.00 %	72.00 %
<b>Discours (1, 0)</b>	68.44 %	73.78 %
<b>Discours (1.5, 0.5)</b>	71.56 %	80.00 %
<b>Sujets (1, 0)</b>	65.24 %	72.41 %
<b>Sujets (1.5, 0.5)</b>	70.00 %	80.67 %
<b>Discours + Sujets (2, 1.5, 0.5)</b>	70.67 %	78.44 %

Tableau 1.4 : Résultats obtenus par (Taboada, Voll, and Brooke 2008)

Les résultats obtenus montrent, une fois le facteur correctif de normalisation appliqué, une nette amélioration des performances, à la fois avec la prise en compte du discours et avec la prise en compte du sujet de la phrase (jusqu'à +8 %). Cette amélioration, même si elle reste importante (+6 % environ), est toutefois plus faible lorsque les deux éléments sont pris en compte. Ceci montre néanmoins l'intérêt de la prise en compte du discours pour la détermination du score d'opinion au niveau du document.

Poursuivant la même idée, (Heerschop et al. 2011) ont également associé à leur système de classification de documents d'opinion PATHOS un parser RST (SPADE, (Soricut and Marcu 2003)). Afin de classer un document en positif ou négatif, leur système se base sur un score d'opinion au niveau du mot. Ce score est déterminé par lexique, après désambiguïsation syntaxique et sémantique : cette dernière est effectuée de manière non supervisée par calcul de similarité sémantique avec le contexte, suivant la méthode proposée par (Zghal, Aaufaure, and Mustapha 2007). Ce score est ensuite pondéré par rapport à la position du mot dans la structure discursive du texte. Les auteurs testent plusieurs méthodes de pondérations :

- Baseline : pas de pondération
- Simple : partant du principe que l'auteur d'un texte garde les éléments les plus pertinents pour la fin de son argumentation (effet de conclusion), les mots sont pondérés par leur position dans le texte : le premier mot a un poids nul, et le dernier mot un poids de 1.

- Spade I : les expressions appartenant à un nucleus sont pondérés avec un poids de 1, ceux appartenant à un satellite avec un poids de 0 (inspiré de (Taboada, Voll, and Brooke 2008))
- Spade II : les expressions appartenant à un nucleus sont pondérés avec un poids de 1.5, ceux appartenant à un satellite avec un poids de 0.5 (inspiré de (Taboada, Voll, and Brooke 2008))
- Extended Spade : les expressions appartenant à un satellite sont pondérées suivant la relation associée à ce satellite. Cette pondération, entre -2 et +2, peut être négative, notamment dans le cas de contrastes. Le poids associé à chaque relation RST considérée est calculé et optimisé via un algorithme génétique. Le Tableau 1.5 décrit les poids obtenus par relation.

Relation RST	Poids associé
(Nucleus)	0.771
Attribution	0.457
Background	0.017
Cause	-0.271
Condition	0.304
Contrast	-0.660
Elaboration	1.400
Enablement	0.956
Explanation	-0.099

*Tableau 1.5 : Poids associés aux nuclei et, par relation, aux satellites*

Ces cinq méthodes sont ensuite testées sur un corpus de 1000 reviews positives et 1000 reviews négatives (Pang and Lee 2004). Si les méthodes Spade I et Spade II ne semblent pas se montrer plus efficace que les baselines sur ces données, la méthode Extended Spade permet une amélioration de la F-mesure de l'ordre de 4 %, à la fois pour les documents positifs et les documents négatifs.

(Zhou et al. 2011) se sont intéressés aux ambiguïtés de polarité à l'intérieur de la phrase. Pour cela, ils proposent d'une part un schéma discursif basé sur certaines relations de la RST permettant de résoudre ces ambiguïtés, et d'autre part une méthode non supervisée, basée sur les SSRs (Semantic Sequential Representations), afin de reconnaître ces relations. Le schéma discursif est basé sur 5 relations, union de relations RST :

- Contrast : union des relations RST Antithesis, Concession, Otherwise et Contrast
- *Condition* telle que décrite en RST
- Continuation : union des relations RST Continuation et Parallel
- Cause : union des relations RST Evidence, Volitional-Cause, Nonvolitional-Cause, Volitional-result et Nonvolitional-result
- *Purpose* telle que décrite en RST

À ces relations sont associées quatre règles :

- (1) Des segments de discours en contraste doivent être de polarités opposées
- (2) Des segments de discours en relation de Continuation doivent porter la même polarité
- (3) La polarité des segments de discours en relation de Contraste, Condition, Cause, et Purpose est déterminée par la polarité du segment en position de nucleus
- (4) La polarité des segments de discours reliés par une Continuation est fixée par l'un ou l'autre des segments

Afin de reconnaître ces relations de discours, les auteurs proposent une approche non supervisée en deux temps : tout d'abord, une grande quantité d'instances de relations de discours marquées – qui ne sont qu'un sous-ensemble des relations de discours – sont extraites pour chaque type de relation. Ces instances sont ensuite décrites en SSR, et des patrons SSRs communs sont déduits de ces instances. Enfin, les patrons SSRs sont utilisés pour repérer les segments de discours, soit directement, soit en tant que traits pour une méthode d'apprentissage supervisée. Expérimentalement, en moyenne, l'utilisation des SSRs pour reconnaître les relations de discours permet une amélioration importante (+12 %) de la F-mesure par rapport à un repérage baseline par marqueurs. L'utilisation des SSRs en tant que traits SVM permet une augmentation encore plus importante (+17 % de F-mesure par rapport à la baseline). Cette augmentation est principalement due à un bien meilleur rappel, au prix d'une baisse de précision. Sachant que la baseline est extrêmement précise (environ 96 %), cette baisse est relativement limitée pour les SSRs seuls (environ 88 %), mais plus importante pour les SSRs utilisés comme traits SVM (environ 76 %). Enfin, la prise en compte de ces informations permet d'améliorer la reconnaissance de polarité au niveau du document de 1 à 2 %, respectivement pour les SSRs seuls ou en tant que traits SVM.

(Polanyi and Van den Berg 2011) proposent quant à eux une étude de l'opinion à travers le framework discrusif LDM. Dans ce framework, le texte est divisé en unités basiques de discours (BDU - Basic Discourse Units), correspondant typiquement à des clauses verbales. Une structure de traits sémantiques est associée à chacun des segments. Ces traits incluent, entre autres, les informations retenues par les auteurs comme pertinents pour l'analyse de l'opinion (polarité, émetteur, sujet). Les unités de discours sont ensuite reliées entre elles par des relations d'attachement de discours pour former un arbre. Les relations d'attachement de discours sont de trois types : subordonnantes, coordonnantes, ou N-aires. Ces dernières peuvent dénoter des structures conditionnelles si/alors, des structures reliant une question à ses multiples réponses, ou encore des structures de contraste. Afin de déterminer la polarité des segments ambigus ou indéterminés de l'arbre de discours, les auteurs proposent deux règles d'héritage :

- descendant (downward sentiment inheritance) : la polarité d'un constituant dominant est transférée aux constituants dominés ne portant pas de polarité ou à la polarité ambiguë.
- ascendant (upward sentiment inheritance) : la polarité d'une liste coordonnée de constituants dominés est transférée au constituant dominant ne portant pas de polarité ou à la polarité ambiguë.



Enfin, (Trnavac and Taboada 2011) ont examiné comment les relations rhétoriques et les marqueurs lexicaux de modalité non-véridique contribuent à l'expression d'une évaluation dans les commentaires web de films ou de livres. Le cadre de travail choisi est celui de l'Appraisal (Martin 2005), comme l'illustre la Figure 1.4.

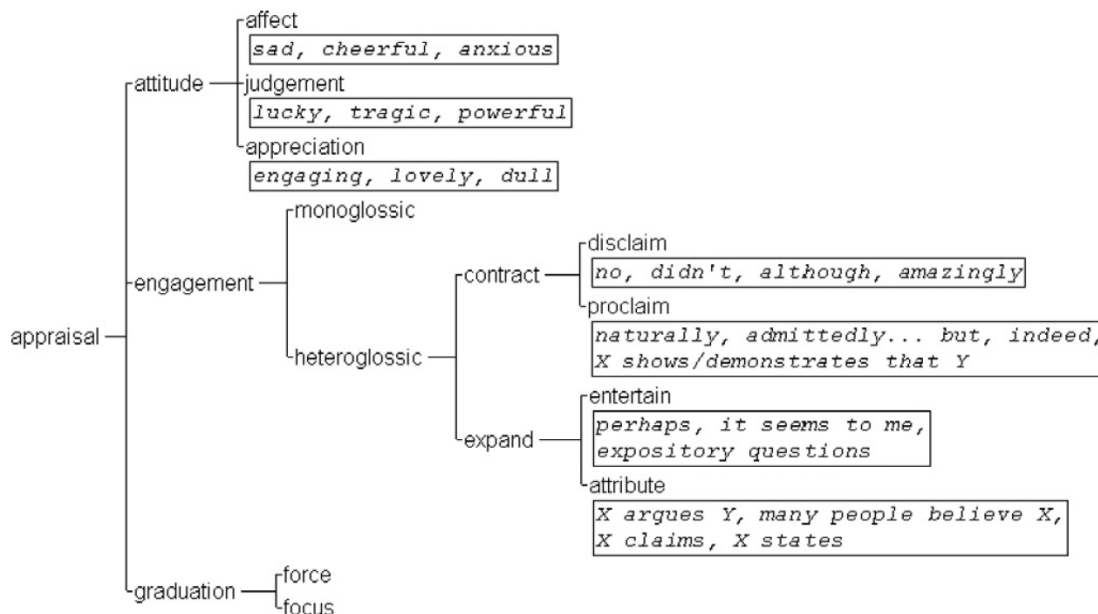


Figure 1.4 : Framework Appraisal (Martin 2005), catégories étudiées par (Trnavac and Taboada 2011)

Les auteures se sont concentrées sur deux principaux types de relations rhétoriques marquées lexicalement, la *Condition* (*conditional*) et la *Concession* (*concessive*), telles que définies par la RST. Pour cela, elles ont extrait automatiquement, via marqueurs de discours, des instances de ces relations apparaissant dans un contexte de non-véridicité. Les auteures prévoient un effet d'atténuation de l'évaluation pour la Condition, et un effet de renversement de la polarité pour la Concession. Après étude sur corpus (sur une partie du Simon Fraser University Corpus<sup>11</sup>, constituée d'environ 100 commentaires web de films et de livres, équilibrés entre positifs et négatifs), les résultats obtenus montrent que chacune des deux relations tend à renverser la polarité. De plus, associées à des marqueurs modaux de non-véridicité, ces relations ont tendance à affecter l'engagement du locuteur associé à l'évaluation.

### 1.3. Bilan et positionnement de nos travaux

L'objectif de la présente thèse est de proposer une chaîne de traitement complète de l'opinion évaluative, permettant de déterminer finement celle-ci au niveau du document, tout en conservant l'information au niveau local. Cette dernière peut en effet illustrer l'opinion globale (notion de « phrase support »), ou encore permettre de créer des liens sémantiques entre documents, suivant par exemples des opinions similaires sur un même sujet (par exemple, deux restaurants dont les critiques vantent les mérites de leur cassoulet). Pour cela, nous nous

<sup>11</sup> [http://www.sfu.ca/~mtaboada/research/SFU\\_Review\\_Corpus.html](http://www.sfu.ca/~mtaboada/research/SFU_Review_Corpus.html)

appuyons sur la structure discursive pour remonter l'information d'opinion du niveau local au niveau du document.

Pour ce faire, nous avons choisi la SDRT comme cadre de travail pour le discours. Ce choix présente plusieurs avantages dans le contexte de l'analyse d'opinion. Tout d'abord, la segmentation en SDRT se fait à un grain plus fin que dans le modèle de (Wolf and Gibson 2006), ou que dans les schémas d'annotation proposé pour la RST (Mann and Thompson 1988; Carlson, Marcu, and Okurowski 2003) ou celui associé au Penn Discourse TreeBank (*PDTB*, (Miltsakaki et al. 2004)). En effet, en RST, les segments de discours sont typiquement des phrases simples, ou des clauses verbales dans le cas de phrases complexes. La RST n'autorise pas les segments imbriqués, segments que la SDRT autorise et qui peuvent être pertinents afin d'isoler l'opinion. Le fait d'isoler au maximum l'opinion au sein de segments séparés est crucial : en effet, cela permet de projeter sur la structure discursive les opinions du texte, qui sont ainsi reliées par les relations repérées.

### Exemples :

- (1) [Ce film, [dont les acteurs jouent à merveille,]<sub>a</sub> n'a pas remporté un franc succès aux Oscars.]<sub>b</sub>
- (2) [Ce film est joyeux,]<sub>a</sub> [mais d'un niais !]<sub>b</sub>

Dans l'exemple 1, le segment imbriqué 1.a porte une opinion qui n'a rien à voir avec celle du segment l'englobant. Dans l'exemple 2, le segment 2.b, dénué de verbe, est en contraste marqué avec la première partie de la phrase. Dans les deux cas, la finesse de découpage est nécessaire afin d'isoler l'opinion au sein d'un segment.

D'autre part, la plupart des théories de représentation du discours définissent des structures hiérarchiques via la construction de segments complexes de façon récursive. En RST, un document est représenté par un arbre orienté où les segments adjacents sont regroupés. Les annotations du Penn Discourse TreeBank se basent également sur l'adjacence des segments, mais, au contraire de la RST, juste pour les relations de discours et les marqueurs explicites et implicites permettant la détection de ces relations. À l'inverse de ces deux théories, le modèle GraphBank de (Wolf and Gibson 2006) et la SDRT vont au-delà des segments de discours adjacents, et autorisent la création de graphes de discours capturant des phénomènes discursifs complexes, tels que, par exemple les attachements de segments à longue-distance, ou les références croisées. La structure discursive résultante, représentant de manière plus adéquate la structure réelle du discours, est alors un graphe, et non nécessairement un arbre (Danlos 2008).

Enfin, la SDRT vise à modéliser l'interface entre Sémantique et Pragmatique en utilisant des relations discursives décrivant le rôle rhétorique joué en contexte par chaque élément, sur la base du seul contenu vériconditionnel. Ainsi, la taxinomie des relations de la SDRT est plus réduite que celle de la RST, ce qui allège aussi bien la reconnaissance que les traitements effectués

Concernant les différents modules de notre chaîne de traitement, le repérage préalable des opinions au niveau local est effectué par lexique, dans la lignée des travaux effectués par (Dubreil et al. 2008; Wilson and Wiebe 2005). Ceci nous permet d'isoler des expressions d'opinion unitaires minimales, telles que définies au chapitre 2. L'étape suivante consiste à étudier l'influence du

contexte local sur les opinions ainsi repérées, en étudiant notamment l'effet des opérateurs sur l'opinion. Nos travaux dans ce domaine sont proches de ceux de (Taboada et al. 2011; Trnavac and Taboada 2011). Ceci nous permet de définir une opinion à un niveau intermédiaire, le niveau du segment de discours : ce niveau a notamment été étudié par (Zirn et al. 2011). Enfin, concernant la remontée de l'opinion au niveau d'un document via le discours, nous avons adopté une approche basée sur l'association de règles de remontée de l'opinion aux relations de discours. En cela, nos travaux sont dans la lignée de (Asher, Benamara, and Mathieu 2008; Asher, Benamara, and Mathieu 2009), et sont proches de (Zhou et al. 2011). Néanmoins, contrairement à ces travaux, nous proposons un jeu de règles plus important, permettant de prendre en compte finement l'effet d'un grand nombre de relations SDRT dans la constitution de l'opinion globale d'un document.

L'architecture globale de la chaîne est décrite dans le chapitre 3. Le chapitre 2 traite quant à lui de la notion d'opinion, et propose un modèle pour son annotation ainsi que celle de son contexte discursif, nous permettant ainsi d'établir une base pour mesurer l'impact du discours sur l'opinion.



## 2. Modèle de l'opinion et annotations manuelles

---

### 2.1. Introduction : motivations et contexte scientifique

Nous avons pu voir précédemment (cf. chapitre 1) que nous nous intéressons dans le cadre de cette thèse à l'opinion évaluative. Afin de formaliser ce que nous entendons par cette notion, et d'étudier comment les opinions sont exprimées dans le discours, nous proposons dans ce chapitre un modèle d'annotation de l'opinion, à la fois au niveau local et au niveau d'un document. Ce modèle est ensuite mis en pratique dans une campagne d'annotation. Cette campagne d'annotation est effectuée dans le cadre du projet CASOAR<sup>12</sup> (Chaîne d'Analyse Sémantique des Opinions, Appréciations et Représentations) : ce projet, dont les partenaires sont l'équipe Melodi du laboratoire IRIT et la société Synapse Développement, bénéficie du financement RAPID<sup>13</sup> de la DGA (Direction Générale de l'Armement). Les travaux décrits dans ce chapitre ont donné lieu à la publication (Benamara et al. à paraître).

#### 2.1.1. Principales campagnes d'annotations de l'opinion

Plusieurs projets ont menés des campagnes d'annotations de l'opinion. Pour l'anglais, le projet de référence est sans nul doute le projet MPQA (Multi-Perspective Question Answering), mené par (Wiebe, Wilson, and Cardie 2005; Wilson, Wiebe, and Hwa 2006). Basé sur le modèle linguistique de Quirk (Quirk et al. 1985), le modèle d'annotation MPQA prévoit une annotation au niveau de l'expression d'opinion, caractérisée par une polarité, une intensité sur 4 niveaux (low, medium, high, extreme), et un caractère explicite (i.e. porteuse d'opinion hors contexte) ou implicite (i.e. non porteuse d'opinion hors contexte) de l'expression d'opinion. Le corpus annoté est composé principalement d'articles de presse. L'exemple ci-dessous est une structure de traits d'annotation MPQA.

**Exemple :** *"The US fears a spill-over," said Xirao-Nima, a professor of foreign affairs at the Central University for Nationalities".*

Explicitations directes d'états privés :

Text anchor: fears

Source: <writer,Xirao-Nima,U.S.>

Intensity: medium

Expression intensity: medium

Attitude type: negative

---

<sup>12</sup> <http://projetcasoar.com/>

<sup>13</sup> <http://www.ixarm.com/Projets-d-innovation-duale-RAPID>

Pour le français, deux projets ont récemment porté sur l'annotation d'opinion. Tout d'abord, le projet Blogoscopie (Daille et al. 2011) propose une annotation d'opinion de blogs. Leur annotation au niveau local est basée sur cinq catégories sémantiques :

- *Opinion* : expression portant une marque de certitude. Cinq valeurs de certitude sont distinguées, de la plus forte à la plus faible : conviction, certitude forte, certitude moyenne, certitude faible, et pressentiment.
- *Appréciation* : opinion évaluative sur un sujet, correspondant à la notion principale sur laquelle nous nous focalisons. Le schéma d'annotation Blogoscopie propose six sous-catégories : explicite positif ou négatif, explicite-exclamatif positif ou négatif, et implicite positif ou négatif. Le caractère explicite ou implicite désigne ici l'attribution explicite ou implicite de l'opinion à son émetteur : cette information prend tout son sens dans le domaine étudié par le projet, à savoir les blogs.
- *Accord/Désaccord* : opinions portant sur une autre opinion. Quatre catégories sont ici retenues : accord total, accord approximatif, rectification (ex : structure en « *certes, mais...* »), désaccord.
- *Jugement* : jugement moral sur une entité.
- *Acceptation/Refus* : Consentement ou non consentement à quelque chose, sans pour autant entraîner l'accord.

Le modèle d'annotation du projet Blogoscopie est principalement centré sur la notion d'émetteur : qu'est-ce que pense l'auteur du blog sur divers sujets, et comment ses opinions interagissent-elles avec les opinions de personnes extérieures. Notre modèle, à l'inverse, se focalise plus sur la notion de sujet de l'opinion et d'évaluation.

Le projet Doxa (Paroubek, Pak, and Mostefa 2010) propose quant à lui une annotation de l'opinion sur trois niveaux : au niveau du document (niveau *macro*), avec l'annotation d'une polarité et d'une intensité sur 3 niveaux, au niveau d'un extrait de texte de 1000 caractères (niveau *meso*), et au niveau d'une expression d'opinion (niveau *micro*), accompagné d'une catégorisation sémantique en 17 catégories, inspirées de (Mathieu 2000; Mathieu 2006). Les annotations du projet Doxa comprennent également le sujet de l'opinion, classé suivant une ontologie du domaine. Deux types de documents ont été annotés dans le cadre de ce projet : des commentaires web portant sur les jeux vidéo (issus du site jeuxvideo.com<sup>14</sup>), et des retours de satisfaction clients de l'opérateur français d'électricité EDF.

Concernant le niveau de l'expression d'opinion, nos travaux reprennent et affinent la plupart des éléments de ces projets : annotation d'une polarité et d'une intensité sur 3 niveaux, et catégorisation sémantique fine. Nos catégories sémantiques, présentées ci-après, sont également dans la lignée des travaux de (Asher, Benamara, and Mathieu 2008; Asher, Benamara, and Mathieu 2009). Notre modèle inclus de plus une analyse de l'opinion sur trois niveaux de granularité :

---

<sup>14</sup> <http://www.jeuxvideo.com/>

- Au niveau de l'expression d'opinion
- Au niveau du segment de discours
- Au niveau du document complet, caractérisé par sa structure discursive

À notre connaissance, il s'agit du seul modèle incluant ces trois niveaux. Concernant application du modèle, la campagne d'annotations inclus deux genres de corpus différents : commentaires web, et réactions d'articles de presse.

### 2.1.2. Déroulement du chapitre

Le chapitre se divise en 5 sous-sections. La section 2.2 traite du modèle local de l'opinion, au niveau d'une expression véhiculant une opinion unitaire. La section 2.3 expose le modèle de l'opinion au niveau du segment de discours. La section 2.4 introduit le modèle au niveau d'un document et de sa structure discursive. La section 2.5 présente un exemple d'annotation, ainsi que le logiciel utilisé pour les annotations. Enfin, la section 2.6 présente le déroulement et les résultats de la campagne d'annotation.

## 2.2. Modèle local de l'opinion

Dans cette section, nous présentons le modèle retenu pour caractériser une opinion au niveau local, c'est-à-dire au niveau d'une portion de phrase portant la subjectivité.

### 2.2.1. Opinion locale : une définition

Au niveau local, nous retenons les éléments constituant une opinion suivant :

- Une expression d'opinion véhiculant une opinion *a priori*, c'est-à-dire indépendamment de son contexte.
- Un sujet sur lequel porte l'opinion
- Un émetteur, qui émet l'opinion
- Un ou plusieurs opérateurs, qui viennent altérer l'opinion *a priori*, portée par l'expression.

Formellement, une opinion locale ou unitaire est donc un quadruplet  $Op = (w, t_w, h_w, ops_w)$ , avec :

- $w$  l'expression d'opinion
- $t_w$  le sujet de l'opinion (en anglais, *Topic*)
- $h_w$  l'émetteur de l'opinion (en anglais, *Holder*)
- $ops_w$  la liste des opérateurs s'appliquant sur l'opinion

La section suivante (2.2.2) présente l'expression d'opinion, et les attributs qui lui sont attachés. La section 2.2.3 présente les éléments du contexte local, à savoir le sujet de l'opinion, l'émetteur de l'opinion, et les opérateurs portant sur l'opinion.

## 2.2.2. L'expression d'opinion et ses caractéristiques

Étant donné une portion de texte dont on reconnaît qu'elle porte explicitement une opinion, nous définissons l'expression d'opinion comme étant la portion du texte minimale porteuse d'opinion.

**Exemples :**

(3) *Ce restaurant est bon.*

"*bon*" porte ici l'opinion, et constitue donc l'expression d'opinion.

Celle-ci doit être minimale :

(4) *Ce restaurant est très bon.*

"*bon*" est ici l'unité de texte minimale à porter l'opinion : "*très*" altère l'opinion, mais ne porte pas d'opinion en lui-même – il y a toujours une opinion si celui-ci est retiré de la phrase. L'expression d'opinion est donc ici aussi "*bon*".

(5) *La viande est très bonne, très tendre.*

Dans la portion de phrase (5), on peut isoler deux opinions distinctes : une opinion positive sur le goût de la viande ("*bonne*"), et une opinion plus spécifique portant sur la tendresse de la viande ("*tendre*"). Là encore, les deux adverbes "*très*" altèrent l'opinion, mais ne sont pas intrinsèquement porteurs d'opinion. Le caractère explicite de l'opinion annotée permet de limiter l'ambiguïté sur les bornes de l'expression d'opinion.

Chaque expression d'opinion possède plusieurs attributs sémantiques associés. Nous avons divisé les expressions d'opinion en quatre catégories sémantiques distinctes, présentées en section 1.2.2.3. Suivant la catégorie de l'expression d'opinion, celle-ci peut également être associée à une polarité et une intensité : ces traits seront présentés respectivement en section 1.2.2.4 et en section 1.2.2.5.

### 2.2.2.1. Catégorisation sémantique

Chaque expression d'opinion peut appartenir à l'une des catégories suivantes, décrites par (Asher, Benamara, and Mathieu 2009) : reportage, jugement, sentiment-appréciation, et conseil.

#### *Reportage*

Les expressions de reportage permettent de relater ou d'introduire les opinions des autres ou les siennes propres.

**Exemple :**

(6) J'affirme que ce restaurant est bon.

La catégorie Reportage possède les sous-catégories suivantes :

- Informer – soutenir : expressions introduisant une information que l'auteur estime vraie et établie. Par exemple : *informer, soutenir, prétendre*.



- Dire – remarquer : expressions de reportage n'indiquant rien sur le point de vue ou l'engagement de l'auteur au sujet de l'opinion émise. Par exemple : *dire, annoncer*.
- Penser – supposer : expressions introduisant une opinion avec un certain degré de subjectivité. L'auteur exprime sa pensée sur quelque chose sans être sûr de sa véracité. Par exemple : *penser, supposer*.

### Jugement

Les expressions de jugement expriment des évaluations normatives d'objets et d'actions, à l'intérieur desquelles on peut distinguer les sous-catégories suivantes :

- Blâmer – louer : expressions indiquant un jugement sur la responsabilité ou le comportement de quelqu'un. Par exemple : *approuver, critiquer*. Ce type est associé à une valeur de polarité, mais pas d'intensité.
- Évaluation : jugements exprimant le caractère intrinsèquement positif ou négatif de quelqu'un ou quelque chose. Par exemple : "*C'est un pur chef d'œuvre*". Ces expressions possèdent une polarité et une intensité.

### Sentiment – appréciation

Les expressions de sentiment – appréciation expriment un sentiment ou une émotion ressentie par une personne. Ces expressions possèdent une polarité, en général définie par leur sous-catégorie, et une valeur d'intensité.

#### Exemple :

(7) J'ai adoré le film.

Parmi ces expressions, nous distinguons plusieurs catégories. Celles-ci peuvent se regrouper suivant leur polarité intrinsèque.

Les sentiments positifs sont divisés entre les classes suivantes :

- Apaisement : par exemple, *apaisant, rassurant*.
- Divertissement – joie : par exemple, *satisfait, réjouissant*.
- Intérêt – passion : par exemple, *éblouissant, séduisant*.

Les sentiments négatifs sont divisés entre les classes suivantes :

- Colère – irritation : par exemple, *agaçant, énervant*.
- Ennui : par exemple, *embêtant, ennuyeux*.
- Offense : par exemple, *révoltant, scandaleux*.
- Tristesse – souffrance : par exemple, *douloureux, nostalgique, tragique*.
- Peur : par exemple, *inquiet, angoissant*.
- Haine : par exemple, *haineux*.
- Déception : par exemple, *décevant, déprimant*.

Les sentiments neutres, dont la polarité dépend du contexte. Par exemple :

(8) Ce film a bouleversé Marie.

(9) La vue de tous ces morts a bouleversé Marie.

Dans la phrase (6), le verbe bouleverser indique une opinion positive sur le film. On s'attend à ce que Marie en parle comme d'un chef-d'œuvre. Par contre, dans la phrase (7), bouleverser revêt une polarité négative, et évoque plutôt un traumatisme.

Ces termes et expressions sont divisés entre les classes suivantes :

- Étonnement : par exemple, *sidérant, stupéfiant*.
- Émotion : par exemple, *émouvant, touchant*.

### **Conseil**

Ces expressions poussent à faire ou à penser quelque chose. Celui qui émet l'opinion se place ici en position d'autorité par rapport au lecteur, et renforce son avis avec d'autres arguments pour le convaincre.

Nous distinguons les catégories de conseil suivantes :

- Recommander : expressions qui véhiculent une opinion, bonne ou mauvaise, et qui essayent de convaincre avec force
- Suggérer : expressions utilisées par l'émetteur d'opinion pour suggérer ou spéculer sans être absolument certain
- Espérer : expressions émettant le sentiment général que certains désirs seront satisfaits. Par exemple : "Espérons que ce film rare touchera un vrai et grand public."

### **Tableau récapitulatif**

Le tableau suivant reprend l'ensemble des catégories sémantiques d'opinion que nous distinguons dans ce modèle.

Catégories	Sous-catégories
Reportage	Informer
	Soutenir
	Dire
	Remarquer
	Penser
	Supposer
Jugement	Blâmer
	Louer
	Évaluation
Sentiment	Irritation – colère
	Apaisement
	Étonnement
	Intérêt – passion
	Haine

	Déception
	Peur
	Offense
	Tristesse – Souffrance
	Ennui
	Divertissement – Joie
	Émotion
Conseil	Recommander
	Suggérer
	Espérer

Tableau 2.1 : Classes sémantiques d'opinion

### 2.2.2.2. Polarité a priori d'une expression d'opinion

Une expression d'opinion possède une polarité, qui peut être soit positive, soit négative, soit neutre. La valeur neutre correspond à une opinion de polarité ambiguë, qui sera éventuellement désambiguïsée par le contexte, notamment discursif. Cette polarité est dite *a priori* ou *hors contexte* : en effet, en contexte, elle peut être altérée, par exemple par un opérateur de négation.

Le Tableau 2.2 présente quelques exemples pour chaque valeur de polarité.

polarité	Exemples	Expression d'opinion
positive	Ce restaurant est <u>bon</u> .	bon
	Ce restaurant est vraiment <u>excellent</u>	excellent
négative	Ce restaurant est <u>mauvais</u> .	mauvais
	Ce restaurant est très <u>moyen</u> .	moyen (emploi négatif)
neutre	Ce restaurant est <u>étonnant</u> .	étonnant

Tableau 2.2 : Exemples illustrant les valeurs de polarité

La phrase :

(10) "Ce restaurant est étonnant."

peut-être interprétée comme positive dans le contexte suivant :

(10a) "Ce restaurant est étonnant. Je n'avais jamais mangé de sushis aussi bons."

Mais peut aussi être interprétée comme négative dans le contexte :

(10b) "Ce restaurant est étonnant. Je ne pensais pas que l'on pouvait manger aussi mal."

Le neutre ne correspond donc pas à une phrase ou une portion de phrase objective, mais bien à une opinion, dont l'interprétation est liée au contexte.

La polarité d'une expression d'opinion est dite "a priori". En effet, il ne s'agit pas nécessairement de la polarité finale en contexte, dans la mesure où un ou plusieurs opérateurs peuvent venir altérer celle-ci.

**Exemple :**

(11) *Ce restaurant n'est pas bon.*

Dans la phrase (11), la polarité a priori de l'expression d'opinion "*bon*" est positive, alors que la polarité en contexte, tenant compte de la négation "*ne ... pas*", est négative. La notion d'opinion en contexte sera détaillée ultérieurement au niveau du segment discursif (cf. section 2.3).

### 2.2.2.3. Intensité a priori d'une expression d'opinion

En plus de la polarité, une expression d'opinion possède également une intensité.

**Exemple :** Considérons les phrases :

(12) Ce restaurant est bon.

(13) Ce restaurant est excellent.

L'expression d'opinion "*excellent*" de la phrase (12) est plus intense que l'expression d'opinion "*bon*" de la phrase (13)

En pratique, nous adopterons une échelle d'intensité discrète sur trois points pour caractériser l'intensité d'une opinion. Le Tableau 2.3 présente cette échelle.

Valeur discrète	Libellé sémantique	Exemple d'expression positive	Exemple d'expression négative	Exemple d'expression neutre
1	intensité basse	correct	décevant	intrigant
2	intensité moyenne	brillant	nul	surprenant
3	intensité forte	excellent	nullissime	bouleversant

Tableau 2.3 : Valeurs d'intensité pour une expression d'opinion

Ces trois valeurs d'intensité nous ont paru suffisantes pour effectuer une analyse fine de l'opinion tout en maintenant des classes d'intensités aisément distinguables pour un juge humain.

Comme pour la polarité, l'intensité propre à l'expression d'opinion est dite *a priori* par opposition à l'intensité en contexte, qui tient compte notamment des opérateurs venant s'appliquer sur l'expression d'opinion.

**Exemple :**

(14) Ce restaurant est très correct.

L'expression d'opinion "*correct*" possède une intensité a priori standard/faible (+1). Néanmoins, dans le contexte de la phrase (14), l'adverbe "*très*" vient altérer cette intensité : l'intensité en contexte est donc plus forte (sur notre échelle, +2).

### 2.2.3. Sujet de l'opinion

Le *sujet* est l'entité (la personne, l'objet, voire le concept) sur laquelle porte l'opinion.

**Exemple :**

(15) Ce restaurant est bon.

Le sujet de l'expression d'opinion "*bon*" est ici "*Ce restaurant*".

Dans le contexte d'un document et d'une application donnés, tous les sujets n'ont pas la même importance. Par conséquent, nous dégageons trois types de sujets :

- Le sujet principal (*main\_topic*)
- Les sujets partiels (*part-of\_topic*)
- Les sujets autres (*other\_topic*)

On peut en particulier dégager un *sujet principal (main\_topic)*, soit parce qu'il est le sujet principal du document (par exemple, dans le cas d'un commentaire web sur un restaurant, le restaurant en question est le sujet principal), soit parce qu'il s'agit d'un sujet sur lequel l'utilisateur de l'application souhaite se focaliser. Dans l'exemple (16), donné plus bas, et dans le contexte d'une critique du restaurant *Sushi House*, la première opinion porte sur le sujet principal, explicitement mentionné par le nom du restaurant.

Toutes les opinions portant sur un sujet donné ne sont pas non plus au même niveau : certaines concernent le sujet dans sa globalité, et d'autres sont focalisées sur un constituant précis du sujet. Ces derniers seront appelés *sujets partiels du sujet principal*, ou *part-of\_topic*. Par souci de concision, nous appellerons dans la suite *Sujet Partiel* les sujets partiels concernant le sujet principal. Dans l'exemple (16), les *plats* et le *service* sont deux exemples de sujets partiels.

Certaines opinions peuvent également concerner un sujet distinct du sujet principal, mais en rapport avec celui-ci. Deux cas peuvent alors se présenter : le sujet peut englober le sujet principal (par exemple, parler des restaurants asiatiques en général englobe le restaurant chinois dont on fait la critique), ou simplement être en rapport avec une caractéristique de celui-ci (par exemple, parler de la qualité du service dans une ville en particulier englobe le service du restaurant dont on fait la critique). Dans le contexte de l'extraction d'opinion sur un sujet donné, les composantes du sujet sans rapport avec le sujet principal ne sont pas pertinentes. Aussi, nous pouvons assimiler ce type de sujet soit au sujet principal pour le premier cas, soit à un sujet partiel pour le second cas. L'exemple (17) présente deux instances de ce phénomène de sujets proches.

Enfin, certaines opinions concernent un sujet qui n'a rien à voir avec le sujet principal. Nous appellerons ce type de sujet *Sujets Autre (other\_topic)*.

**Exemple :** Extrait d'un commentaire web sur le restaurant "*Sushi House*". Les expressions d'opinion sont soulignées, les sujets associés sont en gras.

(16) **Sushi House** est un excellent restaurant japonais. **Les plats** sont délicieux. **Le service** est sympathique, pas désagréable comme dans **certains autres restaurants** dont je tairais le nom.

Expression d'opinion	Sujet de l'opinion	Type du sujet
excellent	Sushi House	Sujet principal
délicieux	(Les) plats	Sujet partiel
sympathique	(Le) service	Sujet partiel
désagréable	certains autres restaurants	Autre sujet

Tableau 2.4 : Sujets associés aux différentes opinions présentes dans l'exemple (14)

**Exemple** : Dans le contexte d'une critique sur un autre restaurant :

(17) **Ce restaurant** est excellent. De toute façon, **les restaurants de cette rue** sont tous bons. **L'ambiance du quartier** est également sympathique.

Expression d'opinion	Sujet de l'opinion	Type du sujet
excellent	Ce restaurant	Sujet principal
bons	Les restaurants de cette rue	Sujet principal
sympathique	L'ambiance du quartier	Sujet partiel

Tableau 2.5 : Sujets associés aux différentes opinions présentes dans l'exemple (15)

### Sujet implicite

Lorsque le document possède un sujet principal, celui-ci n'est pas toujours lexicalisé. Par convention, nous considérons que dans le cas où le sujet n'est pas explicité, l'opinion est toujours liée au sujet principal.

**Exemple** : Dans le cas d'un commentaire web sur un film :

(18) Époustouflant ! Magistral ! Un chef-d'œuvre !

Les trois opinions portent toutes sur le sujet principal, c'est-à-dire le film.

### 2.2.4. Émetteur de l'opinion

L'émetteur de l'opinion est la personne ou le groupe de personnes à l'origine de l'opinion. Cette notion est différente de celle de rédacteur ou locuteur : ce dernier peut tout à fait rapporter une opinion émise par une autre personne.

**Exemple** :

(19) **Pierre** pense que le Stade Toulousain a fait un beau match.

Pierre est l'émetteur de l'opinion "beau". Dans ce cas, on peut raisonnablement supposer que Pierre n'est par contre le rédacteur de la phrase.

De même que pour le sujet, on distingue parmi les émetteurs :

- L'émetteur principal (main\_holder), associé au document (i.e. le locuteur)
- Les autres émetteurs (other\_holder).

L'émetteur identifié de l'opinion peut également être un groupe de personnes incluant l'émetteur principal, comme dans l'exemple (20). Dans ce cas, comme pour la notion de sujet, on assimilera l'émetteur à l'émetteur principal.

**Exemple :**

(20) Nous avons regardé le Rugby hier. **Pierre** pense que le stade Toulousain a fait un beau match. En tout cas, sa télévision, neuve, est magnifique.

Expression d'opinion	Sujet de l'opinion	Émetteur	Type d'émetteur
beau	match	Pierre	Autre émetteur
magnifique	la télévision (de Pierre)	Implicite (locuteur)	Émetteur principal

Tableau 2.6 : Émetteurs associés aux différentes opinions présentes dans l'exemple (20)

(21) **On** a tous adoré ce restaurant.

Expression d'opinion	Sujet de l'opinion	Émetteur	Type d'émetteur
adoré	ce restaurant	On	Émetteur principal (assimilable)

Tableau 2.7 : Émetteurs associés aux différentes opinions présentes dans l'exemple (21)

### Émetteur implicite

L'émetteur n'est pas toujours explicité lorsqu'il s'agit du locuteur, comme nous pouvons le voir dans la troisième phrase de l'exemple (20).

## 2.2.5. Opérateurs portant sur l'opinion

Nous avons vu dans la section précédente que certains termes ou expressions ont un impact sur les caractéristiques de l'opinion. Nous distinguons dans notre modèle d'annotation les opérateurs suivants : opérateurs de négation, d'intensité, de modalité, et de restriction. Cette section détaille chacune de ces catégories.

### Opérateurs de négation

Les opérateurs de négation ont pour effet principal de renverser la polarité d'une expression d'opinion. La négation peut être exprimée via des marqueurs de différentes catégories grammaticales :

- Des adjectifs. Par exemple, *aucun*.
- Des pronoms. Par exemple, *personne* ("Personne n'a applaudi le dernier groupe de la soirée."), *rien* ("Sur la carte, rien n'est alléchant.").
- Des adverbes. Par exemple, *ne ... pas* ("Les sushis n'étaient pas frais"), *non* ("Un non-bonheur absolu.")
- Des prépositions. Par exemple, *sans* ("*Les plats sont sans goût.*").
- Des conjonctions. Par exemple, *ni ... ni* ("Les plats n'ont ni goût, ni saveur.")

### Opérateurs d'intensité

Les opérateurs d'intensité ont pour effet principal d'augmenter (*Intensifier+*) ou d'atténuer (*Intensifier-*) l'intensité d'une opinion. Ces opérateurs sont principalement des adverbes de degré, qui apportent une indication de quantité aux expressions sur lesquelles ils portent.

Nous distinguons les deux sous-types suivants :

- Intensifier+ : ces opérateurs augmentent l'intensité. Par exemple, très (l'intensité de "Ce restaurant est très bon." est plus forte que celle de "Ce restaurant est bon.")
- Intensifier- : ces opérateurs diminuent l'intensité. Par exemple, plutôt (l'intensité de "Ce restaurant est plutôt bon." est plus faible que celle de "Ce restaurant est bon.")

### Opérateurs de modalité

Une forme modale peut être employée en tant qu'opérateur pour nuancer l'opinion. Un modal modifie ainsi l'énoncé d'une proposition en le présentant comme obligatoire, possible, souhaitable, ou encore vrai de facto. De ce fait, un opérateur de modalité a une action sur le degré de certitude ou sur l'intensité d'une opinion.

**Exemple** : dans la phrase :

(22) Il est nécessaire que ce restaurant soit bon.

La proposition "*ce restaurant soit bon*" est altérée par l'opérateur de nécessité.

Syntaxiquement, les opérateurs de modalité peuvent être :

- Des adverbiaux. Par exemple, probablement ("Ce restaurant est probablement bon."), nécessairement ("Ce restaurant est nécessairement bon."), peut-être, il est possible que...
- Des temps verbaux. Par exemple, le conditionnel ("*Ce restaurant serait bon.*").
- Des auxiliaires de modalité. Par exemple, pouvoir ("Ce restaurant peut être bon."), devoir ("Ce restaurant doit être bon.).
- Des verbes d'attitude propositionnelle. Par exemple, *douter, croire, savoir.*

### Opérateurs de restriction

Ces opérateurs induisent une restriction sur le contexte de validité de l'expression d'opinion. Les effets de ces opérateurs sont divers. Nous pouvons, entre autres, citer :

- Effet de nuance temporelle. Souvent, il s'agit d'adverbes de temps, comme par exemple *souvent* ("*Les plats sont souvent excellents.*") ou *parfois* ("*Parfois, les plats sont bons.*"). D'autres constructions syntaxiques sont néanmoins envisageables : par exemple, "*Bien le soir.*"
- Effet de nuance de quantification. Par exemple, tous ("Tous les cocktails sont bons."), ou uniquement ("Ici, uniquement les sushis sont bons.")
- Effet de restriction sur le destinataire du message. Par exemple : "Restaurant très bien pour les enfants.", "Je le recommande aux groupes d'amis."



### 2.2.6. Schéma récapitulatif du modèle d'annotation au niveau local

Schématiquement, notre modèle de l'opinion au niveau local peut se représenter comme suit.

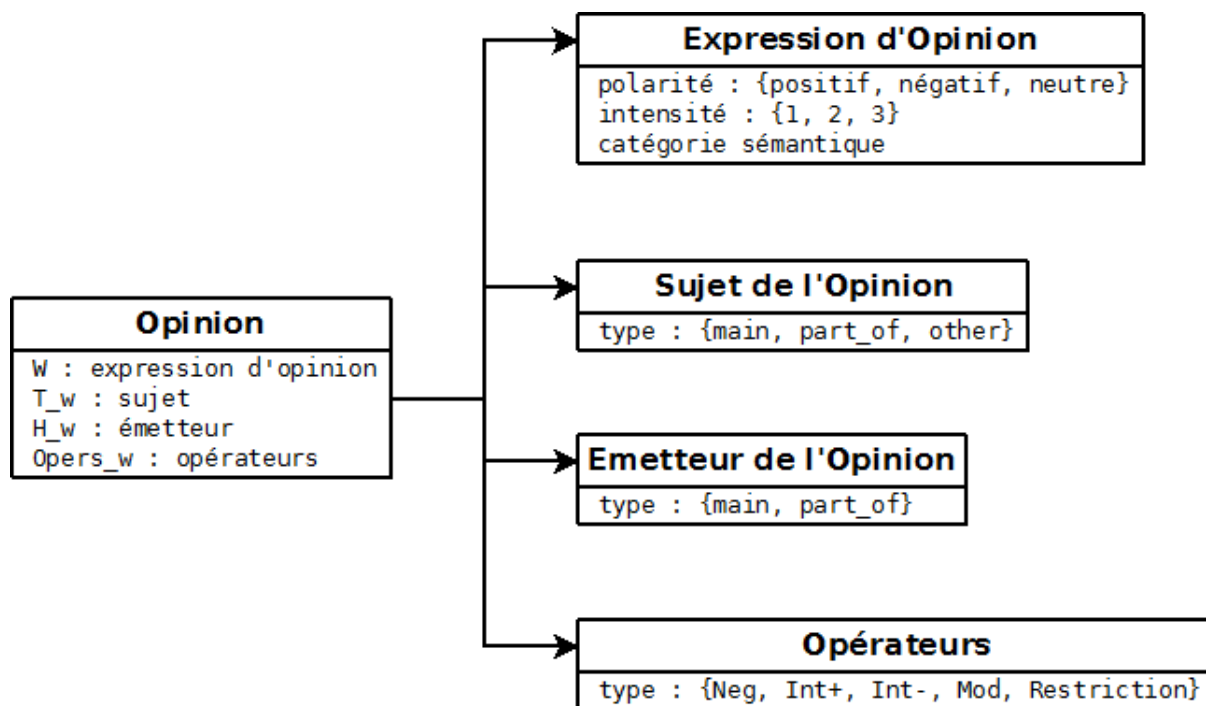


Figure 2.1 : Schéma récapitulatif de la structure d'une opinion au niveau local

### 2.3. L'opinion dans son contexte : niveau du segment discursif

Comme annoncé en introduction, notre modèle d'annotation ne se limite pas à la notion locale d'expression d'opinion, et comprend la structure discursive du texte. Cette section illustre les avantages de cette prise en compte, en montrant les limites de la notion d'expression d'opinion seule d'une part, et du traitement d'un texte comme un ensemble non ordonné d'opinions d'autre part.

Nous avons vu dans la section précédente une définition de la notion d'expression d'opinion, et comment l'annoter dans un texte. Néanmoins, toutes les opinions sont-elles exprimées dans un texte via une expression explicite d'opinion ? La réponse est non, et l'exemple suivant illustre ce fait :

(23) Les spectateurs ont tous quitté la salle avant la fin du film.

La phrase (23) évoque un fait, mais véhicule néanmoins une opinion sur le film, qui est d'ailleurs très forte. Dans cette phrase, il n'y a pas d'expression d'opinion à proprement parler : aucun segment de la phrase isolé ne constitue une opinion a priori. Cela illustre la nécessité de considérer un niveau intermédiaire entre l'expression et le document afin de prendre en considération toutes les opinions du texte. Ce niveau est celui du segment discursif.

### 2.3.1. Segments discursifs et opinion

Notre objectif est de calculer l'opinion globale d'un texte, à partir des opinions contenues dans les segments et de la structure discursive du texte. Une définition de la notion d'unité élémentaire de discours (UDE, ou en anglais EDU – Elementary Discourse Unit) a été proposée dans le cadre du projet ANNODIS<sup>15</sup> (Péry-Woodley et al. 2009; Afantenos et al. 2010; Afantenos et al. 2012) :

« La première étape dans l'annotation des relations de discours est de segmenter un texte en segments, appelés EDU – Elementary Discourse Units, qui doivent correspondre aux unités à mettre en relation avec des relations rhétoriques. L'EDU prototypique est une proposition indépendante. En général, une EDU correspond à la description d'un événement ou d'un état unique. »

Néanmoins, dans cette définition et dans le corpus annoté pour le projet ANNODIS, il est tout à fait possible d'avoir plusieurs opinions dans un même segment, ce qui empêche de calculer les caractéristiques d'opinion au niveau du segment. En effet, considérons les deux phrases suivantes, comportant chacune une seule UDE :

- Ce restaurant est bon mais cher.
- Ce restaurant est cher mais bon.

En raison des contrastes locaux introduits par "*mais*", la phrase 1 met l'accent sur "*cher*" et est donc plutôt négative, et la phrase 2 met l'accent sur "*bon*" et est donc plutôt positive.

Il apparaît donc difficile de remonter une opinion unifiée au niveau de l'UDE, telle que définie dans le projet ANNODIS. Pire : afin de mener à bien cette tâche, il serait nécessaire de réaliser une étude des relations locales à l'intérieur du segment.

L'UDE ANNODIS n'est donc pas une unité de texte satisfaisante pour l'étude de l'opinion globale. Afin de capturer les interactions locales, il nous faut donc définir un niveau de segment plus fin que l'UDE ANNODIS.

Ces segments sont des sous-segments des segments ANNODIS : il s'agit donc d'un sur-découpage, venant en aval du découpage ANNODIS, et non pas d'un redécoupage complet. Le découpage suit les règles suivantes :

- Une UDE peut être redécoupée si elle contient plusieurs opinions (nous ne sommes pas intéressés à capturer les interactions fines entre éléments objectifs)
- Dans le cas de plusieurs opinions en conjonction ou disjonction dans une même UDE, celle-ci doit être redécoupée au niveau de la jonction, c'est-à-dire après la virgule, ou avant la conjonction de coordination.

---

<sup>15</sup> L'objectif du projet ANNODIS est la construction d'un corpus annoté et outillé, pour servir de référence pour l'étude de l'organisation du discours en français.

Par la suite, les termes UDE ou segments de discours correspondront à cette définition spécifique à l'opinion.

### 2.3.1.1. Caractéristiques d'opinion associées à une UDE

Nous avons vu dans la section précédente qu'une expression d'opinion possédait plusieurs traits associés. De même, nous pouvons associer au niveau du segment plusieurs attributs correspondant à l'opinion véhiculée.

#### *Subjectivité*

Une UDE peut être subjective ou objective, i.e. représenter une opinion ou non. De plus dans le cas d'un segment subjectif, celui-ci peut être explicite ou implicite. Nous distinguons cinq catégories principales de subjectivité.

#### Segments objectifs purs (O)

Les segments objectifs purs (O) ne comportent pas d'opinion. Plus explicitement, ils ne comportent ni élément lexical subjectif, ni opinion implicite.

#### **Exemple :**

(24) [ Ce rôle est incarné par Jean Dujardin. ]<sub>a</sub>

#### Segments subjectifs évaluatifs explicites (SE)

Les segments subjectifs explicites (SE) comportent au moins une opinion évaluative explicite, c'est-à-dire au moins une expression d'opinion. Leur polarité peut être positive, négative, ou neutre hors contexte (i.e. d'une polarité pouvant être interprétée comme positive ou négative suivant le contexte).

#### **Exemple :**

(25) [ Ce rôle est incarné par un Jean Dujardin exceptionnel. ]<sub>a</sub>

#### Segments subjectifs non évaluatifs (SN)

Les segments subjectifs non évaluatifs (SN) sont des segments subjectifs qui introduisent l'opinion. Ils comportent une expression de reportage, en général un verbe d'introduction de la parole.

#### **Exemple :**

(26) [ Je pense ]<sub>a</sub> [ que ce film mérite son oscar. ]<sub>b</sub>

Ici, le premier segment est subjectif non évaluatif.

#### **Remarque importante :**

Les segments subjectifs non-évaluatifs ne rentrent pas dans la notion de polarité neutre hors contexte vue précédemment, puisqu'ils ne comportent pas d'opinion.

#### Segments subjectifs implicites (SI)

Les segments subjectifs implicites (SI) sont des segments porteurs d'opinion, mais ne contenant pas d'expression d'opinion : l'opinion, positive ou négative, est sous-entendue dans le contenu

propositionnel factuel. Privés de leur contexte, il est impossible de détecter l'opinion, et ces segments deviennent objectifs.

**Exemple :**

(27) [ Le film aux huit oscars sort en DVD mardi. ]<sub>a</sub>

Ici, le segment (27.a) présente des faits (le film sort en DVD mardi, le film a été primé à la cérémonie des oscars), mais la sélection de ces faits (i.e. la caractérisation du film par ses récompenses, et non pas par le nom de son réalisateur par exemple) induit une opinion positive.

### Segments subjectifs mixtes (SEI)

Les segments subjectifs mixtes comportent à la fois une opinion explicite, et une opinion implicite distincte de la précédente. Souvent, le sujet de l'opinion explicite est différent et n'a rien à voir avec le sujet de l'opinion implicite.

**Exemple :** Dans le contexte d'un commentaire web de film :

(28) [ Les spectateurs sont sortis bruyamment de la salle. ]<sub>a</sub>

Le segment (28.a) comporte les deux opinions suivantes :

Opinion	Polarité	Sujet	Émetteur
bruyamment	négative	(Les) spectateurs	<locuteur>
<implicite>	négative	<film>	(Les) spectateurs

Tableau 2.8 : Opinions du segment (28.a)

### Polarité

La polarité d'un segment peut être positive ou négative pour les segments SE, SI, et SEI. Elle dépend, pour un segment explicite SE, des expressions d'opinion constituant le segment, ainsi que des opérateurs s'appliquant sur celles-ci, en particulier les opérateurs de négation.

Les segments SN et O n'ont pas de polarité associée (valeur *no\_polarity* en pratique).

### Intensité

L'intensité d'un segment est une valeur, sur la même échelle vue précédemment au niveau de l'expression d'opinion ( $\{1; 2; 3\} = \{faible, moyenne, forte\}$ ) pour les segments subjectifs. Une intensité de 0 est arbitrairement associée aux segments objectifs.

L'intensité dépend, pour un segment explicite, des expressions d'opinion constituant le segment, ainsi que des opérateurs s'appliquant sur celles-ci.

## 2.3.2. Structure discursive globale

Une fois l'annotation et la caractérisation de l'opinion au niveau du segment effectuée, la dernière étape est d'annoter l'opinion au niveau du document dans sa globalité. Pour cela, nous prenons en compte deux éléments dans notre modèle : la note globale, i.e. l'opinion globale associée au document, et la structure discursive, qui lie entre elles les différentes opinions annotées au niveau du segment.

Dans cette section, nous décrivons brièvement les différents éléments de la structure discursive que nous repérons.

### 2.3.2.1. Note globale du document

Nous avons demandé aux annotateurs de donner une note associée au document décrivant l'opinion globale véhiculée par le texte. Cette note va de -3 (document véhiculant une opinion très négative) à +3 (document véhiculant une opinion très positive), sur une échelle similaire à celle employée pour l'intensité d'une expression.

Au-delà de la différence d'échelle, cette note est distincte de la note en étoiles donnée par le rédacteur pour certains types de documents : en effet, la note attribuée par les annotateurs ne se base que sur le contenu du document, alors que les étoiles peuvent varier suivant un ressenti non décrit explicitement par le rédacteur. Les annotateurs n'ont, le cas échéant, pas accès au nombre d'étoiles données par le rédacteur.

### 2.3.2.2. Relations rhétoriques

Pour la partie discursive, notre manuel d'annotation s'inspire de celui développé dans le cadre du projet ANNODIS (Péry-Woodley et al. 2009; Afantenos et al. 2010; Afantenos et al. 2012).

Les segments discursifs ainsi repérés sont reliés entre eux par des relations rhétoriques, explicitant les liens discursifs structurant le texte.

Ces liens peuvent soit être explicités par des marqueurs spécifiques de discours (par exemple *car*, *mais*, *cependant*,...). L'exemple suivant illustre ce cas.

**Exemple** : deux segments en relation de contraste marqué :

[Ce restaurant est bon,]<sub>1</sub> [mais très cher.]<sub>2</sub>

Ces liens peuvent également être implicites, tout en étant indispensable à la bonne compréhension du texte, comme dans l'exemple suivant.

**Exemple** : deux segments en relation d'explication non marquée :

[L'équipe de France de handball a perdu lamentablement hier.]<sub>1</sub> [Elle avait trop de blessés.]<sub>2</sub>

Notre objectif lors de l'annotation de ces relations est de voir en quoi le discours peut être utile lors de l'analyse de l'opinion, et non de constituer une banque de données annotée discursivement qui vérifierait l'adéquation entre la SDRT et l'intuition des annotateurs. Par conséquent, en plus d'une définition précise de la notion de relation de discours, notre manuel d'annotation détaille clairement les contraintes que doivent respecter les annotateurs lors de l'étiquetage du document (par exemple, le principe de frontière droite et l'absence de cycles dans le graphe final).

### Types sémantiques des relations

Nous distinguons les types de relations suivantes dans notre annotation discursive pour l'analyse d'opinion :

#### Relations de causalité

- Explication (subordonnante) : le segment source explique le segment cible
  - o **Exemple** :
    - [L'équipe de France de handball a perdu lamentablement hier.]<sub>1</sub> [Elle avait trop de blessés.]<sub>2</sub>
    - Explication(1, 2)
- But (coordonnante) : le segment cible décrit le but des actions ou évènements évoqués dans le segment source
  - o Exemple :
    - [Ce restaurant a fait de gros travaux]<sub>1</sub> [afin d'améliorer l'accessibilité aux personnes à mobilité réduite.]<sub>2</sub>
    - But(1, 2)
- Résultat (coordonnante) : le segment cible exprime le résultat des actions ou évènements décrit dans le segment source
  - o Exemple :
    - [Ce restaurant a fait de gros travaux]<sub>1</sub> [dans le but d'améliorer l'accessibilité aux personnes à mobilité réduite.]<sub>2</sub> [La salle a donc été complètement remaniée.]<sub>3</sub>
    - Résultat(1, 3)

#### Relations structurelles

- Parallèle (coordonnante) : les deux segments reliés par cette relation sont reliés par une construction parallèle ou symétrique.
  - o Exemple :
    - [Les entrées sont mauvaises,]<sub>1</sub> [les plats sont infâmes,]<sub>2</sub> [les desserts sont catastrophiques.]<sub>3</sub>
    - Parallèle(1, 2), Parallèle(2, 3)
- Contraste (coordonnante<sup>16</sup>) : le contenu de deux segments sont placés en opposition, en contraste.
  - o Exemple :
    - [Ce restaurant est bon,]<sub>1</sub> [mais très cher.]<sub>2</sub>
    - Contraste(1, 2)
- Continuation (coordonnante) : le contenu du segment cible s'inscrit dans la continuité discursive du contenu du segment source.
  - o Exemple :
    - [Le filet mignon était excellent,]<sub>1</sub> [et parfaitement cuit.]<sub>2</sub>

---

<sup>16</sup> (Asher and Vieu 2005) montrent néanmoins la possibilité que, sous certaines conditions, cette relation soit subordonnante. Nous ne traiterons pas ici ce cas particulier.

- Continuation(1, 2)

### Relations logiques

- Conditionnelle (coordonnante) : la véracité de la cible est soumise à la condition de la véracité de la source
  - Exemple :
    - [Si j'avais su,]<sub>1</sub> [je ne serais pas venu.]<sub>2</sub>
    - Conditionnelle(1, 2)
- Alternation (coordonnante-) : les segments source et cible sont en disjonction.
  - Exemple :
    - [Soit le chef était ivre,]<sub>1</sub> [soit je ne sais plus reconnaître les épices.]<sub>2</sub>  
[Dans tous les cas, le curry avait goût de cumin.]<sub>3</sub>
    - Alternation(1, 2)

### Discours rapporté

- Attribution (subordonnante) : le segment cible est attribué à son émetteur par le segment source.
  - Exemple :
    - [Pierre a affirmé]<sub>1</sub> [que ce restaurant était le meilleur de Toulouse.]<sub>2</sub>
    - Attribution(1, 2)

### Relations de narration

- Arrière-plan (subordonnante) : le segment cible fournit des informations concernant l'arrière-plan (le contexte général de narration) du segment source
  - Exemple :
    - [Je suis entré dans le restaurant.]<sub>1</sub> [Une bonne odeur flottait dans la salle.]<sub>2</sub>
    - ArrierePlan(1, 2)
- Narration (coordonnante) : deux segments décrivent deux évènements successifs
  - Exemple :
    - [J'ai commandé mon plat.]<sub>1</sub> [Le serveur me l'a rapidement apporté.]<sub>2</sub>
    - Narration(1, 2)
- Flashback (coordonnante) : une narration relie le segment source et le segment cible dans l'ordre temporel inverse
  - Exemple :
    - [Je suis allé dans ce restaurant Lundi dernier.]<sub>1</sub> [J'y étais déjà venu deux ans auparavant.]<sub>2</sub>
    - Flashback(1, 2)
- Encadrement (subordonnante) : le segment cible (adverbial ou groupe prépositionnel en tête de phrase) sert de cadre temporel ou spatial pour le segment source
  - Exemple :
    - [Hier au soir,]<sub>1</sub> [nous avons mangé dans cet excellent restaurant thaï.]<sub>2</sub>
    - Encadrement(1, 2)

### Relations d'élaboration

- Elaboration (subordonnante) : le segment cible décrit un sous-évènement du segment source.
  - o Exemple :
    - [Ce restaurant est très bon.]<sub>1</sub> [Ma pizza en particulier était délicieuse.]<sub>2</sub>
    - Elaboration(1, 2)
- Elaboration d'entité (subordonnante) : le segment cible décrit une entité présente dans le segment source
  - o Exemple :
    - [Baptiste a adoré sa pizza.]<sub>1</sub> [C'est un grand fan de quatre fromages.]<sub>2</sub>
    - E-Elaboration(1, 2)

### Autre relations

- Correction (subordonnante) : le segment cible apporte une correction au contenu du segment source
  - o Exemple :
    - [Ce restaurant est très bon]<sub>1</sub> [(quoique le très soit un peu exagéré)]<sub>2</sub>
    - Correction(1, 2)
- Commentaire (subordonnante) : le segment cible commente ce qui est décrit dans le segment source
  - o Exemple :
    - [Ce restaurant est sur les allées Jean Jaurès,]<sub>1</sub> [ce qui est un emplacement très pratique.]<sub>2</sub>
    - Commentaire(1, 2)

#### 2.3.2.3. Graphe du discours

L'annotation discursive est un processus incrémental. En considérant les segments élémentaires dans l'ordre de lecture du texte, l'annotation consiste à rechercher, pour le segment courant, à quel(s) segment(s) et par quelle(s) relation(s) attacher ce nouveau segment aux segments précédemment lus. Une fois tous les segments attachés, nous obtenons un graphe dont les nœuds sont les segments et les arêtes les relations.

#### *Attachement : relations coordonnantes et subordonnantes*

Comme annoncé en section précédente, une relation rhétorique peut être soit coordonnante (les segments sont au même niveau de granularité logique dans le discours), soit subordonnante (le segment cible est situé sur un niveau de granularité plus fin au niveau du discours). Dans le graphe de discours, les relations coordonnantes sont représentées par une flèche horizontale. Les relations subordonnantes sont quant à elles représentées par une flèche verticale.

Au cours de la construction du graphe de discours, n'est considéré accessible pour attacher un nouveau segment que la frontière droite du graphe, c'est-à-dire le dernier segment attaché en position cible, ou un de ses subordonnants.



### Segments complexes

Certains segments peuvent constituer une unité avec une cohérence forte, et peuvent ainsi être considérés en relation globale avec un autre segment. Ces segments sont alors groupés en une seule unité dite segment complexe.

#### Exemple :

[Ce restaurant n'est pas accueillant,]<sub>1</sub> [car la décoration est froide]<sub>2</sub> [et le service peu aimable.]<sub>3</sub>

Nous avons ici le segment complexe [2,3], et les deux relations :

- Explication(1, [2,3])
- Continuation(2, 3)

Graphiquement, un segment complexe peut se représenter par une boîte englobant plusieurs nœuds. Il est possible de créer des relations sortant d'un segment complexe (i.e. des relations dont le segment source est englobé dans un segment complexe  $C$ , mais dont le segment cible n'est pas inclus dans  $C$ ), mais pas de créer des relations entrantes (i.e. des relations dont le segment cible est englobé dans un segment complexe  $C$ , mais dont le segment source n'est pas inclus dans  $C$ ). La figure suivante montre un exemple de relation entrante, interdite en SDRT.

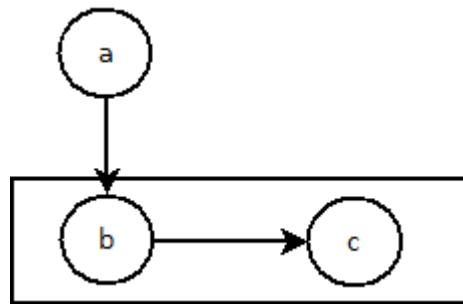


Figure 2.2 : Exemple de relation entrante – interdite – le segment  $a$  est relié par une relation subordonnante avec le segment  $b$ , alors que celui-ci est inclus dans le segment complexe  $(b,c)$

## 2.4. Exemple d'annotation

L'exemple suivant est extrait de notre corpus de commentaires web. La segmentation a été effectuée de manière automatique, puis rectifiée manuellement. Une note globale de +3 a été affectée à ce document en suivant le manuel. Le texte a été retranscrit tel quel, et contient donc, comme la plupart des documents annotés, quelques erreurs orthographiques et grammaticales.

[Je viens de découvrir le Sushiyaki.]\_1 [On y mange très bien.]\_2 [Je vous recommande les sashimis de saumon.]\_3 [Ils sont meilleurs que ceux]\_4 [que j'ai mangés le mois dernier au Japan.]\_5 [Les glaces au thé vert m'ont rappelé mon enfance au pays.]\_6 [En bref, une cuisine saine]\_7 [et très savoureuse.]\_8 [Le restaurant devrait figurer sur le site de l'Office de Tourisme de Toulouse.]\_9

Nous allons tout d'abord présenter l'outil ayant servi pour les annotations d'opinion et de discours, puis les annotations sur ce document

### 2.4.1. Outil d'annotation

L'annotation a été effectuée via le logiciel Glozz<sup>17</sup>. Glozz est un outil qui propose une interface dédiée à l'annotation, développé dans le cadre du projet ANNODIS (Péry-Woodley et al. 2009).



Figure 2.3 : Interface générale de Glozz

Glozz permet de distinguer trois types d'éléments : les *unités*, les *relations*, et les *schémas*.

Les *unités* correspondent à l'annotation d'une portion de texte simple. Chaque *unité* possède un ensemble d'attributs. Dans notre modèle d'annotation, les *expressions d'opinion*, les *opérateurs*, le *sujet de l'opinion*, l'*émetteur de l'opinion* sont des *unités* Glozz.

Les *schémas* permettent de regrouper plusieurs *unités* entre elles. Comme les *unités*, les *schémas* peuvent posséder des attributs. Dans le contexte de notre modèle d'annotation, nous introduisons le *schéma* Glozz *Opinion*, reliant l'expression d'opinion à son sujet, à son émetteur, et aux opérateurs venant s'appliquer sur elle. Une *unité* peut appartenir à plusieurs *schémas*.

Les *relations* permettent de relier deux *unités* entre elles de manière asymétrique (i.e. avec un segment source et un segment cible). Cet objet est particulièrement adapté aux relations discursives, mais également pour relier une anaphore (par exemple d'expression d'opinion) à l'unité à laquelle elle se réfère.

**Remarque** : Il est également possible d'attacher des relations à un schéma. La structure devient néanmoins moins lisible via l'outil de Glozz, ce qui peut amener des erreurs lors du processus d'annotation.

<sup>17</sup> [www.glozz.org](http://www.glozz.org)

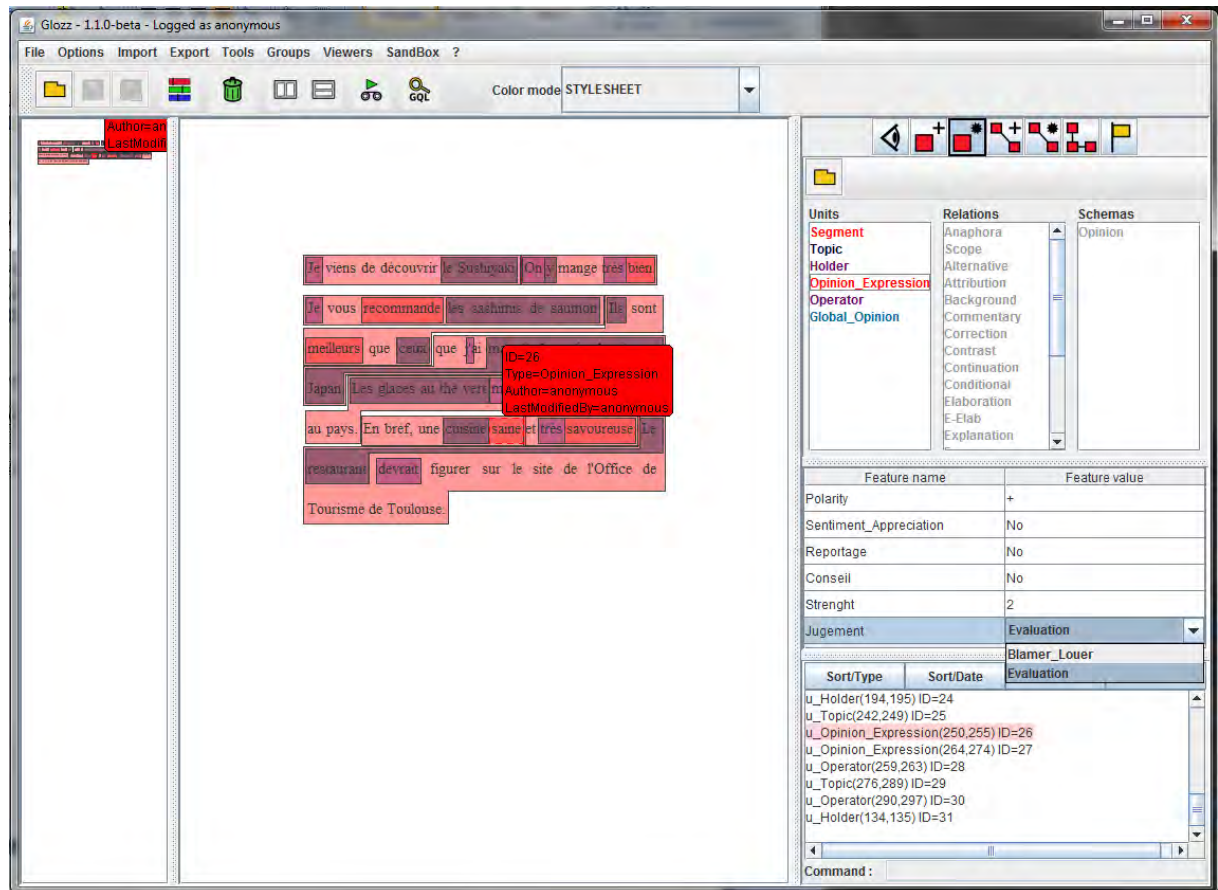


Figure 2.4 : Annotation de l'exemple sous Glozz

### 2.4.2. Annotation du document

Au sein de chaque segment, nous pouvons repérer les éléments suivants. Chaque élément est associé à un identifiant (id) pour simplifier la retranscription des relations.

- [Je viens de découvrir le Sushiyaki.]\_1

id	Portion de texte concernée	Type d'annotation	Attributs
1.1	Je	Émetteur	type=main_holder
1.2	Sushiyaki	Sujet	type=main_topic

- [On y mange très bien.]\_2

id	Portion de texte concernée	Type d'annotation	Attributs
2.1	On	Émetteur	type=main_holder
2.2	mange	Sujet	type=partof_topic
2.3	très	Opérateur	type=Intensifieur+
2.4	bien	Expression d'Opinion	type=Jugement/Évaluation ; polarité=+ ; intensité=2

Les éléments suivant ont été regroupés en schémas Glozz (cf. définition section précédente) :

- Opinion(2.1, 2.2, 2.3, 2.4)

- [Je vous recommande les sashimis de saumon.]\_3

id	Portion de texte concernée	Type d'annotation	Attributs
3.1	Je	Émetteur	type=main_holder
3.2	recommande	Expression d'Opinion	type=Conseil/Recommander ; polarité=+ ; intensité=2
3.3	les sashimis de saumon	Sujet	type=partof_topic

Les éléments suivant ont été regroupés en schémas :

- Opinion(3.1, 3.2, 3.3)

- [Ils sont meilleurs que ceux]\_4

id	Portion de texte concernée	Type d'annotation	Attributs
4.1	Ils	Sujet	type=partof_topic
4.2	meilleurs	Expression d'Opinion	type= Jugement/Évaluation ; polarité=+ ; intensité=better
4.3	ceux	Sujet	type=other_topic

Les éléments suivant ont été regroupés en schémas :

- Opinion(4.1, 4.2, 4.3)

- [que j'ai mangés le mois dernier au Japan.]\_5

id	Portion de texte concernée	Type d'annotation	Attributs
5.1	Je	Émetteur	type=main_holder
5.2	mangés le mois dernier au Japan	Sujet	type=other_topic

- [Les glaces au thé vert m'ont rappelé mon enfance au pays.]\_6

id	Portion de texte concernée	Type d'annotation	Attributs
6.1	Les glaces au thé vert	Sujet	type=partof_topic
6.2	m'	Émetteur	type=main_holder

- [Bref, une cuisine saine]\_7

id	Portion de texte concernée	Type d'annotation	Attributs
7.1	Une cuisine	Sujet	type=partof_topic
7.2	saine	Expression d'Opinion	type= Jugement/Évaluation ; polarité=+ ; intensité=2

Les éléments suivant ont été regroupés en schémas :

- Opinion(7.1, 7.2)

- [et très savoureuse.]\_8

id	Portion de texte concernée	Type d'annotation	Attributs
8.1	très	Opérateur	type=Intensifieur+
8.2	savoureuse	Expression d'Opinion	type=Jugement/Évaluation ; polarité=+ ; intensité=2

Les éléments suivant ont été regroupés en schémas :

- Opinion(8.1, 8.2)

- [Le restaurant devrait figurer sur le site de l'Office de Tourisme de Toulouse.]\_9

id	Portion de texte concernée	Type d'annotation	Attributs
9.1	restaurant	Sujet	type=main_topic
9.2	devrait	Opérateur	type=Modalité

Le Tableau 2.9 suivant reprend les valeurs d'opinions annotées pour chacun des segments.

Segment	Subjectivité	Polarité	Intensité
1	O	/	/
2	SE	+	3
3	SE	+	2
4	SEI	+	3
5	O	/	/
6	SI	+	2
7	SE	+	2
8	SE	+	3
9	SI	+	2

Tableau 2.9 : Valeurs de subjectivité et d'opinion pour chacun des segments

Enfin, la structure rhétorique du document a été annotée de la façon suivante (les segments complexes sont indiqués par la liste entre crochets de leurs constituants :

- ArrièrePlan(1, [2,3,4,5,6])
- Elaboration(2, [3,6])
- Continuation(3,6)
- Explication(3,[4,5])
- E-Elab(4,5) % entité = sashimis de saumon
- Résultat([2,3,4,5,6], [7,8,9])
- Continuation(7,8)
- Résultat([7,8], 9)

Cette structure peut être représentée par le graphe orienté suivant :

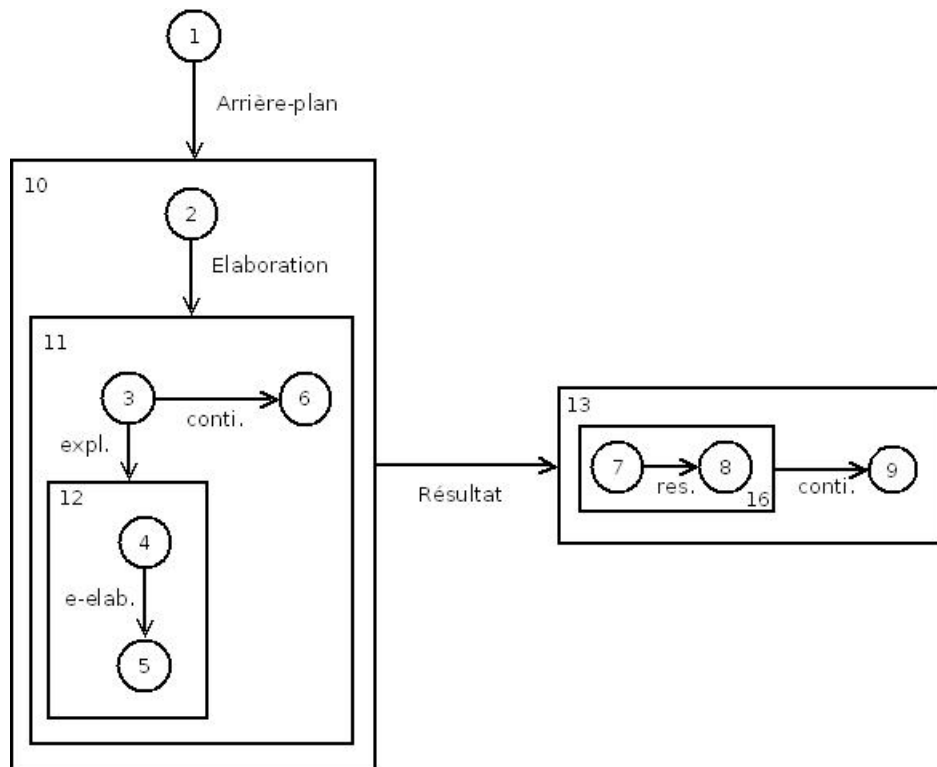


Figure 2.5 : Graphe de discours correspondant à l'annotation discursive de l'exemple

## 2.5. En pratique : processus d'annotation manuelle

Dans le cadre du projet CASOAR, le modèle d'annotation présenté dans ce chapitre a été mis en pratique dans une campagne d'annotations.

### 2.5.1. Déroulement pratique

#### 2.5.1.1. Types des corpus annotés

Nous avons sélectionné, dans le cadre du projet CASOAR, deux types de corpus afin de mettre en application le schéma d'annotations. Les corpus choisis devaient satisfaire plusieurs critères :

- Variété des opinions rencontrées : opinions lexicalement explicites, mais également des opinions implicites
- Longueur des documents : les documents ne doivent être ni trop long, ni trop court. En effet, l'annotation étant une tâche fastidieuse, il est difficile pour un annotateur de garantir une cohérence d'annotation sur un texte long, en particulier pour les annotations discursives. À l'inverse, un document trop court n'aura pas une structure rhétorique suffisante pour être exploitée.
- Possibilité d'avoir plusieurs documents traitant d'un même sujet : cela permettra d'étudier, ultérieurement, la détermination d'une opinion globale pertinente sur un sujet dans le cas d'une collection de documents

- Contrainte de temps : étant donné le budget alloué à l'annotation dans le projet CASOAR (3 mois à temps plein pour le français), l'annotation ne doit pas se faire sur plus de trois genres différents afin d'avoir suffisamment de documents annotés pour chaque genre.

Nous avons donc sélectionné deux corpus répondant à ces critères.

Le premier corpus choisi est un corpus de commentaires web (reviews) sur des restaurants, films, livres, et jeux vidéo. Chaque commentaire contient une évaluation, le plus souvent en termes de nombres d'étoiles puis le texte du commentaire ; dans le cadre de l'annotation, nous nous focalisons sur le texte : les étoiles ne sont pas transmises à l'annotateur. Les textes sont courts, typiquement un paragraphe de long. Ils contiennent parfois un nombre non négligeable de fautes d'orthographe, et le style utilisé comprend de temps en temps des abréviations, des émoticônes, et des marques de ponctuations emphatiques variées (par exemple, « !!! » ou encore « !!?? »). Les opinions émises concernent l'objet principal du commentaire (un film donné, un produit d'une marque donnée) mais aussi des parties-de ou des propriétés de cet objet (comme le scénario ou les acteurs pour un film, ou encore le service ou la décoration pour un restaurant). Un exemple de commentaires extraits du site Allociné<sup>18</sup>, est présenté en Figure 2.6.



**Alain Danet**  
 [Ajouter à mes amis](#) | [Discuter](#)  
 Aucun point  
 Membre depuis 26 jours

**Sa note :** ★★★★★ (5)

**Sa critique :** Intrigue solide , effets spéciaux épatants , réelle épaisseur psychologique des personnages , Un réalisateur dévoué au film (et au cinéma tout simplement ) qui nous fait oublier le 3ème opus sans âme du yes man Brett Rattner . Tout simplement le meilleur épisode de la franchise. Un commencement qui .....commence fort . Un blockbuster comme on aimerait en voir plus souvent ,

Ecrit le 19 juin à 19h51 - Signaler un abus

---

**Benjamin M.**  
 [Ajouter à mes amis](#) | [Discuter](#)  
 Aucun point  
 Membre depuis 174 jours

**Sa note :** ★★★★★ (4)

**Sa critique :** Je ne suis pas du tout un fan de X-Men mais cet épisode est sans conteste le meilleur de la saga. J'ai apprécié l'ambiance guerre froide. Les effets spéciaux sont bien mis en valeur car utilisés avec parcimonie. On regrettera des longueurs ainsi qu'un manque d'action pour un film de ce genre.

Ecrit le 19 juin à 19h48 - Signaler un abus

Figure 2.6 : Exemple de commentaires extraits du site Allociné

<sup>18</sup> <http://www.allocine.fr/>

Ces documents, riches en opinion par nature, correspondent à un type de documents étudiés par la communauté scientifique, notamment en anglais (citons notamment (Pang, Lee, and Vaithyanathan 2002; Pang and Lee 2005; Hu and Liu 2004; Popescu and Etzioni 2007; Kennedy and Inkpen 2006)), mais également à un besoin industriel d'exploitation et de prise en compte de retours clients.

Ensuite, nous avons extrait un corpus de réactions aux articles de presse de l'édition en ligne du journal LeMonde<sup>19</sup> : il s'agit des commentaires postés par les utilisateurs (abonnés) sur les articles mis en ligne. Ces textes sont assez courts et ne dépassent généralement pas deux paragraphes ; ils ont néanmoins en général une structure rhétorique bien construite. En plus de pouvoir commenter un article de presse, un utilisateur peut également répondre aux commentaires postés précédemment, donnant ainsi à ce corpus une structure un peu hybride entre le texte et le dialogue. La Figure 2.7 présente un exemple de ce type de commentaires, réagissant à l'article : "Le protectionnisme détruirait nos emplois et affaiblirait notre croissance" daté du 18/06/2011.

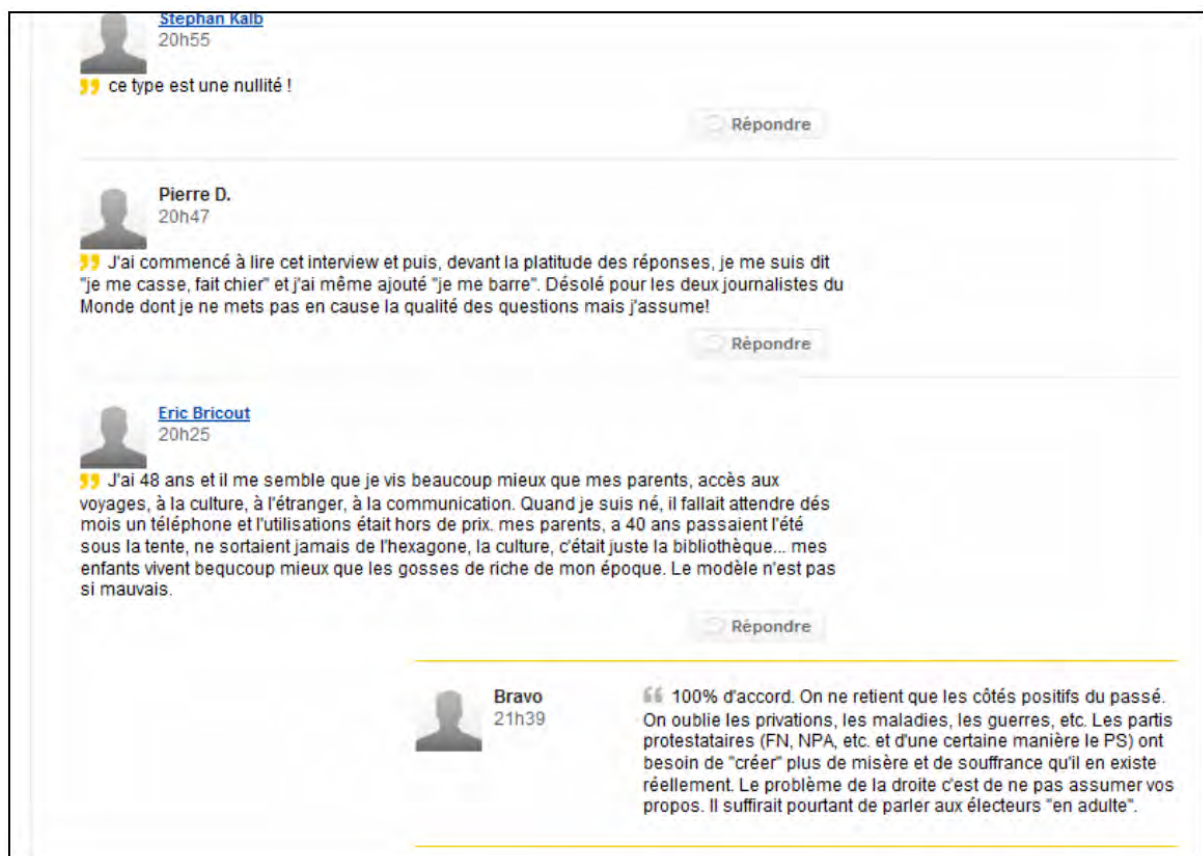


Figure 2.7 : Exemple de réactions à un article du Monde

En pratique, les réactions ont été extraites automatiquement entre le 15 et le 21 Août 2011, dans les catégories société, politique, actualité internationale. Seules les réactions de plus de trois phrases d'une part, et liées à un article ayant reçu au moins 10 réactions différentes d'autre part, sont retenues. Ceci nous a permis de sélectionner des réactions riches en opinion (les articles

<sup>19</sup> [www.lemonde.fr](http://www.lemonde.fr)



ayant de nombreuses réactions sont en général des articles appelant à un débat d'idées) et à la structure rhétorique suffisamment développée. Dans la littérature, ce type de corpus a été étudié dans le cadre de la campagne anglophone MPQA (Wiebe, Wilson, and Cardie 2005), ainsi que, entre autres travaux, par (Kim and Hovy 2006b).

Les deux corpus annotés présentent une certaine complémentarité : dans le cadre des commentaires de produits (*reviews*), les opinions sont majoritairement explicites, et souvent très intenses. Les réactions aux articles de presse présentent à l'inverse de nombreuses opinions implicites, et une structure rhétorique plus riche ; les commentaires web sont également lexicalement plus mesurés.

### 2.5.1.2. Guide pour les annotateurs

Afin de guider les annotateurs pendant leur annotation, un manuel d'annotation ainsi qu'un guide d'annotation leur ont été proposés. Le premier expose les éléments linguistiques à repérer et les consignes que doivent suivre les annotateurs pour cette tâche. Le second guide les annotateurs dans l'environnement logiciel. Ces deux manuels ont été réalisés sous forme de *Wiki*, ce qui permet un parcours du document à la recherche d'informations spécifiques aisé pour l'utilisateur, et une maintenance facile pour les membres du projet (tout le monde à toujours accès à la dernière version).

Afin de faciliter le travail des annotateurs, et ainsi de limiter les erreurs d'annotations, deux artefacts d'annotation ont été introduits dans ce manuel :

- Une relation d'anaphore, permettant à l'annotateur d'indiquer que certaines annotations font référence à d'autres éléments précédemment annotés. Ces éléments anaphoriques, qui peuvent être des expressions d'opinion, mais aussi des sujets de l'opinion ou des émetteurs de l'opinion héritent de toutes les propriétés de l'élément référé.

#### Exemples :

- "Le plat de pâtes était bonne. La pizza quatre fromages l'était également."
  - o Anaphore("également", "bonne") entre expressions d'opinion
- "J'ai apprécié le tajine d'agneau. Celui-ci était par ailleurs très copieux."
  - o Anaphore("Celui-ci", "le tajine d'agneau") entre sujets de l'opinion
- "Pierre n'a pas aimé le saumon. Il l'a trouvé trop sec."
  - o Anaphore("Il", "Pierre") entre émetteurs de l'opinion

- Une relation rhétorique *Unknown* permettant de gérer l'incertitude de l'annotateur. Dans le cas où l'annotateur ne sait pas quelle relation est adaptée pour relier deux segments, mais qu'il sent qu'une relation existe, il est invité à ajouter une relation de ce type. Cela permet de limiter les erreurs d'annotations au niveau du discours, et permet de revenir ultérieurement sur ces points de doute, soit par l'annotateur lui-même après avoir annoté tout le document, soit au moment de l'établissement du *corpus gold*.

### 2.5.1.3. Annotation exploratoire

Afin de tester le modèle d'annotation, nous avons procédé à l'annotation manuelle de 115 commentaires web extrait du site Allociné<sup>20</sup>. Trois personnes, membres du projet CASOAR, ont participé à cette phase d'annotation.

Cette phase s'est déroulée en trois temps :

- Segmentation discursive manuelle des textes en unités discursives élémentaires
- Annotation du contenu d'opinion de chaque unité discursive
- Annotation d'un sous-sensé restreint du corpus au niveau discursif

L'annotation a été effectuée par consensus : les annotateurs ont pu discuter entre eux pour se mettre d'accord sur une annotation commune et sur la version finale du manuel et du guide d'annotation. Au cours de cette phase exploratoire, 1430 segments discursifs ont été annotés, soit une moyenne de 12.5 segments par document.

Cette phase d'annotation exploratoire a permis de dégager plusieurs aspects cruciaux pour la phase opérationnelle. Tout d'abord, il est primordial d'avoir une segmentation discursive correcte afin de pouvoir comparer les annotations d'opinion et les annotations de relations de discours produites par les différents annotateurs naïfs. Ainsi et afin d'éviter les erreurs liées à la segmentation, nous avons décidé que la phase de segmentation discursive serait effectuée en amont par les membres du projet. Les textes annotés par les annotateurs naïfs ont donc été segmentés avant la phase d'annotation opérationnelle. Ensuite, même si nous avons pu confirmer la stabilité du modèle d'annotation proposé, les annotations produites comprenaient un certain nombre d'erreurs liées au non respect des contraintes d'annotation. Ces erreurs sont notamment issues du fait que ces contraintes sont souvent difficiles à vérifier graphiquement, tout particulièrement dans le cas de documents longs et/ou denses. Un outil externe de vérification des contraintes a donc été développé, permettant de vérifier :

- La conformité des annotations d'opinion avec le type du segment discursif, par exemple :
  - o un segment objectif ne doit pas contenir d'expression d'opinion
  - o un segment subjectif explicite doit contenir au moins une annotation d'opinion
- La conformité des annotations de discours avec les règles SDRT de l'attachement, et notamment :
  - o Contrainte de frontière droite
  - o Caractère connexe du graphe de discours produit
  - o Non-chevauchement de segments complexes

---

<sup>20</sup> [www.allocine.fr](http://www.allocine.fr)

#### 2.5.1.4. Annotation opérationnelle

La campagne d'annotation opérationnelle s'est déroulée sur 6 mois, entre le 17 avril 2011 et le 17 octobre 2011. Pour cette campagne, trois annotateurs, initialement naïfs du domaine, ont été mobilisés. La campagne s'est déroulée en trois temps

- Pendant la première phase d'entraînement, pendant une semaine, les trois annotateurs ont annoté les mêmes textes tout en discutant de leur annotation. L'objectif était que les annotateurs se familiarisent avec l'outil Glozz d'une part, et avec le modèle d'annotation et les concepts manipulés d'autre part.
- Ensuite, pendant une seconde phase d'entraînement d'une semaine, les annotateurs ont annoté des textes distincts (10 textes), et n'ont pas discuté entre eux. Après analyse des annotations produites, des recommandations ont été données aux annotateurs quand cela était nécessaire, afin de limiter les erreurs et d'éviter les mauvaises interprétations.
- Enfin, dans un troisième temps, les annotateurs ont annotés le corpus opérationnel.

Une fois les annotateurs entraînés, le temps nécessaire à l'annotation d'un texte sur les trois niveaux – expression, segment et discours – a été évalué à environ 1h.

#### 2.5.2. Statistiques d'annotations

Nous présentons dans cette section plusieurs statistiques sur les corpus annotés. Ce travail a été réalisé dans le cadre du projet CASOAR, en collaboration avec Farah Benamara et Vladimir Popescu.

Les corpus annotés ont été séparés en trois parties, chaque partie étant annotée par deux annotateurs distincts. Chaque document est donc doublement annoté.

Notre objectif étant d'étudier l'interaction entre la structure discursive d'une part, et l'opinion annotée dans le document, nous nous focalisons dans cette section sur les statistiques suivantes :

- Analyse quantitative des opinions (cf. section 2.5.2.1)
- Accords inter-annotateurs sur la subjectivité au niveau du segment de discours (cf. section 2.5.2.2)
- Corrélation de la note globale du document avec les segments de discours explicites et implicites (cf. section 2.5.2.3)
- Analyse quantitative des relations discursives présentes dans les documents d'opinion (cf. section 2.5.2.4)

D'autres mesures peuvent bien sûr être réalisées sur ces annotations (par exemple, accords inter-annotateurs sur les caractéristiques des opinions au niveau local, sur l'émetteur et le sujet de l'opinion, ou sur la structure discursive annotée), mais n'entre pas dans le cadre de cette étude. Nous nous focalisons ici sur les éléments nouveaux de notre schéma d'annotation.

De même, l'objectif de cette étude n'étant pas la validation du schéma discursif (cf. les résultats du projet d'ANNODIS (Afantenos et al. 2012) concernant une telle étude), nous n'analyserons pas dans ce chapitre les accords inter-annotateurs sur les relations de discours.

### 2.5.2.1. Annotations d'opinion : analyse quantitative

Le tableau ci-dessous expose les statistiques générales caractérisant les deux corpus annotés.

	<b>Reviews</b>	<b>Le Monde</b>
<b>Total segments</b>	2751	1739
- dont segments SE	1482 (53.9 %)	574 (33.0 %)
- dont segments SN	121 (4.4 %)	79 (4.5 %)
- dont segments SI	587 (21.3 %)	607 (34.9 %)
- dont segments O	441 (16.0 %)	455 (26.2 %)
- dont segments SEI	120 (4.4 %)	24 (1.4 %)
- dont segments positifs	1237 (45.0 %)	375 (21.6 %)
- dont segments négatifs	934 (34.0 %)	857 (49.3 %)
- dont segments sans polarité, ambigu, ou neutre	580 (21.1 %)	507 (29.2 %)
<b>Segments avec une expression d'opinion</b>	1449	548
<b>Segments avec au moins 2 expressions d'opinion</b>	70	39
<b>Total Expressions d'opinion</b>	1596	632
<b>Segments avec au moins un opérateur</b>	778	264

Tableau 2.10 : Annotation opérationnelle : statistiques générales

Nous pouvons observer plusieurs différences importantes entre les deux types de corpus annotés.

Concernant le type de subjectivité, le poids des segments contenant des opinions implicites (*SI* et *SEI*) est très différent d'un genre à l'autre. En effet, alors que dans les commentaires web les expressions sont majoritairement exprimées via des expressions d'opinion explicites, on peut remarquer que les réactions aux articles de presse contiennent majoritairement des opinions implicites. Dans les deux cas, nous pouvons observer un taux similaire, assez faible, de segments purement objectifs (environ 30 %). Les documents annotés sont donc, comme pressentis, majoritairement constitués de segments porteurs d'opinion.

Concernant la polarité des documents, nous pouvons également observer un biais dépendant du corpus. Si les commentaires web contiennent majoritairement des opinions positives, les réactions de presse contiennent quant à elle une grande majorité de segments négatifs.

Dans leur grande majorité (plus de 90 % pour chaque corpus), les segments contenant des expressions d'opinion ne contiennent qu'une seule expression d'opinion. Ceci permet de valider

l'intérêt de notre segmentation en UDE comme passerelle directe entre l'expression d'opinion et la structure discursive du document.

Enfin, les opérateurs portant sur l'opinion sont présents en quantité non-négligeable, notamment dans le corpus de commentaires web, où 28 % des segments en contiennent au moins une. Ceci montre bien la nécessité pour une chaîne de traitement automatique d'intégrer ce type d'informations, en particulier pour les commentaires web.

### 2.5.2.2. Accords inter-annotateurs : catégories de subjectivité des segments

Les tableaux ci-dessous présentent les kappas de Cohen entre les annotateurs A1, A2 et A3. L'échelle va de -1 pour un désaccord complet à +1 pour un accord complet.

	<b>A1 – A2</b>	<b>A1 – A3</b>	<b>A2 – A3</b>
<b>Segments SE</b>	0.55	0.63	<b>0.69</b>
<b>Segments SN</b>	0.22	0.57	<b>0.60</b>
<b>Segments SI</b>	0.26	0.34	<b>0.53</b>
<b>Segments O</b>	0.47	0.57	<b>0.58</b>

Tableau 2.11 : Accords inter-annotateurs pour le corpus "Reviews"

	<b>A1 – A2</b>	<b>A1 – A3</b>	<b>A2 – A3</b>
<b>Segments SE</b>	0.23	0.25	<b>0.45</b>
<b>Segments SN</b>	0.17	0.14	<b>0.66</b>
<b>Segments SI</b>	0.08	0.15	<b>0.32</b>
<b>Segments O</b>	0.20	0.44	<b>0.40</b>

Tableau 2.12 : Accords inter-annotateurs pour le corpus "Le Monde"

Nous observons, pour les deux types de corpus, un meilleur accord sur les segments comportant une opinion explicite (SE) que sur les segments comportant une opinion implicite (SI). Par ricochet, les segments objectifs (O) ont également un accord plus faible : en effet, les segments implicites non détectés comme tels par un annotateur sont le plus souvent annotés en segments objectifs. Ceci est cohérent avec la nature des segments implicites, dont l'interprétation hors contexte n'est pas subjective.

Nous avons également pu constater un grand nombre d'inconsistances dans l'annotation produite par l'annotateur A1, ce qui se ressent dans les accords avec les deux autres annotateurs. L'accord entre A2 et A3 est, quant à lui, satisfaisant au regard de la difficulté de la tâche.

### 2.5.2.3. Opinion globale du document : lien avec les opinions implicites

Nous avons mesuré la corrélation entre les opinions portées par les segments implicites et la note globale du document afin de déterminer le poids de l'implicite dans chaque genre de corpus. Cette information est déterminante, car elle permet de déterminer si la prise en compte des opinions implicites est nécessaire pour la constitution d'une chaîne d'analyse de l'opinion sur un genre donné, ou si le seul repérage des expressions d'opinion explicitement présentes permet de déterminer avec une précision satisfaisante l'opinion globale véhiculée par un document donné.

La corrélation est mesurée en termes de coefficient de Pearson<sup>21</sup> entre deux vecteurs : la note attribuée au document par l'annotateur d'une part, et le type et la polarité du segment d'autre part. Plus précisément, nous avons constitué 4 vecteurs distincts :

- Expl\_Pos : nombre de segments explicites positifs pour chaque document
- Expl\_Neg : nombre de segments explicites négatifs pour chaque document
- Impl\_Pos : nombre de segments implicites positifs pour chaque document
- Impl\_Neg : nombre de segments implicites négatifs pour chaque document

Pour chacun de ces vecteurs, nous calculons le coefficient de Pearson. Celui-ci, entre -1 et +1, est interprété en termes de valeur absolue : plus la valeur absolue du score est élevée, plus la corrélation est importante. Naturellement, la corrélation entre la note et les vecteurs de décompte de segments négatifs sera négative (les deux valeurs ont une sémantique de la croissance opposée), tandis qu'elle sera positive entre les segments positifs et la note globale.

Les graphes ci-après présentent les résultats obtenus pour chaque corpus.

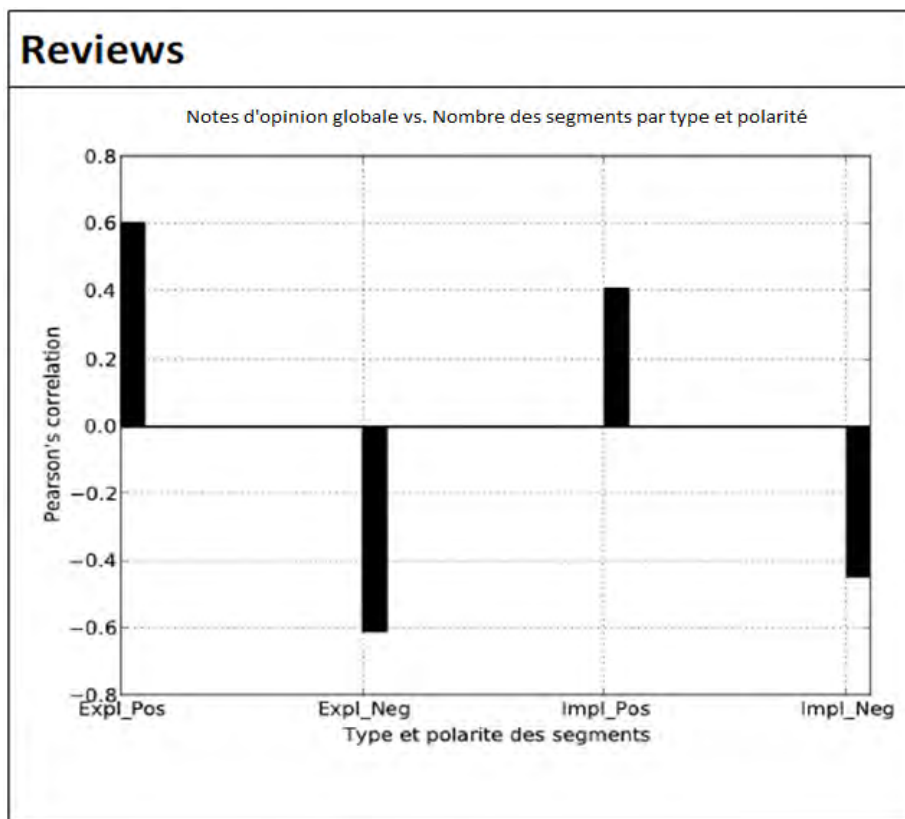


Figure 2.8 : Corrélation entre note globale et opinions explicites/implicites pour le corpus de commentaires web

<sup>21</sup> La corrélation de Pearson suit la formule suivante : 
$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$
.

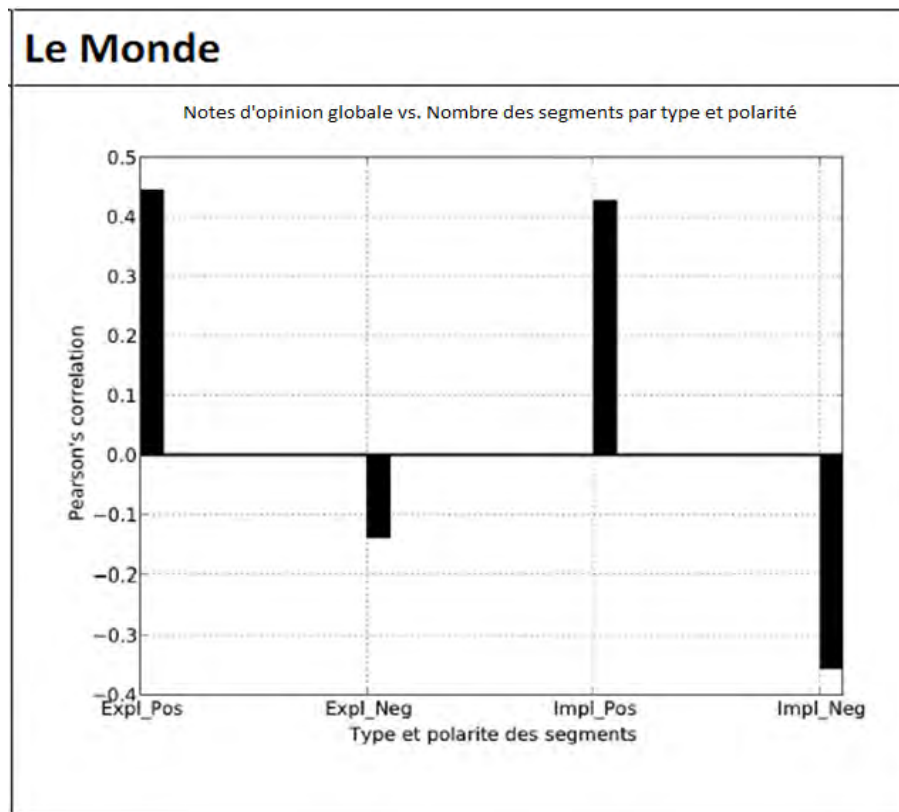


Figure 2.9 : Corrélation entre note globale et opinions explicites / implicites pour le corpus de réactions aux articles de presse

Pour le corpus de commentaires web, nous observons une plus forte corrélation entre opinions explicites et note globale qu'entre opinion implicite et note globale. Ceci tend à indiquer que dans ce genre de document, les opinions sont principalement émises grâce à des mentions explicites. Par conséquent, le repérage des opinions implicites semble moins essentiel pour ce type de document. Attention toutefois, les opinions implicites ne sont pas non plus décorréliées avec la note globale. Nous étudierons plus en détail au chapitre 6 la meilleure stratégie à adopter vis-à-vis de la subjectivité des segments pour le calcul de l'opinion globale.

À l'inverse, pour le corpus de réactions aux articles de presse, la corrélation entre note globale et segments implicites n'est pas plus faible que celle entre explicites et note globale. Ceci indique donc une importance plus forte des segments implicites, qui ne devront donc, a priori, pas être négligé pour une reconnaissance automatique correcte de la note globale. Ceci peut s'expliquer par une raison sociale : les opinions émises dans les avis de presse, souvent très négatives, sont exprimées de manière implicite d'une part par politesse envers l'interlocuteur, et d'autre part pour que ces opinions prennent l'aspect de faits : ainsi, ils peuvent prendre plus de poids lors d'un débat d'idées.

Nous avons également testé de pondérer le poids de chaque segment dans les vecteurs de comptage par l'intensité du segment. Les quatre segments deviennent alors :

- Expl\_Pos : somme des intensités des segments explicites positifs pour chaque document

- Expl\_Neg : somme des intensités des segments explicites négatifs pour chaque document
- Impl\_Pos : somme des intensités des segments implicites positifs pour chaque document
- Impl\_Neg : somme des intensités des segments implicites négatifs pour chaque document

Les graphes ci-après montrent les résultats obtenus :

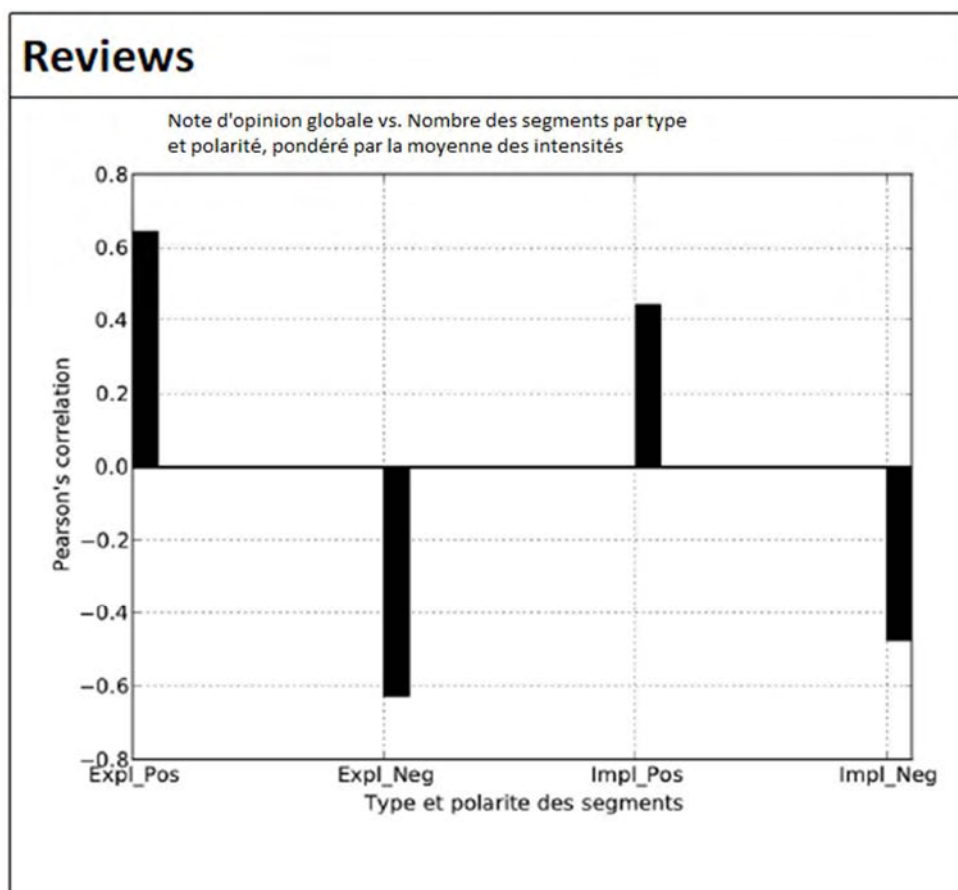


Figure 2.10 : Corrélation entre note globale et opinions explicites/implicites + intensités, corpus de commentaires web



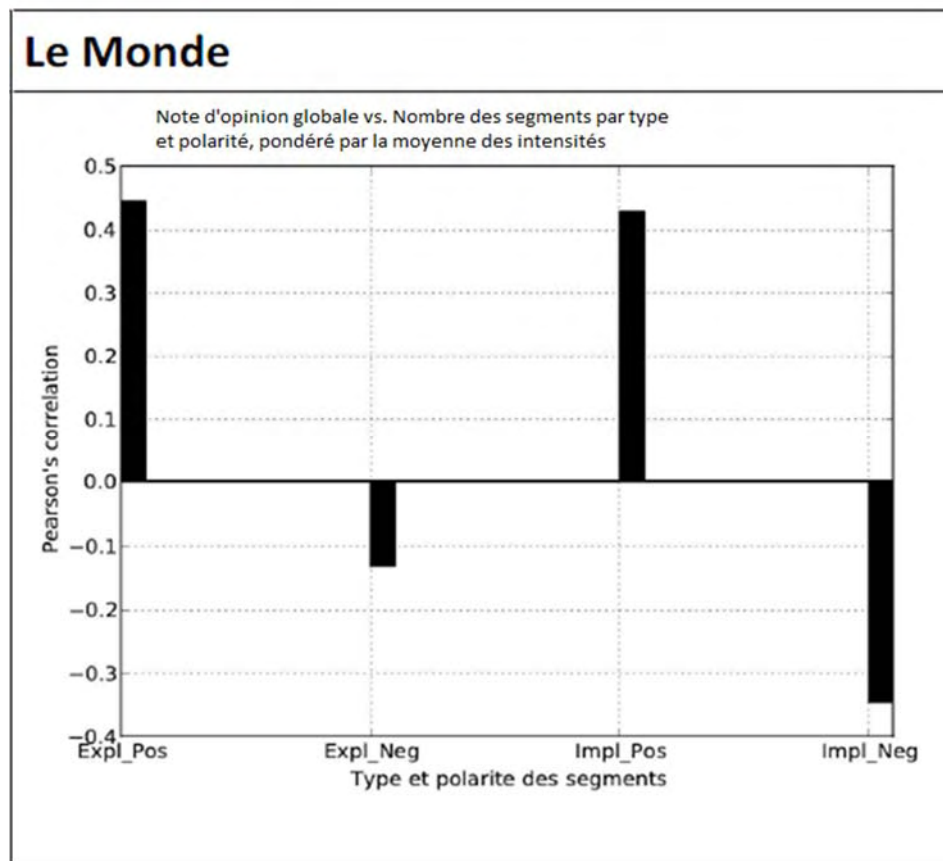


Figure 2.11 : Corrélation entre note globale et opinions explicites/implicites + intensités, corpus de réactions aux articles de presse

Les résultats obtenus sont sensiblement similaires à ceux obtenus sans prise en compte des intensités. L'intensité de l'opinion n'impacte donc pas la corrélation du type de subjectivité de l'opinion avec la note globale.

#### 2.5.2.4. Corrélation entre note globale et polarité de chaque segment

Nous nous sommes ensuite intéressés à la corrélation directe entre note globale et polarité du segment. Ceci nous permet de vérifier l'existence, parfois observée sur certain type de corpus, d'un « biais de politesse », c'est-à-dire si les rédacteurs ont tendance à nuancer leur opinion quant celle-ci est négative. En termes de corrélation, cela impliquerait une meilleure corrélation entre la note et le nombre global d'opinions positives qu'entre la note et le nombre total d'opinions négatives.

Les figures suivantes montrent les résultats obtenus :

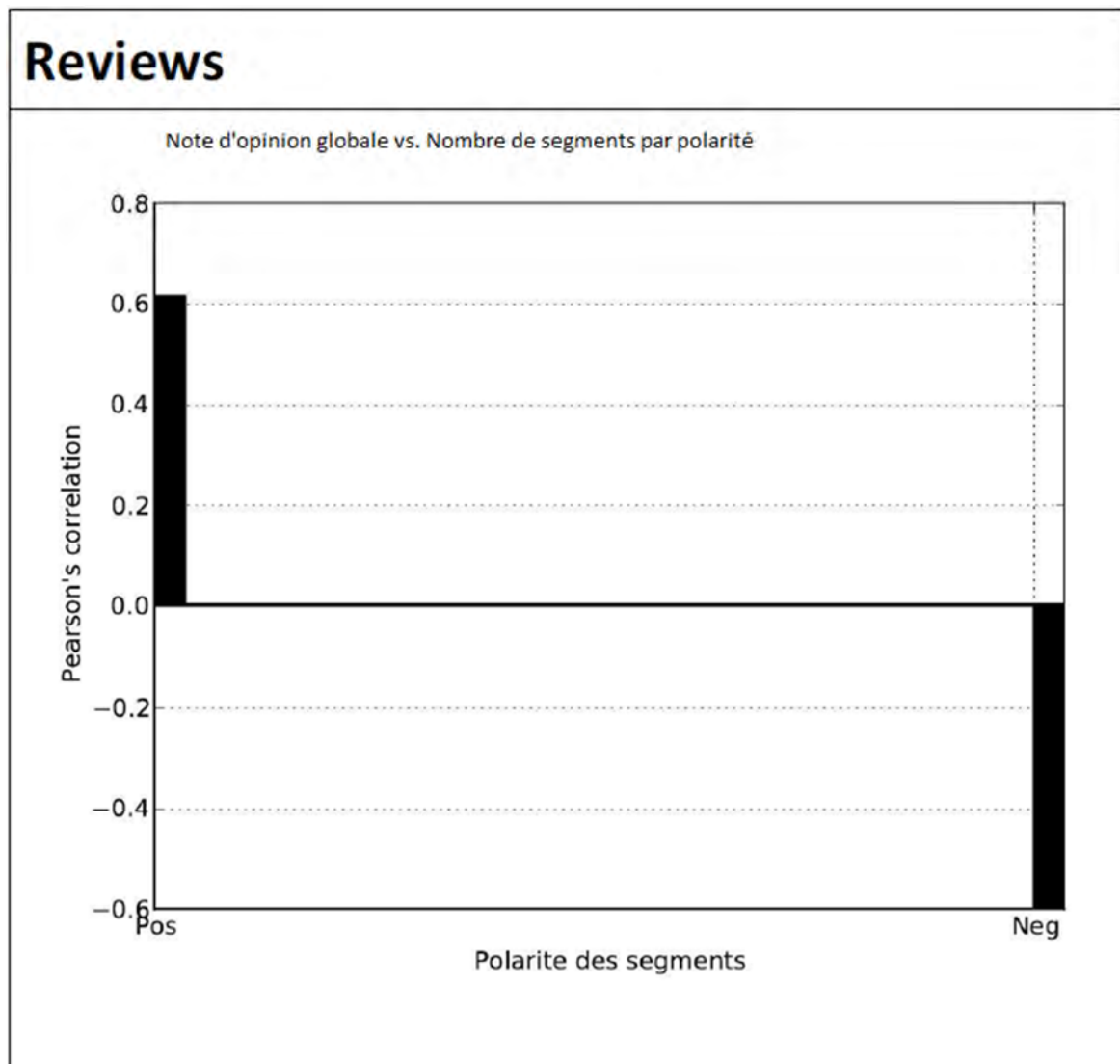


Figure 2.12 : Corrélation entre note global et nombre de segments pondéré par leurs intensités -  
Reviews

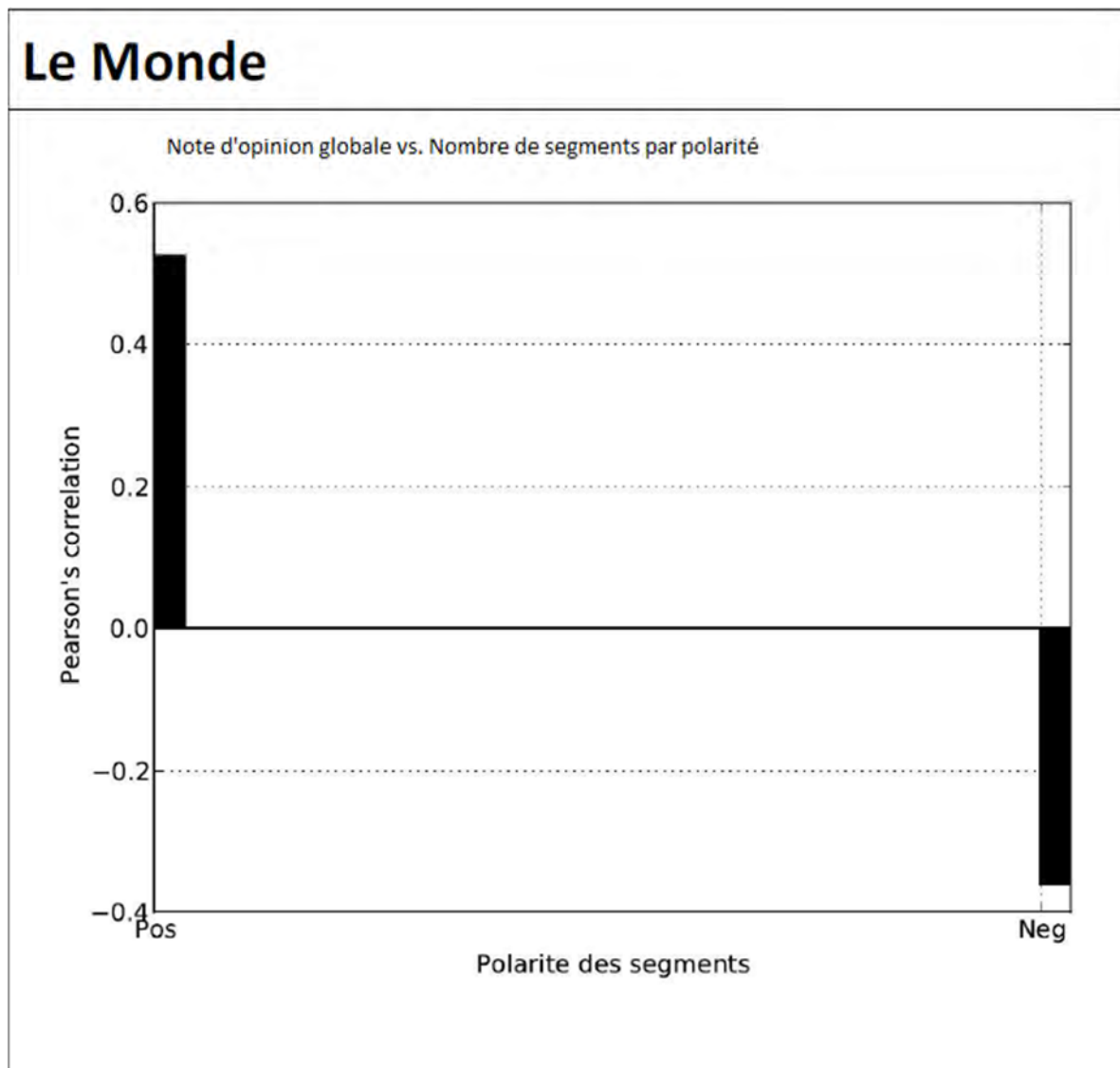


Figure 2.13 : Corrélation entre note globale et nombre de segments pondéré par leurs intensités - LeMonde

Nous pouvons observer que le biais de politesse est effectivement présent sur les corpus, en particulier sur le corpus de réactions aux articles de presse. Il est également présent sur le corpus de commentaires web, mais est beaucoup moins marqué.

#### 2.5.2.5. Annotation discursive : analyse quantitative

Au total, 2539 relations de discours ont été annotées pour le corpus de commentaires web, et 1536 pour celui de réactions aux articles de presse. Afin de mieux comprendre le contexte d'apparition des opinions dans chaque genre de corpus, nous nous sommes intéressés à la fréquence d'apparition de chacune des relations de discours, présentées dans le Tableau 2.13 ci-après :

	Reviews	Le Monde
<b>Explication</b>	2.9 %	3.7 %
<b>But</b>	1.2 %	3.3 %
<b>Résultat</b>	4.3 %	3.5 %
<b>Parallèle</b>	0.7 %	1.2 %
<b>Contraste</b>	10.0 %	7.6 %
<b>Continuation</b>	30.0 %	23.0 %
<b>Conditionnelle</b>	0.6 %	2.3 %
<b>Alternative</b>	0.4 %	1.6 %
<b>Attribution</b>	3.1 %	4.3 %
<b>Arrière-plan</b>	1.1 %	0.8 %
<b>Narration</b>	0.8 %	1.0 %
<b>Flashback</b>	0.1 %	0.1 %
<b>Encadrement</b>	2.2 %	2.6 %
<b>Elaboration</b>	12.0 %	11.0 %
<b>E-Elaboration</b>	6.9 %	9.8 %
<b>Commentaire</b>	18.0 %	24.0 %
<b>Correction</b>	1.3 %	0.4 %
<b>Unknown<sup>22</sup></b>	3.0 %	0.8 %

Tableau 2.13 : Fréquence de chaque relation de discours annoté

Les relations les plus fréquentes dans les corpus sont les relations de continuation, ce qui était attendu. Les relations d'explication, d'Elaboration, d'Elaboration d'entité, de Commentaire, et de Contraste sont également présentes en quantité relativement importante. Ces résultats sont globalement stables d'un genre à un autre, à l'exception de la relation *Unknown*, qui est sensiblement plus fréquente dans les réactions aux articles de presse. Ceci peut être dû à deux facteurs :

- La qualité de rédaction : les réactions aux articles de presse sont en général bien écrites, et comportent donc une structure discursive assez claire. Les commentaires web peuvent à l'inverse être écrits par des rédacteurs assez jeunes, et rédigés de manière plus simple, moins élaborée.
- L'ordre d'annotation : les annotateurs ont commencé par les annotations de commentaires web, et ont fini par les annotations de réactions aux articles de presse. Par conséquent, leur expérience du domaine étant plus élevée lors de l'annotation du second corpus, ils ont eu moins recours à la relation *Unknown*.

Le tableau suivante présente la fréquence des relations par type de segments source et cible (i.e. segment élémentaire – EDU – ou complexe – CDU).

---

<sup>22</sup> Cette relation sert à indiquer que l'annotateur pressent une relation entre les deux segments, mais n'arrive pas à en déterminer la nature. Cette relation est considérée par défaut comme coordonnante.

	Reviews			Le Monde		
	EDU-EDU	EDU-CDU <sup>23</sup>	CDU-CDU	EDU-EDU	EDU-CDU <sup>24</sup>	CDU-CDU
<b>Explication</b>	37 %	43 %	20 %	51 %	27 %	22 %
<b>But</b>	79 %	17 %	4 %	75 %	25 %	0 %
<b>Résultat</b>	31 %	42 %	27 %	35 %	37 %	28 %
<b>Parallèle</b>	35 %	24 %	41 %	73 %	20 %	7 %
<b>Contraste</b>	33 %	33 %	34 %	31 %	31 %	38 %
<b>Continuation</b>	68 %	13 %	19 %	72 %	16 %	12 %
<b>Conditionnelle</b>	61 %	31 %	8 %	50 %	36 %	14 %
<b>Alternative</b>	70 %	30 %	0%	70 %	15 %	15 %
<b>Attribution</b>	69 %	29 %	2 %	42 %	52 %	6 %
<b>Arrière-Plan</b>	68 %	20 %	12 %	50 %	50 %	0 %
<b>Narration</b>	44 %	39 %	17 %	42 %	17 %	41 %
<b>Flashback</b>	100 %	0 %	0 %	0 %	100 %	0 %
<b>Encadrement</b>	35 %	58 %	7 %	59 %	31 %	10 %
<b>Elaboration</b>	15 %	49 %	36 %	17 %	55 %	28 %
<b>E-Elaboration</b>	61 %	33 %	6 %	60 %	35 %	5 %
<b>Commentaire</b>	33 %	38 %	29 %	21 %	38 %	41 %
<b>Correction</b>	33 %	40 %	27 %	40 %	40 %	20 %
<b>Unknown</b>	68 %	9 %	23 %	50 %	30 %	20 %

Tableau 2.14 : Répartition fréquentielle des types de segments reliés par type de relation

Nous pouvons remarquer que les relations ont majoritairement au moins une CDU dans leurs arguments (i.e. elles partent d'une CDU ou arrivent sur une CDU). Ceci souligne l'importance des segments complexe dans les types de documents annotés. Ceci est d'autant plus vrai que cela concerne des relations jouant intuitivement un rôle crucial pour la sémantique d'opinion du document : par exemple, moins de 35 % des contrastes repérés relient deux EDU, et ceci pour les deux genres de corpus. Il en va de même pour les relations de Résultat, d'Elaboration, ou encore de Commentaire.

#### 2.5.2.6. Annotation discursive : lien avec la subjectivité

Nous nous sommes ensuite demandé quels étaient les liens entre les relations repérées et la subjectivité des segments reliés. Plus précisément : quel type de relations relient des segments source et cible tous deux subjectifs / tous deux objectifs ? Le Tableau 2.15 suivant présente, par corpus, pour les 10 relations les plus fréquentes, la proportion d'instance préservant la subjectivité (i.e. source et cible sont toutes deux porteuses d'opinion (SE, SI, SEI) ou non porteuses d'opinion (O, SN)).

<sup>23</sup> Cette catégorie inclut les relations EDU -> CDU ainsi que les relations CDU -> EDU.

<sup>24</sup> Cette catégorie inclut les relations EDU -> CDU ainsi que les relations CDU -> EDU.

	Reviews		Le Monde	
	Subjectivité préservée	Subjectivité non préservée	Subjectivité préservée	Subjectivité non préservée
<b>Continuation</b>	81 %	19 %	79 %	21 %
<b>Commentaire</b>	61 %	39 %	75 %	25 %
<b>Elaboration</b>	50 %	50 %	82 %	18 %
<b>Contraste</b>	76 %	24 %	76 %	24 %
<b>E-Elaboration</b>	63 %	37 %	65 %	35 %
<b>Résultat</b>	81 %	19 %	47 %	53 %
<b>Attribution</b>	14 %	86 %	18 %	82 %
<b>Parallèle</b>	100 %	0 %	73 %	27 %
<b>Explication</b>	76 %	24 %	78 %	22 %
<b>Encadrement</b>	39 %	61 %	47 %	53 %

Tableau 2.15 : Préservation de la subjectivité par type de relation

Seules les attributions et les encadrements ne préservent pas majoritairement la subjectivité dans les deux corpus. Les autres relations préservent en majorité la subjectivité, ce qui est plutôt en accord avec les hypothèses de (Pang and Lee 2004) comme quoi les phrases subjectives ont tendance à apparaître ensemble. Un cas limite est observé pour la relation d'Elaboration dans le cadre du corpus de commentaires web, qui préserve la subjectivité dans 50 % des cas.

#### 2.5.2.7. Annotation discursive : lien avec la polarité

Nous nous sommes enfin demandé quels étaient les liens entre les relations repérées et la polarité des segments reliés. Plus précisément : quel type de relations relie des segments source et cible tous deux positifs / tous deux négatifs ? Le Tableau 2.16 suivant présente, par corpus, pour les 10 relations les plus fréquentes, la proportion d'instance préservant la polarité.

	Reviews		Le Monde	
	Polarité préservée	Polarité non préservée	Polarité préservée	Polarité non préservée
Continuation	97 %	3 %	90 %	10 %
Commentaire	82 %	18 %	96 %	4 %
Elaboration	100 %	0 %	100 %	0 %
Contraste	15 %	85 %	59 %	41 %
E-Elaboration	71 %	29 %	86 %	14 %
Résultat	100 %	0 %	100 %	0 %
Attribution	50 %	50 %	100 %	0 %
Parallèle	100 %	0 %	100 %	0 %
Explication	80 %	20 %	83 %	17 %
Encadrement	100 %	0 %	86 %	14 %

Tableau 2.16 : Préservation de la subjectivité par type de relation

Les résultats observés pour la polarité sont également plutôt conformes aux hypothèses de départ, à l'exception notable de la relation de *Contraste* pour le corpus de réactions aux articles de presse, où l'on observe une forte proportion de polarité préservée entre le segment source et le segment cible. Ceci peut s'expliquer par certaines confusions des annotateurs dans le cadre de

tournure de phrase du type : "*Les finances sont mauvaises, mais la vie culturelle l'est aussi.*", relativement courantes dans ce corpus, et où le marqueur "*mais*" induit en erreur l'annotateur : la relation ici est bien un *Parallèle*, et non un *Contraste*.

## 2.6. Conclusion

Nous avons présenté dans ce chapitre un modèle d'annotation de l'opinion sur trois niveaux.

Tout d'abord, au niveau local, où nous repérons l'expression d'opinion et ses attributs (polarité, intensité), le sujet de l'opinion, l'émetteur de l'opinion, et les éventuels opérateurs venant s'appliquer sur l'expression d'opinion, et altérer ses caractéristiques (polarité, intensité, certitude associée). Ensuite au niveau du segment de discours, où nous annotons le type de subjectivité et les attributs d'opinion hérités du contenu ancré dans le segment (expressions d'opinion et opérateurs pour les segments explicites). Enfin, au niveau du document, où nous annotons la note globale associée au document, et où nous relierons l'opinion au graphe de discours SDRT du document.

Ce modèle d'annotation a été validé par une campagne d'annotation, réalisée dans le cadre du projet CASOAR, où ont été doublement annotés deux corpus de genre différents : un corpus de commentaires web (*web reviews*), et un corpus de réactions d'internautes aux articles de presse de l'édition en ligne du journal *Le Monde*. Nous avons pu observer quelques différences entre ces deux genres de documents, et notamment une plus grande importance des opinions implicites pour les réactions aux articles de presse. Nous avons enfin pu analyser l'impact du discours sur l'opinion, et avons observé une forte influence de celui-ci à la fois sur la polarité et sur la subjectivité. Fort de ces résultats, nous pouvons envisager des approches basées sur le discours pour l'analyse de la subjectivité (cf. chapitre 5 au niveau du segment) et la détermination de l'opinion globale (cf. chapitre 7).

Une campagne d'annotation complémentaire pour la langue anglaise est actuellement en phase de finalisation : le corpus choisi est constitué de commentaires web extraits du site Metacritic<sup>25</sup>.

---

<sup>25</sup> <http://www.metacritic.com/>





## 3. Architecture du système : l'analyse fine de l'opinion locale comme support de remontée de l'opinion globale

---

### 3.1. Introduction

Dans le chapitre précédent, nous venons de voir comment structurer l'opinion de manière à l'annoter et à raisonner sur ces annotations. Nous avons également pu voir que cette structure était pertinente relativement au jugement d'annotateurs humains. De ce fait, nous avons isolé comment, à plusieurs niveaux, l'opinion pouvait être extraite, et quels éléments devaient être reconnus pour caractériser celle-ci.

Afin de mettre en pratique cette extraction, nous proposons ici une architecture sous forme de *pipeline* : chaque module exploite les résultats du module précédent afin de produire une nouvelle annotation, ajoutant des informations au niveau de granularité courant, ou remontant l'information à un niveau de granularité plus élevé.

#### 3.1.1. Contexte scientifique

Comme nous avons pu le voir précédemment, extraire une information d'opinion correspondant à un document est une problématique étudiée depuis plusieurs années. Certains travaux se sont attachés à raisonner directement au niveau du document, en proposant une classification à ce niveau : nous pouvons notamment citer les travaux de (Pang, Lee, and Vaithyanathan 2002).

D'autres travaux, plus récents, ont adoptés une approche différente : l'opinion est tout d'abord extraite le plus finement possible au niveau local, afin d'être par la suite remontée au niveau du document, via un *pipeline* de processus. (Hu and Liu 2004) ont ainsi proposé une chaîne de traitement de l'opinion, orientée sur l'extraction d'opinion sur les composantes d'un produit donné dans les commentaires web de consommateurs. Ils se basent pour cela sur une classification au niveau de la phrase, et un module de présentation des résultats agrégeant simplement les opinions de même polarité sur une même caractéristique du produit. (Popescu and Etzioni 2007) ont proposé une chaîne d'extraction d'opinion (OPINE), également orientée sur les caractéristiques de produit, présentant les opinions en les ordonnant par ordre d'intensité. (Ding, Liu, and Yu 2008) proposent une amélioration de ce processus en prenant en compte certains éléments de contexte dans le cas où la polarité observée serait ambiguë : opérateurs de négation, coordinations locales, et marques de contraste entre phrases adjacentes.

Plusieurs travaux se déroulant en parallèle de cette thèse ont également visé à mettre en place une chaîne d'extraction de l'opinion. Pour l'anglais, (Taboada, Voll, and Brooke 2008; Taboada et

al. 2011) proposent la chaîne de traitement SO-CAL, permettant également d'extraire l'opinion au niveau du document à partir d'une prise en compte fine des effets linguistiques des opérateurs. (Somasundaran 2010) présente une architecture centrée sur la notion de relation de discours entre opinions. À partir d'un texte découpé en segments (Dialog Acts), elle propose plusieurs modules visant à reconnaître la polarité au niveau d'un segment de phrase, et à raffiner celle-ci en fonction des interactions discursives entre les segments. (Vernier 2011) a proposé une séquence de modules permettant la reconnaissance d'une opinion au niveau local (i.e. au niveau de l'expression d'opinion), ainsi que le rattachement de celle-ci au sujet de l'opinion.

### 3.1.2. Positionnement et spécificités de nos travaux

Nos travaux se placent dans le cadre d'une analyse fine au niveau local permettant la remontée d'informations au niveau du document. Comme (Somasundaran 2010), nous nous appuyons sur une analyse discursive du texte afin d'améliorer la détection des opinions unitaires au niveau local et de mieux comprendre comment celles-ci contribuent à véhiculer une opinion globale au niveau du document. Comme vu précédemment (cf. chapitre 1), nous nous plaçons dans le contexte théorique de la SDRT (Asher and Lascarides 2003). De ce fait, nos travaux diffèrent de ceux de (Somasundaran 2010), qui s'appuient sur un cadre de travail discursif propre, dédié à l'opinion. L'utilisation de la SDRT nous permet de nous reposer, pour l'analyse du texte, sur un ensemble de résultats théoriques concernant les relations et interactions entre segments de discours. Le choix de la SDRT par rapport à la RST, utilisée notamment par (Taboada, Voll, and Brooke 2008; Heerschop et al. 2011; Zhou et al. 2011), nous place dans le contexte d'une structure discursive en graphe, permettant de modéliser certaines finesses supplémentaires.

### 3.1.3. Objectifs et déroulement du chapitre

Dans ce chapitre, nous allons poser les fondations informatiques qui vont nous permettre d'effectuer automatiquement notre annotation, et de répondre aux besoins spécifiques d'industrialisation de la chaîne. La section 2.2 précisera les spécifications techniques relatives à la chaîne, et décomposera le problème en différentes sous-tâches et modules associés. La section 2.3 abordera les problématiques pratiques de l'implémentation, les principales motivations du choix d'environnement, et la conception du langage de règles utilisé.

## 3.2. Architecture globale du système

La tâche principale à laquelle ce système vise à répondre est de déterminer, pour un document  $Doc$  de sujet principal  $t_{doc}$  du document, la valeur de l'opinion globale,  $ValOp_{doc | t_{doc}}$ .

Pour cela, le système proposé se base sur une analyse poussée au niveau local, et sur une remontée progressive de l'information, notamment via la structure discursive du document. Il suit l'architecture globale suivante :

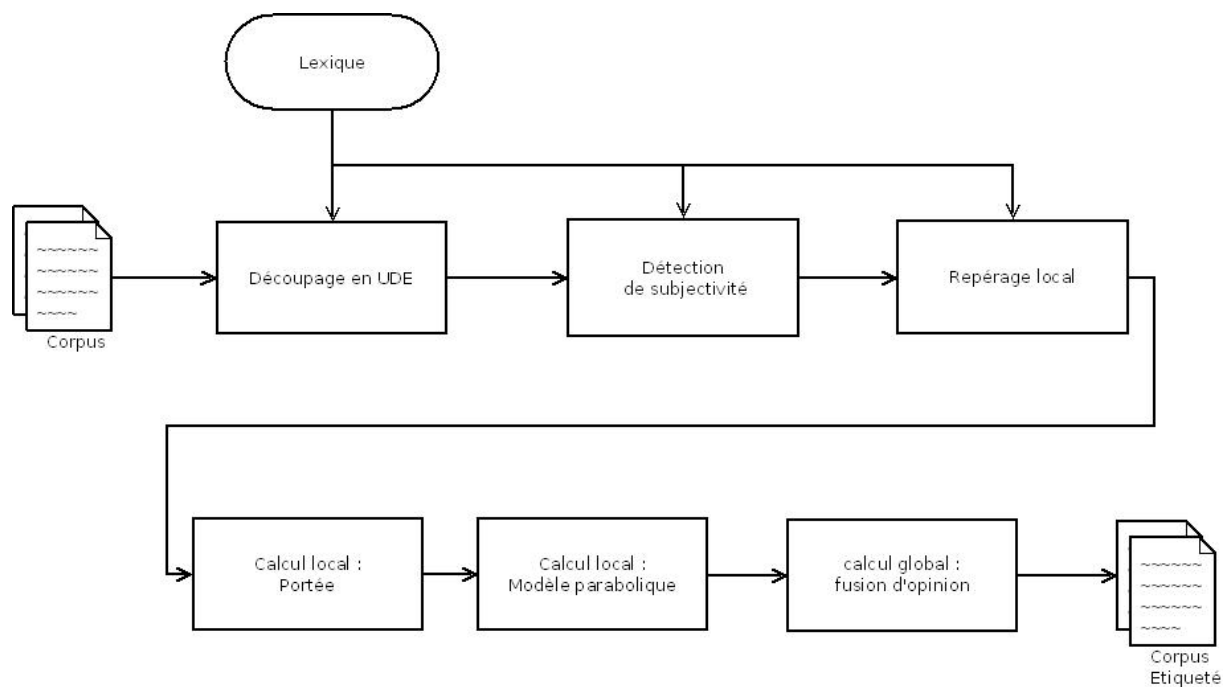


Figure 3.1 : Architecture globale du système

Cette section présente une description de chacun des modules réalisés ou intégrés. Certains de ces modules, correspondant aux travaux de recherche de cette thèse, feront ensuite l'objet d'une description détaillée dans les chapitres suivants.

### 3.2.1. Lexiques

Comme l'illustre la figure précédente, le lexique d'opinion constitue la clé de voûte de la chaîne d'extraction de l'opinion. En effet, tous les modules reposent, directement ou indirectement, sur le repérage des termes d'opinions :

- Le module de segmentation repose directement sur le repérage d'expressions d'opinion afin de proposer des unités discursives fines (cf. section 3.4.2)
- Le module de détection de subjectivité exploite le repérage d'expressions d'opinion en tant que trait pour la classification (cf. chapitre 5)
- Le module de calcul de l'opinion locale repose directement sur la détection de l'opinion au niveau de l'expression, à partir de laquelle sera déterminée l'opinion au niveau du segment
- Le module de calcul de l'opinion globale repose sur l'opinion au niveau du segment, et donc indirectement sur le repérage des expressions d'opinion

Par conséquent, toute erreur dans le lexique sera propagée dans la totalité de la chaîne de traitement.

Le lexique final adopté pour la chaîne de traitement est décrit dans le chapitre 4. Le module de repérage des éléments de ce lexique fonctionne en combinaison avec l'analyse effectuée par l'analyseur Cordial de Synapse Développement (Laurent, Nègre, and Séguéla 2009) : les termes

sont lemmatisés, et désambiguïsés syntaxiquement et sémantiquement suivant leur contexte afin d'opérer une reconnaissance la plus fine possible des expressions d'opinion.

### 3.2.2. Module de segmentation discursive

Nous avons pu voir au chapitre 2 qu'il était possible et utile de remonter l'opinion au niveau du segment de discours. Ce module vise à automatiser le découpage d'un document en segments discursifs.

**Exemple** : texte découpé discursivement, extrait du corpus de commentaires web du projet CASOAR. Les segments sont délimités par des crochets, et numérotés.

```
[Je viens de découvrir le Sushiyaki.]1 [On y mange très bien.]2 [Je vous recommande les sashimis de saumon.]3 [Ils sont meilleurs que ceux]4 [que j'ai mangés le mois dernier au Japon.]5 [Les glaces au thé vert m'ont rappelé mon enfance au pays.]6 [En bref, une cuisine saine]7 [et très savoureuse.]8 [Le restaurant devrait figurer sur le site de l'Office de Tourisme de Toulouse.]9
```

Formellement, cette tâche consiste à déterminer l'ensemble des segments de discours  $Segs_{Doc}$  du document  $Doc$ .

#### 3.2.2.1. Principe

La tâche de segmentation discursive peut être résolue en repérant dans un texte les bornes des différents segments le constituant. En effet, deux segments discursifs peuvent être imbriqués, mais ne peuvent pas se recouper partiellement. Par conséquent, un ensemble bien formé d'ouverture et fermeture de segments équivaut à un et un seul ensemble de segments discursifs.

**Exemple** : Étant donné le repérage automatique des bornes suivant :

```
[Celle-ci, [prudente,] a préféré conserver l'anonymat.]
```

La seule segmentation possible est :

```
[prudente, ]1  
[Celle-ci, prudente, a préféré conserver, en attendant, l'anonymat.]2
```

Le module utilisé dans la chaîne de traitement est une adaptation de celui proposé par (Afantenos et al. 2010) dans le cadre du projet ANNODIS.

#### Classification préliminaire

L'objectif du module de segmentation est de classer chaque mot/*Token* du texte dans l'une des classes suivantes :

- R (Right) : le *Token* marque un début de segment discursif
- L (Left) : le *Token* marque une fin de segment discursif
- N (Nothing) : le *Token* n'est ni un début, ni une fin de segment discursif
- B (Both) : le *Token* est à la fois un début et une fin de segment discursif (i.e. le *Token* constitue un segment discursif d'un seul mot)

En sortie de classification, on obtient un fichier possédant un nombre non balancé d'ouvertures et de fermetures de segments. Les segments sont à partir de là reconstitués, en reliant chaque borne à la borne complémentaire la plus proche.

En pratique, le prototype du classifieur a été réalisé via le package *MEGAM*<sup>26</sup> (classifieur *Maximum Entropy*). Les traits de classification ont été adaptés afin de prendre en entrée l'analyse syntaxique de Synapse Développement : en plus des traits originels proposés par (Afantenos et al. 2010) – N-grammes, Début et fin de phrase, Début et fin de *chunk*, Fonction syntaxique, *Part-of-speech* – le module inclut les deux traits suivants :

- Le *token* considéré marque-t-il le début d'une proposition ?
- Le *token* marque-t-il la fin d'une proposition ?

Ces deux traits supplémentaires permettent un gain de l'ordre de 1 % en f-mesure sur le corpus ANNODIS.

### *Découpage complémentaire sémantique*

Dans le contexte de l'analyse d'opinion, nous sommes intéressés à repérer des relations discursives très locales, permettant ainsi de projeter les relations au niveau des opinions. Par conséquent, dans le cas de clauses locales identifiant des opinions distinctes, nous avons décidé de resegmenter les UDE, comme nous l'avons décrit au chapitre 2.

**Exemples** : La phrase suivante est segmentée suivant la définition issue du projet Annodis (1), et suivant notre définition (2), permettant de capturer la relation locale entre les opinions – ici, relation de continuation entre les segments (2.a) et (2.b).

(2.a) [ Ce film est majestueux et brillant. ]<sub>a</sub> est resegmenté en :

(2.b) [ Ce film est majestueux ]<sub>a</sub> [ et brillant. ]<sub>b</sub>

Ce découpage, plus simple que la tâche de classification présentée ci-avant, est effectué par un jeu de règles, présenté en Annexe A.1. Le taux de segments nécessitant resegmentation est de l'ordre de 12 % sur le corpus de reviews Casoar.

### 3.2.2.2. Résultats

#### *Classification en EDU "Annodis"*

Afin d'évaluer notre classifieur, nous avons tout d'abord effectué une validation croisée sur 10 dossiers (*10-fold cross validation*) sur le corpus Annodis (Péry-Woodley et al. 2009). Ceci nous permet de comparer nos résultats avec ceux obtenus par (Afantenos et al., 2010). Ensuite, nous avons testé l'utilisation du corpus Annodis complet en tant que corpus d'entraînement, et le corpus de commentaires web du projet Casoar en tant que corpus de test. Enfin, nous avons effectué une validation croisée sur le corpus de commentaires web Casoar seul.

---

<sup>26</sup> <http://www.cs.utah.edu/~hal/megam/>

Les tableaux suivants présentent les résultats obtenus pour les classes R, L et N. Les résultats pour la classe B ne sont pas présentés, en raison du nombre trop faible d'instances, tant dans les corpus d'entraînement que dans ceux de tests.

	Rappel	Précision	F1
10-fold cross validation, Annodis	0.858	0.913	0.885
Train. Annodis ; Test Casoar Reviews	0.791	0.927	0.853
10-fold cross validation, Casoar Reviews	0.925	0.942	0.933

*Tableau 3.1 : Résultats pour la classe R (borne ouvrante d'un ou plusieurs segments)*

	Rappel	Précision	F1
10-fold cross validation, Annodis	0.872	0.894	0.883
Train. Annodis ; Test Casoar Reviews	0.752	0.917	0.827
10-fold cross validation, Casoar Reviews	0.941	0.952	0.946

*Tableau 3.2 : Résultats pour la classe L (borne fermante d'un ou plusieurs segments)*

	Rappel	Précision	F1
10-fold cross validation, Annodis	0.976	0.967	0.972
Train. Annodis ; Test Casoar Reviews	0.978	0.926	0.952
10-fold cross validation, Casoar Reviews	0.982	0.977	0.980

*Tableau 3.3 : Résultats pour la classe N (Token n'étant pas une borne de segment)*

Les résultats obtenus sur le corpus Annodis sont cohérents avec ceux obtenus par (Afantenos et al. 2010) sur ces mêmes données, ce qui valide l'adaptation de l'analyse syntaxique de Synapse développement à la tâche. Nous pouvons aussi remarquer que l'utilisation du corpus Annodis comme corpus d'entraînement pour annoter un nouveau corpus induit une chute significative du rappel : le classifieur, une fois entraîné, est donc plutôt sensible aux changements dans le type du corpus. Néanmoins, la validation croisée sur les données des reviews de Casoar donne de très bons résultats : le classifieur semble donc bien adapté au domaine, pourvu que l'entraînement ait été effectué sur un corpus similaire.

### **Segmentation complémentaire "Casoar"**

Nous avons testé la segmentation complémentaire sur le corpus de commentaires web Casoar, dont la segmentation manuelle est compatible avec notre notion d'UDE. Les tableaux suivants présentent les résultats obtenus, respectivement lorsque l'UDE est connue comme à redécouper (information de subjectivité issue de l'annotation manuelle – tableau 3.4), ou déterminée par lexicque (tableau 3.5). Enfin, le Tableau 3.6 décrit les résultats obtenus lorsque l'on corrige préalablement l'orthographe du corpus, de façon automatique. Cette correction orthographique a été obtenue via le correcteur d'orthographe et de grammaire Cordial<sup>27</sup> de Synapse

<sup>27</sup> [http://www.synapse-fr.com/correcteur\\_orthographe\\_grammaire.htm#cordial\\_pro](http://www.synapse-fr.com/correcteur_orthographe_grammaire.htm#cordial_pro)

Développement. Le correcteur a été paramétré pour ne corriger que les erreurs d'orthographe, et la fonctionnalité de correction automatique permet de remplacer automatiquement un terme fautif si le remplacement est jugé statistiquement « sûr » par le moteur de correction.

	Rappel	Précision	F1
Bornes des segments <sup>28</sup>	0.976	0.968	0.972
Reconnaissance globale des UDE <sup>29</sup>	0.821	0.732	0.774

*Tableau 3.4 : Résultats pour la sur-segmentation (information de subjectivité donnée par oracle)*

	Rappel	Précision	F1
Bornes des segments	0.961	0.977	0.969
Reconnaissance globale des UDE	0.751	0.772	0.762

*Tableau 3.5 : Résultats pour la sur-segmentation (information de subjectivité déterminée par lexique)*

	Rappel	Précision	F1
Bornes des segments	0.963	0.979	0.971
Reconnaissance globale des UDE	0.763	0.791	0.777

*Tableau 3.6 : Résultats pour la sur-segmentation après correction orthographique automatique (information de subjectivité déterminée par lexique)*

Nous pouvons remarquer que les résultats obtenus sont comparables à ceux issus de la pré-segmentation par classification. La correction orthographique préalable apporte également une amélioration significative des résultats, ce qui montre la relative sensibilité du module de segmentation à l'orthographe du document.

### 3.2.3. Détection de la subjectivité

Ce module consiste à déterminer si un segment de discours revêt un caractère objectif (O) ou subjectif, explicite (SE, SN) ou implicite (SI), évaluatif (SE, SI) ou non.

Formellement, cette tâche consiste à associer à tout segment *seg* du document une valeur de subjectivité  $subj_{seg} \in \{SE, O, SI, SN\}$ .

Nous proposons au chapitre 4 une méthode de classification automatique des segments de discours, basées sur deux classifieurs SVM parallèles.

### 3.2.4. Repérage local

Ce module, relativement simple étant donné les ressources mentionnées précédemment, consiste à repérer dans le texte expressions d'opinion et opérateurs.

En pratique, l'annotation se fait après analyse syntaxique et désambiguïisation sémantique par l'analyseur de Synapse Développement, à partir des informations de lemme, de nature

<sup>28</sup> Toute classe comprise.

<sup>29</sup> i.e. reconnaissance correcte de chacune des deux bornes pour une UDE.

grammaticale, et de sens présentes dans le lexique. Ce repérage est également effectué en parallèle de la détection de subjectivité, la présence de termes subjectifs étant un des traits considérés par le module.

Les termes ou expressions considérés comme neutre hors contexte (i.e. dont la polarité dépend du contexte) ne sont, de base, pas pris en compte dans l'extraction. Ce choix, qui maximise la précision de la chaîne de traitement, n'est pas l'unique possibilité proposée dans la littérature. Ainsi, (Ding, Liu, and Yu 2008) ont proposé des règles contextuelles associées à la détection du sujet de l'opinion (caractéristiques de produits commentés) permettant de désambigüiser dynamiquement certaines entrées. Néanmoins, deux éléments limitent fortement l'impact sur le rappel de ce choix :

- La désambigüisation sémantique limite fortement le nombre d'entrées neutres hors contexte : l'information de sens introduit en général l'information de polarité
- Dans le cas où la polarité dépend du domaine, il est possible d'introduire un lexique complémentaire, prenant le pas sur le lexique principal, décrivant les entrées spécifiques au domaine.

### 3.2.5. Module de calcul de l'opinion au niveau du segment

Le calcul local de l'opinion consiste à déterminer, à partir des expressions d'opinion repérée dans un segment, et des opérateurs présents dans la phrase, les valeurs de polarité, d'intensité, et de certitude du segment.

Cette tâche peut se diviser en deux parties, la détermination de la portée des opérateurs, et le modèle computationnel de l'application d'une séquence d'opérateurs sur une expression d'opinion.

#### *Portée d'un opérateur*

Déterminer la portée d'un opérateur est un problème syntaxique relativement complexe. Nous aborderons ce sujet dans la première partie du chapitre 5.

Formellement, cette tâche consiste à associer, pour chaque élément  $w_{op} \in W_{Doc}$ , l'ensemble  $\{oper \in Opers_{Doc} \mid oper \in opers_{w_{op}}\}$  des opérateurs s'appliquant sur cette expression, et à ordonner cet ensemble pour en déduire la liste  $opers_{w_{op}}$  des opérateurs dans leur ordre d'application.

Exemple :

- « Ce restaurant n'est pas très bon. »

Dans cet exemple, nous avons  $w_{op} = \text{"bon"}$ , et  $opers_{w_{op}} = \{\text{"ne...pas"}, \text{"très"}\}$ . Nous pouvons également représenter cet ordre d'application de manière fonctionnelle par :  $\text{ne...pas(très(bon))}$ .



### **Modèle computationnel : modèle parabolique**

Une étude linguistique des effets des opérateurs, et le modèle proposé pour appliquer les opérateurs – modèle parabolique – seront abordés en seconde partie du chapitre 5.

Formellement, cette tâche consiste, pour tout segment discursif  $seg$  subjectif, à calculer une valeur d'opinion  $ValOp_{seg|t_i}$  à partir de chaque opinion unitaire  $Op_i = (w_i, t_i, h_i, ops_i)$ .

#### **3.2.6. Fusion au niveau du document : calcul d'une opinion globale**

La fusion globale consiste à présenter un score d'opinion relatif à l'ensemble du document (score "résumé"). Nous étudierons diverses stratégies afin de calculer ce score dans le chapitre 6.

Formellement, cette tâche consiste à calculer la valeur d'opinion globale  $ValOp_{doc|t_{doc}}$  à partir de l'ensemble  $\{ValOp_{seg|t_i} \mid seg \in Segs_{Doc}\}$ , et de l'ensemble des relations de discours  $RelS_{Doc}$ .

#### **3.2.7. Remarques sur la détection du sujet de l'opinion**

Certains modules, en particulier la fusion au niveau du document, peuvent exploiter le lien entre l'opinion et son sujet. La problématique de recherche du sujet de l'opinion est un problème de recherche ouvert, qui ne sera pas traité dans le cadre de cette thèse. Ainsi, dans les chapitres concernés, les sujets seront considérés comme donnés, soit par annotation manuelle, soit par module de détection "parfait".

En pratique, dans le contexte de la chaîne de traitement, le repérage du sujet de l'opinion est effectué *ad hoc*, en partant d'une ontologie et/ou d'une liste de sujets possibles pour chaque document.

### **3.3. Environnement technique : vers un langage de règles compatible avec une chaîne industrialisée**

La plate-forme d'exécution est l'environnement sur lequel tournera la chaîne de traitement. Ce choix est important, car il conditionnera plusieurs éléments :

- Facilité de prototypage pour les expérimentations
- Caractère industrialisable de la plate-forme finale
- Facilité de maintenance

La facilité de prototypage est une notion en partie subjective, mais repose aussi sur des éléments concrets : dispose-t-on facilement d'outils complémentaires accessibles via la plate-forme, ou faut-il développer des passerelles spécifiques ? Le langage utilisé permet-il la manipulation directe d'abstractions spécifiques au traitement du langage (par exemple, la notion de mots ou Tokens) ?

Le caractère industrialisable de la chaîne demande de répondre à plusieurs questions. Tout d'abord, la portabilité : est-ce que la chaîne sera utilisable dans l'environnement technique spécifique d'un client ? Ensuite, la légèreté : le client devra-t-il installer des plates-formes

intermédiaires ? De même, il faut considérer les problématiques d'optimisation : la chaîne de traitement s'exécutera-t-elle dans un temps raisonnable ? A-t-on la possibilité d'effectuer des optimisations sur le code exécuté, ou est-on tributaire de composants peu efficaces ? Une autre problématique est celle du passage à l'échelle (scalabilité) : peut-on traiter de gros fichiers ? Peut-on traiter un nombre important de fichiers ? Une exécution en parallèle est-elle possible ? Enfin se pose le problème de la licence : les composants utilisés peuvent-ils être utilisés dans le cadre d'une chaîne propriétaire (par exemple, avec une licence de type LGPL<sup>30</sup>), où l'utilisation des composants impliquent-ils que la chaîne complète soit Open Source (par exemple, avec une licence de type GPL<sup>31</sup>) ?

La facilité de maintenance regroupe deux aspects : la maintenance corrective (i.e. le fait de corriger ou d'améliorer les fonctionnalités déjà implémentées), et l'ajout de nouvelles fonctionnalités. Contrairement à ce que l'on pourrait attendre intuitivement, la facilité de maintenance est tout à fait distincte de la facilité de prototypage. En effet, même si les éléments facilitant le prototypage auront naturellement tendance à faciliter la maintenance, d'autres éléments entrent en compte. Est-ce que la technologie utilisée est facilement maîtrisable, y compris pour un non informaticien ? L'ajout d'une fonctionnalité peut-elle se faire sans impacter les fonctionnalités existantes ?

Cette section a pour objectif de répondre à ces questions pour les différents choix qui s'offrent à nous, et ainsi de déterminer un choix si possible optimal pour constituer la plate-forme d'extraction d'opinion.

### 3.3.1. Plate-forme d'exécution

#### 3.3.1.1. Présentation de l'existant

Nous présentons dans cette section trois plateformes existantes permettant d'effectuer des traitements chaînés sur des données textuelles :

- La plate-forme *GATE*
- La plate-forme *UIMA*
- Le framework interne de Synapse Développement

#### ***GATE***

*GATE (General Architecture for Text Engineering)*<sup>32</sup> est une plate-forme dédiée au traitement du langage développée par l'université de Sheffield (Cunningham, Maynard, and Bontcheva 2011).

Dans le cadre du projet Rhécitas (Tanguy et al. 2009), le composant d'analyse syntaxique de Synapse Développement a été adapté pour *GATE*. En pratique, il consiste en un plugin *GATE* effectuant la passerelle vers la bibliothèque C++ d'étiquetage. Les entrées et sorties de ce plugin

---

<sup>30</sup> <http://www.gnu.org/copyleft/lesser.html>

<sup>31</sup> <http://www.gnu.org/licenses/gpl.html>

<sup>32</sup> <http://gate.ac.uk/>

ont été conçues pour pouvoir se substituer au TreeTagger (Schmid 1994; Schmid 1995), présent dans la distribution de GATE.

Le Tableau 3.7 présente les avantages et inconvénients de cette plateforme.

Prototypage	Outils disponibles	Excellent : de nombreux modules ont été développés pour ou adaptés à GATE, y compris certains composants Synapse. L'interface comprend directement des outils d'évaluation (précision, rappel, F1)
	Abstraction des objets	Excellente : la plate-forme a été développée pour manipuler des documents
Industrialisation	Portabilité	Bonne (versions 4 et ultérieures) : développée en Java
	Légèreté	Mauvaise : nécessite l'installation de la plate-forme complète
	Possibilités d'optimisation	Mauvaise : prévue pour le prototypage, le temps d'exécution d'une chaîne est parfois long
	Scalabilité	Moyenne : peu de maîtrise sur les tailles maximales de fichiers et de corpus.
	Licence	licence LGPL : logiciel Open-Source utilisable dans une chaîne comprenant des composants propriétaires
Maintenance	Facilité	Mitigée (cf. section suivante sur Jape) Pour : langage Jape permet une maintenance corrective aisée Contre : il est fréquemment nécessaire de développer certaines parties en Java – connaissance de l'API interne nécessaire.
	Adaptabilité	Excellente : les plugins, qu'ils soient développés en Jape ou en Java, sont indépendants

Tableau 3.7 : Analyse des avantages et inconvénients de GATE

En conclusion, GATE présente des avantages très intéressants pour le prototypage et dans une moindre mesure la maintenance, de part ses nombreux modules disponibles et son langage de règles. Néanmoins, dans le contexte d'une thèse CIFRE avec perspectives d'industrialisation, GATE présente de gros inconvénients dont il faut tenir compte, notamment en matière d'optimisation et de légèreté de déploiement.

### Apache UIMA

UIMA (Unstructured Information Management Architecture)<sup>33</sup> est une plate-forme de traitement de données non structurées, et en particulier de traitement du langage. Historiquement, le développement d'UIMA a commencé chez IBM, avant que le projet ne passe sous le contrôle de la fondation Apache.

<sup>33</sup> <http://uima.apache.org/>

Afin d'évaluer la plate-forme UIMA, nous avons effectué un portage du plugin d'analyse syntaxique GATE vers l'environnement UIMA. Celui-ci est donc disponible de la même manière que pour GATE.

Le Tableau 3.8 analyse les avantages et inconvénients de cette plateforme.

Prototypage	Outils disponibles	Bon : même si moins de modules sont disponibles de base pour UIMA que pour GATE, de nombreux projets Open Source mettent à disposition des passerelles vers UIMA.
	Abstraction des objets	Excellente : la plate-forme a été développée pour manipuler, entre autres possibilités, des documents
Industrialisation	Portabilité	Bonne : chaîne en Java ou C++, disponible Linux/Windows
	Légèreté	Plutôt mauvaise : nécessite l'installation de la plate-forme complète, qui peut néanmoins se faire oublier à l'exécution.
	Possibilités d'optimisation	Correcte : l'exécution des modules peut être efficace, l'API fournissant des solutions d'optimisation (entre autres, possibilités d'adapter les index d'annotations)
	Scalabilité	Correcte - Moyenne : peu de maîtrise sur les tailles maximales de fichiers et de corpus. Exécution parallèle envisageable.
	Licence	Apache License 2.0 : compatible avec une utilisation dans une chaîne propriétaire
Maintenance	Facilité	Mauvaise : nécessité de connaître l'API interne pour maintenir les modules. Langage de règles limité à la déclaration d'expressions régulières.
	Adaptabilité	Excellente : les modules sont indépendants.

Tableau 3.8 : Analyse des avantages et inconvénients de Apache UIMA

UIMA minimise les défauts de GATE concernant l'industrialisation tout en gardant l'avantage d'une plate-forme intermédiaire et des modules disponibles. Néanmoins, la lourdeur de l'installation et l'absence de langage de règles constituent des freins importants.

### *Environnement interne Synapse Développement*

Synapse Développement développe plusieurs composants logiciels de traitement du langage pour un usage industriel. Parmi ceux-ci, le composant d'analyse syntaxique « Etiquette » permet l'appel, via fonction callback C++, de code métier spécifique. Les modules exécutés dans ce contexte ont accès directement aux données d'analyse syntaxique, sans passer par l'export des sorties en

fichier *csv*<sup>34</sup>. Grâce à ce mécanisme, il est donc possible d'intégrer directement la chaîne de traitement au module d'étiquetage.

Le Tableau 3.9 présente les avantages et inconvénients de cet environnement.

Prototypage	Outils disponibles	Mitigée : Pour : accès direct aux modules de Synapse Développement Contre : des passerelles sont à développées pour la plupart des outils externes.
	Abstraction des objets	Excellente : les structures de données internes permettent la gestion de texte.
Industrialisation	Portabilité	Correcte : chaîne en C++, disponible Linux/Windows sous réserve de non-utilisation de fonctions systèmes dans les modules spécifiques d'extraction d'opinion.
	Légèreté	Bonne : ne nécessite que l'installation de la chaîne de traitement.
	Possibilités d'optimisation	Excellente : accès direct aux structures internes, permettant une optimisation poussée
	Scalabilité	Bonne : maîtrise de la taille maximale des fichiers (segmentation en sous-fichier envisageable). Exécution parallèle envisageable.
	Licence	OK dans le contexte de la thèse CIFRE.
Maintenance	Facilité	Mauvaise : de base, nécessité de maintenir les modules en C++ avec l'API interne.
	Adaptabilité	Moyenne : les modules sont compilés avec le reste de la chaîne.

Tableau 3.9 : Analyse des avantages et inconvénients de l'environnement interne Synapse

L'utilisation de la chaîne de traitement propriétaire possède de nets avantages en termes d'industrialisation : la chaîne de traitement est plus légère, et les modules sont optimisables en termes de performances. L'aspect propriétaire de la chaîne peut être un inconvénient suivant le point de vue : dans le contexte de la thèse CIFRE, cela est tout à fait en accord avec les spécifications. Néanmoins, des problématiques de maintenance sont à anticiper.

### Conclusion

Dans le cadre de cette thèse, afin de réaliser les prototypes permettant d'évaluer indépendamment chacun des étages de la chaîne de traitement, l'utilisation d'une plate-forme externe semble adaptée. Les expérimentations présentées dans le chapitre 3 sont réalisées sous GATE, et celles présentées dans le chapitre 4 sont réalisées sous UIMA.

Dans une optique d'industrialisation, une chaîne basée directement sur l'environnement interne de Synapse Développement semble par contre plus adaptée aux problématiques clients. Il reste

<sup>34</sup> *Comma Separated Values*, format de fichier permettant de décrire simplement un tableau de données en séparant le contenu de chaque cellule par un séparateur donné, généralement une virgule, un point-virgule, ou une tabulation.

néanmoins à améliorer les questions de maintenance, afin de permettre à un non informaticien (par exemple, quelqu'un disposant d'une formation de linguistique pure) de corriger *a posteriori* certains défauts dans le comportement. Pour cela, nous allons étudier dans la section suivante les différentes possibilités de langage de règles associé à la chaîne.

### 3.3.2. Langage de règles

Afin de faciliter le travail de développement et de maintenance des modules, notamment liés à l'opinion, il est important de disposer d'un langage de règles permettant de décrire efficacement les modèles linguistiques développés.

Pour évaluer le langage de règles, il convient de regarder les éléments suivants :

- Adéquation au problème : le langage est-il bien adapté pour raisonner sur du langage naturel ?
- Simplicité d'utilisation : le langage est-il intuitif ? Ce critère est très subjectif, je n'indiquerai dans cette section que mon ressenti personnel, à titre informatif.
- Expressivité du langage : est-ce que le langage permet de décrire efficacement les comportements observés ?
- Portabilité du code : est-ce que le langage peut être interprété ou compilé vers plusieurs plates-formes ?

Cette section passe en revue différentes possibilités de langage de règles disponibles. D'autres existent, mais, à notre connaissance, ne sont pas spécifiquement adaptés au traitement du langage.

#### 3.3.2.1. RegExp Annotator (UIMA)

La plate-forme UIMA permet de déclarer des règles simples de traitement du langage naturel via le module "RegExp Annotator" (annoteur d'expressions régulières). Celui-ci, via un XML de description, permet d'annoter des éléments repérés dans un texte via des expressions régulières sans passer par un module UIMA Java ou C++.

#### *Syntaxe*

Le RegExp Annotator utilise des fichiers XML décrivant les différentes expressions régulières devant être satisfaites, et les annotations à appliquer. La syntaxe du XML de description est disponible à l'adresse :

<http://uima.apache.org/downloads/sandbox/RegexAnnotatorUserGuide/RegexAnnotatorUserGuide.html>

#### *Conclusion*

Le tableau ci-après récapitule les avantages et inconvénients de l'annoteur d'expressions régulières d'UIMA.

Adéquation	Moyenne : niveau d'abstraction limité
Simplicité	Bonne / Moyenne : Intuitif pour toute personne familière avec les expressions régulières Le format XML est, à mon avis, un peu lourd à manipuler
Expressivité	Moyenne : il devient vite difficile d'exprimer des règles complexes seulement à l'aide d'expressions régulières.
Portabilité	Correcte : pour peu que l'on porte l'analyseur de fichier XML, le code pourrait être portable.

Tableau 3.10 : Avantages et inconvénients du RegExp Annotator d'UIMA

### 3.3.2.2. Jape

Jape (Cunningham, Maynard, and Tablan 2000) est le langage de règles associé à la plate-forme GATE. Il s'agit d'un langage interprété permettant de raisonner sur les annotations stockées au sein de la plate-forme GATE par les précédents modules d'une chaîne de traitement, et de produire de nouvelles annotations.

Jape, comme la plate-forme GATE, est orienté traitement du langage naturel, et est donc bien adapté pour des problématiques d'extraction de l'opinion. Il est de plus relativement simple à lire, ce qui simplifie les opérations de maintenance et de débogage.

Il existe deux *plugin* GATE permettant d'exécuter du code Jape au sein de la plate-forme : le Montreal Transducer et le Jape Compiler. Ces deux plugins varient légèrement dans leurs comportements<sup>35</sup> : un même fichier Jape peut donc donner des résultats différents suivant la version de l'interpréteur ou le compilateur utilisé. Depuis la version 5.1, le Montreal Transducer n'est plus maintenu dans les répertoires chargés par défaut par GATE.

#### Syntaxe

La syntaxe Jape est complètement orientée traitement du langage, et permet de définir des règles se basant sur les annotations des modules précédents.

Un guide de la syntaxe Jape est disponible sur le site internet de la plate-forme GATE à l'adresse suivante : <http://gate.ac.uk/sale/tao/splitch8.html#x12-2060008>

#### Limitations

En dépit de ces avantages, Jape possède quelques limitations gênantes. En particulier, s'il est possible d'affecter à une variable un élément repéré dans la partie condition de la règle, il n'est pas possible d'y faire référence dans la suite de la condition. Cela limite l'expressivité du langage : par exemple, il n'est pas possible d'écrire une condition où l'on vérifie que deux mots consécutifs partagent le même trait sémantique. Cela force le concepteur de la règle à effectuer un décrochage vers du code Java, prévu dans le langage, pour effectuer directement, dans la partie action, la vérification nécessaire.

<sup>35</sup> <http://gate.ac.uk/sale/tao/splitch8.html#x12-2480008.10>

**Exemple** : Règle Jape permettant d'apparier plusieurs annotations consécutives possédant la même valeur pour l'attribut "numProp" (regroupement de termes appartenant à la même proposition). Faute de moyen de faire référence à la valeur numProp du premier objet rencontré, cette règle doit être dupliquée pour chaque valeur possible de numProp (ici, numProp doit être égal à 1) si l'on ne souhaite pas faire intervenir de code Java.

```
Rule: PropRule1
(
    ({Token.numProp=="1"})+
    (PONCT)?
):prop
-->
:prop.Proposition = {kind="prop", rule=PropRule,
Verb=:prop.Token.verbeProp, Num=1}
```

Le code Java ainsi écrit exploite l'API interne de GATE, notamment pour accéder aux annotations. Celle-ci, bien que tout à fait fonctionnelle, est bien évidemment spécifique à la plate-forme. Par conséquent, le code Jape avec décrochage Java est lié à GATE, et est donc très difficilement portable vers une plate-forme différente.

### Conclusion

Le Tableau 3.11 présente les avantages et inconvénients de l'utilisation de Jape dans le cadre de la constitution de la chaîne d'extraction.

Adéquation	Excellente : langage spécifique permettant de déclarer des règles à un bon niveau d'abstraction.
Simplicité	Bonne / Moyenne : le langage Jape en lui-même est tout à fait intuitif les décrochages Java sont fréquents, et nécessitent la connaissance de l'API interne Java
Expressivité	Moyenne : décrochages Java nécessaires
Portabilité	Mauvaise : fichiers de règles spécifiques à GATE : les décrochages sont directement écrits suivant l'API interne GATE.

Tableau 3.11 : Avantages et inconvénients de Jape

### 3.3.2.3. Myéline : un langage de règles à visée industrielle

Myéline est un langage de règles développé dans le cadre de cette thèse et du projet CASOAR, soutenu par la DGA. Spécifique au traitement du langage, il répond à un besoin de Synapse Développement allant au-delà de l'analyse d'opinion : il permet, comme Jape, de déclarer des règles métiers exploitant des primitives sur le langage – analyse syntaxique, lexiques – ce qui permet de développer des modules dont la maintenance et l'évolution sont grandement simplifiées. Le principal avantage de Myéline est sa grande expressivité : le fait de pouvoir faire appel aux variables dès leur affectation dans la partie *conditions* permet de corriger le défaut de Jape relevé en section précédente.

Le langage Myéline a été développé prioritairement avec pour plate-forme cible le module d'analyse syntaxique de Synapse Développement : le prototype du compilateur Myéline a été



développé pour cette plate-forme cible. Dans une optique d'industrialisation, une compatibilité avec UIMA est également prévue, mais n'est pas encore disponible.

### *Éléments de syntaxe*

Une règle Myéline se compose de deux parties. La partie condition regroupe les contraintes sur le texte rencontré pour déclencher l'application de la règle. La partie action décrit ce qui est effectué lorsque la règle est déclenchée – typiquement, la création d'une annotation.

#### **Exemple** : Structure d'une règle Myéline

```
Rule: <RuleName>
{
    <Condition>
}
=>
{
    <Action>
}
```

L'unité de base d'une condition de règle Myéline est le Token. La partie condition est séparée en plusieurs sous-conditions, chacune entre parenthèses. Chaque nouvelle sous-condition concerne le Token suivant immédiatement celui concerné par la condition précédente.

#### **Exemple** : Détail structure condition

```
Rule: <RuleName>
{
    (<Contenu Sous-Condition 1 - concerne le premier Token>)
    (<Contenu Sous-Condition 2 - concerne le second Token>)
    ...
    (<Contenu Sous-Condition N - concerne le N-ième Token>)
}
=>
...
```

Dans chaque sous-condition, il est possible de faire référence au Token courant via le mot-clé Token. Plusieurs primitives lui sont associées : entre autres, pos (nature grammaticale), fonction (fonction syntaxique), lemma (lemme)... Le tableau suivant regroupe les principales primitives du langage associées à un Token.

Token.pos	Part of Speech (nature grammaticale)
Token.lemma	Lemme du mot
Token.function	Fonction grammaticale
Token.startOffset	Offset de début du mot
Token.endOffset	Offset de fin du mot
Token.string	Chaîne de caractères du mot
Token.numProp	Numéro de la proposition en cours

Tableau 3.12 : Quelques primitives usuelles du Token

Un Token ou un trait peut être stocké dans une variable. Une fois une variable initialisée, celle-ci peut-être utilisée directement : par conséquent, il est possible de vérifier dès la partie condition des propriétés entre Tokens successifs.

**Exemple** : Utilisation de variable pour résoudre le problème de regroupement de proposition. Syntaxiquement, l'opérateur ":" permet l'affectation dans une variable, et les variables sont identifiés par un '\$'. L'opération d'affectation présente dans la sous-condition étoilée (fermeture de Kleene (Ebbinghaus, Flum, and Thomas 1994, 656) est réalisée pour chaque nouveau Token découvert correspondant à la contrainte. La directive #define permet de définir un type d'Annotation, la directive #RULENAME est une macro donnant le nom de la règle en cours.

```
#define Proposition(num, ruleName)

Rule: NumProp
{
    (Token:$first && Token.numProp:$num)
    (Token:$last && Token.numProp==$num)*
    (Token:$last && Token.pos==PCT)?
}
=>
{
    Proposition($first.startOffset, $end.endOffset, $num, #RULENAME);
}
```

Chacun des objets manipulés possède un type : le tableau suivant décrit les types Myéline, et leur équivalent en C++ (plate-forme cible Synapse Etiquette). Le type de chaque variable est calculé dynamiquement.

Type Myéline	Type C/C++
Token	int (identifiant interne)
Int	int
String	char *
Bool	int ( 0 ou 1 )
Char	char
StoredTag	myeline::Annotation *
List	myeline::List *

Tableau 3.13 : Types Myélines et leurs équivalents pour la plate-forme cible Synapse Étiquette

Il est possible de faire appel, dans la partie *condition* comme dans la partie *action*, à des portions de code source externe, spécifique à la plate-forme cible : cette possibilité est nécessaire afin de permettre des fonctionnalités avancées n'ayant pas vocation à faire partie du langage, comme par exemple une connexion à une base de données externes. Cet appel est possible via la déclaration de fonctions externes. Le type des paramètres comme celui de la valeur de retour de la fonction doivent être déclarés en Myéline, ce qui permet l'intégration directe de la fonction au langage. Ainsi, si l'on dispose d'une implémentation spécifique à chaque plate-forme pour toutes les fonctions externes d'un module Myéline, le même code Myéline est compilable vers chacune de ces plates-formes.

**Exemple** : Reprise de la règle Myéline précédente, avec ajout d'une fonction externe permettant de tracer l'application de la règle (par exemple, pour tracer dans un journal l'application des règles). Syntactiquement, la directive `import` permet de déclarer une fonction externe.

La fonction externe de cet exemple – `TraceRule` – doit exister pour la plate-forme cible, mais le concepteur de la règle n'a pas à maîtriser cette implémentation : celle-ci peut avoir été effectuée en amont par un informaticien – de même, cet informaticien n'a pas à maîtriser les détails linguistiques des règles appelantes.

```
#define Proposition(num, ruleName)
import Bool TraceRule(Token, Token, String)

Rule: NumProp_Trace
{
    (Token:$first && Token.numProp:$num)
    (Token:$last && Token.numProp==$num)*
    (Token:$last && Token.pos==PCT)?
}
=>
```

```
{
    Proposition($first.startOffset, $end.endOffset, $num, #RULENAME);

    TraceRule($first, $last, #RULENAME) ;
}
```

### Réalisation technique

Le compilateur Myéline a été réalisé en *Java* via l'analyseur lexical *JFlex*<sup>36</sup>, et l'analyseur syntaxique *Cup*<sup>37</sup>. Le compilateur génère un module *C++* pour la plate-forme Synapse Etiquette, directement intégrable dans un projet *Microsoft Visual Studio*. Pour la plate-forme *UIMA*, le compilateur génère un annotateur (*Analysis Engine*) *Java*, ainsi que les types d'annotations associés.

### Conclusion

Le tableau suivant présente les avantages et inconvénients de Myéline.

Adéquation	Excellente : langage spécifique permettant de déclarer des règles à un bon niveau d'abstraction.
Simplicité	Bonne : Je pense que le langage Myéline est relativement intuitif, mais je ne suis pas le mieux placé pour l'évaluer. Les décrochages en langage source sont empaquetés dans une déclaration Myéline : le concepteur de la règle n'est pas tenu de connaître le langage source et l'API interne de la plate-forme cible.
Expressivité	Bonne
Portabilité	Excellente (théoriquement) – Correcte (en pratique) : Le code est portable, moyennant l'existence d'un compilateur pour la plate-forme cible.

Tableau 3.14 : Avantages et inconvénients de Myéline

Myéline a été spécifiquement conçu pour corriger certains défauts de Jape, et répondre ainsi aux besoins de Synapse Développement dans le cadre de nos travaux de recherche. Les spécifications du langage et le compilateur du langage ont été réalisés dans le cadre de cette thèse.

### 3.3.3. Choix final d'architecture

En regard des contraintes techniques et des besoins d'industrialisation de la chaîne de traitement, il a été fait le choix d'utiliser la plate-forme interne de Synapse Développement, et de développer le langage Myéline pour faciliter la maintenance de la chaîne de traitement.

Le Tableau 3.15 résume les caractéristiques de ce choix.

<sup>36</sup> <http://jflex.de/>

<sup>37</sup> <http://www2.cs.tum.edu/projects/cup/>

Prototypage	Outils disponibles	Mitigée : Pour : accès direct aux modules de Synapse Développement Contre : des passerelles sont à développées pour la plupart des outils externes.
	Abstraction des objets	Excellente : les structures de données internes permettent la gestion de texte.
Industrialisation	Portabilité	Bonne - Correcte : Les modules Myéline sont potentiellement portables Les modules et fonctions en C++ sont spécifiques, disponible Linux/Windows sous réserve de non-utilisation de fonctions systèmes dans les parties spécifiques à l'extraction d'opinion.
	Légèreté	Bonne : ne nécessite que l'installation de la chaîne de traitement.
	Possibilités d'optimisation	Excellente : accès direct aux structures internes, permettant une optimisation poussée
	Scalabilité	Bonne : maîtrise de la taille maximale des fichiers (segmentation en sous-fichier envisageable). Exécution parallèle envisageable.
	Licence	OK dans le contexte de la thèse CIFRE.
Maintenance	Facilité	Bonne : Myéline assouplit considérablement la maintenance corrective, qui peut désormais être effectuée sans connaissances poussées de l'API interne.
	Adaptabilité	Correcte : les modules sont compilés vis à vis du reste de la chaîne, mais sous forme de bibliothèques indépendantes.

Tableau 3.15 : Récapitulatif des avantages et inconvénients de l'environnement d'exécution final

Dans le cadre des expérimentations préliminaires, les algorithmes ont néanmoins parfois été préalablement testés sur d'autres plates-formes, pour des raisons pratiques. En effet, ces prototypes n'ont pas à répondre aux problématiques d'industrialisation évoquées dans cette section. De plus, le développement de Myéline s'est réalisé tout au long de la thèse, et la plate-forme finale n'était pas nécessairement disponible au moment des expérimentations.

### 3.4. Conclusion

Dans cette section, nous avons pu voir l'architecture générale de la chaîne d'extraction de l'opinion. Cette chaîne comprend :

- un module de segmentation discursive, adapté de (Afantenos et al. 2010) pour l'analyseur syntaxique de Synapse Développement (Laurent, Nègre, and Séguéla 2009). L'évaluation de ce module montre des performances similaires à celles de (Afantenos et al. 2010) sur le corpus segmenté discursivement du projet ANNODIS

(Péry-Woodley et al. 2009), et légèrement supérieures sur le corpus de commentaires web issus du projet CASOAR (cf. chapitre 2).

- Un module de détection de la subjectivité des segments discursifs, détaillé au chapitre 5.
- Un module de repérage local des expressions à l'intérieur de chaque segment basé sur un lexique d'opinion présenté au chapitre 4.
- Un module de calcul local permettant de remonter les attributs de l'opinion du niveau de l'expression au niveau du segment en prenant en compte les opérateurs s'appliquant sur les opinions. Ce module peut-être décomposé en deux sous-tâches, l'une de détermination de la portée de chaque opérateur, l'autre de calcul des attributs résultants de l'application du ou des opérateurs. Ces deux composantes seront décrites au cours du chapitre 6.
- Un module de calcul de l'opinion globale au niveau du document, permettant la fusion des informations d'opinion en une seule valeur globale. Ce module sera abordé au chapitre 7.

Dans le cadre du contrat de thèse CIFRE, l'architecture utilisée devait répondre à plusieurs contraintes, notamment en matière d'industrialisation de la chaîne, et de confort de maintenance. Pour cela, après étude des différentes options disponibles, nous nous sommes décidés pour une architecture intégrée à la chaîne d'analyse syntaxique de Synapse Développement. Afin de permettre une maintenance corrective et adaptative aisée, nous avons également décidé de nous lancer dans la conception de notre propre langage de règles, Myéline. Ce langage, spécialisé dans le traitement du langage naturel, possède deux principaux avantages sur le langage Jape de la plateforme GATE : il est d'une part légèrement plus expressif, en permettant des références croisées via des variables dans la partie condition de la règle, et d'autre part potentiellement plus portable, en introduisant une couche intermédiaire de déclaration pour les usages de code externe spécifique à la plateforme.







## 4. Génération automatique d'un lexique d'opinion

---

### 4.1. Introduction

Devant le travail nécessaire pour constituer manuellement une ressource linguistique, l'intérêt des méthodes de génération automatique de lexique est transparent : pouvoir créer une ressource, à coût extrêmement réduit et en un temps négligeable. Dans le contexte de cette thèse *CIFRE*, la société Synapse Développement met à disposition de nombreuses ressources linguistiques propriétaires pour le Français, de très bonne qualité, et liées à une désambiguïsation syntaxique et sémantique fine. Les méthodes d'extension automatique de lexique permettraient donc de profiter du travail déjà réalisé sur ces ressources.

Néanmoins, le lexique d'opinion peut être considéré comme la clé de voûte de toute la chaîne d'extraction de l'opinion : intervenant dès les premières briques de la chaîne, toute erreur issue de celui-ci sera répercutée dans les étapes ultérieures. Il convient donc de déterminer s'il est possible de constituer automatiquement un lexique de haute qualité, notamment en matière de précision – pour éviter les erreurs visibles qui dégraderaient la qualité ressentie de la chaîne.

Pour répondre à cette interrogation, nous allons étudier et évaluer deux méthodes de génération automatique de lexiques de polarité (i.e. de lexique de termes porteurs d'opinion avec leur polarité associée) à partir de ressources dictionnairiques génériques :

- l'une basée sur une taxinomie classant tous les mots du Français (section 4.3) – travaux publiés dans (Chardon 2010)
- l'autre basée sur le dictionnaire de synonymes de Synapse Développement (section 4.4).

Nous étudierons ensuite la possibilité d'étendre ces méthodes pour la constitution d'un lexique distinguant finement les intensités et/ou les catégories sémantiques d'opinion définies au chapitre 2. Ceci nous permettrait d'obtenir un lexique de richesse sémantique suffisante pour notre chaîne de traitement. La section 4.5 décrit les travaux effectués dans cette optique, ainsi que les résultats obtenus.

Enfin, après analyse des résultats et afin de minimiser les erreurs dans le lexique, nous avons opté pour un lexique manuel, constitué dans le cadre du projet CASOAR. La section 4.6 présente ce lexique.

### 4.2. Contexte scientifique

Afin de générer un lexique, le principe général est de s'appuyer sur une ébauche construite manuellement, appelé lexique germe ou termes germes. Suivant les méthodes, cette ébauche peut être constituée d'un ensemble très restreint de termes – par exemple, (Kamps et al. 2004)

sont partis des deux seuls adjectifs "good" et "bad" – ou d'un ensemble plus conséquent. Pour étendre ce lexique germe, deux principales catégories de méthodes sont utilisées dans la littérature.

Les premières sont les méthodes basées sur corpus, qui n'utilisent que les informations présentes dans le ou les corpus utilisés, annotés ou non. (Hatzivassiloglou and McKeown 1997) ont utilisé les conjonctions pour étendre leur lexique, partant de la supposition que deux adjectifs en conjonction (par exemple, reliés par un « *et* ») ont une polarité similaire, et que deux mots en contraste (par exemple reliés par un « *mais* ») ont une polarité opposée. Turney et Littman ont proposé deux scores calculant la polarité d'un terme étant donné son contexte : le *SO-PMI* (*Semantic Oriented – Pointwise Mutual Information*, (Turney and Littman 2002)) basé sur les cooccurrences d'un mot avec les termes du lexique germe polarisé, et le *LSA* (*Latent Semantic Analysis*, (Turney and Littman 2003)), calculant la proximité des contextes d'utilisation de deux mots. (Kanayama and Nasukawa 2006) ont étendu cette méthode au Japonais, obtenant une bonne précision. (Harb et al. 2008) ont utilisé, sur corpus spécifique extrait dynamiquement du web, les collocations entre adjectifs et mots germes pour constituer une liste d'adjectifs d'opinion polarisés. Ces méthodes requièrent généralement un corpus assez conséquent. Il est d'ailleurs en général possible de prendre la totalité du web indexé par les moteurs de recherche comme corpus pour appliquer les méthodes présentées, notamment au moyen d'opérateurs fournis par les moteurs de recherche : par exemple, l'opérateur NEAR d'Altavista (opérateur aujourd'hui abandonné) (Turney and Littman 2003; Taboada, Anthony, and Voll 2006), l'opérateur AND de Google (Taboada, Anthony, and Voll 2006), ou encore l'opérateur d'exclusion de termes de Google (Harb et al. 2008). Néanmoins, (Gamon and Aue 2005) ont montré que l'utilisation d'un corpus spécifique au domaine donnait, en général, de meilleurs résultats. Pour le français, nous pouvons citer les travaux de (Vernier and Monceaux 2010), qui proposent une méthode d'extension de lexique de polarité par tests sémantiques en deux étapes : détection de termes potentiellement subjectifs inconnus, en se basant sur l'hypothèse qu'un adjectif non subjectif sera rarement intensifié par un adverbe, puis attribution d'une polarité en se basant sur les contrastes avec les termes du lexiques germes, à la manière de (Hatzivassiloglou and McKeown 1997).

Les secondes méthodes, dites basées sur dictionnaire, consistent à augmenter un lexique germe via des liens sémantiques et/ou syntaxiques, présents dans des ressources linguistiques externes. (Kobayashi, Inui, and Inui 2001) ont proposé une méthode permettant d'étendre un lexique germe en appliquant un jeu de 52 règles manuellement constituées sur les *gloses* d'un dictionnaire – cette méthode, de par les règles utilisées est spécifique au langage cible, ici le Japonais. (Kamps et al. 2004) ont utilisé la relation de synonymie de *WordNet* (Miller et al. 1990; Fellbaum 1998) pour calculer une distance entre adjectifs. La distance relative avec les mots germes "good" et "bad" a permis la classification d'adjectifs entre positif et négatif. (Kim and Hovy 2005) ont quant à eux directement étendu leurs lexiques à partir de synonymes et antonymes de ces mots germes. (Esuli and Sebastiani 2006b) ont amélioré ces méthodes en effectuant au préalable une désambiguïsation syntaxique. Plusieurs travaux ont exploités les phrases de description des *synsets* de *WordNet* (*gloses*) : (Andreevskaja and Bergler 2006) ont classé ces derniers entre positif, négatif, et neutre (objectif), en testant la présence de mot germes dans la description du *synset* ; (Esuli and Sebastiani 2006a) ont quant à eux exploités la proximité entre différentes

*gloses*. (Chesley et al. 2006) ont classé les adjectifs suivant le nombre d'occurrences de mots germes dans leur page de description dans le dictionnaire collaboratif *Wiktionary*<sup>38</sup>, section antonymes mise à part. Afin de constituer un lexique privilégiant la précision au détriment du rappel, (Kaji and Kitsuregawa 2007) se sont appuyés sur des patrons d'extraction fins, basé sur les particules spécifiques au Japonais. Enfin, (Ding, Liu, and Yu 2008) ont proposé une extension de leurs lexiques se basant sur les synonymes et antonymes des termes, mais avec une limitation de contexte : les synonymes et antonymes n'appartenant pas au lexique source ne sont étiquetés que lorsqu'ils apparaissent dans un contexte similaire à celui du terme originel.

Quelques travaux ont également proposé une approche mixte, basée à la fois sur dictionnaire et sur corpus. Parmi eux, (Takamura, Inui, and Okumura 2005) et (Takamura, Inui, and Okumura 2007) proposent une adaptation de modèles issus de la physique – résolution du spin d'un électron appartenant à un ensemble en interaction électromagnétique – afin d'étendre un lexique germe restreint. Aux interactions physiques entre électrons sont substitués les liens entre termes dans diverses ressources lexicales, augmentés de liens de cooccurrence.

### **4.3. Génération d'un lexique de polarité via une taxinomie générale du Français**

Pour le français, peu d'études basées sur dictionnaire ont été menées à notre connaissance. Disposant, à Synapse Développement, d'une ressource linguistique fine, nous avons souhaité tester ce type d'approche.

Pour cette étude, nous avons choisi de nous baser exclusivement sur les adjectifs. En effet, ceux-ci constituent la grande majorité des expressions d'opinion relevées dans le corpus d'évaluation (cf. section 4.3.3). Ce choix a également été adopté dans la littérature pour d'autres études : citons notamment (Vernier and Monceaux 2010).

#### **4.3.1. Ressource utilisée**

Pour élaborer notre ressource lexicale sur les adjectifs d'opinion, nous avons utilisé une ressource lexicale générique. Celle-ci propose, pour 158 320 syntagmes (divisés en 174 810 sens), un ou plusieurs liens vers une taxinomie de concepts structurée hiérarchiquement sur quatre niveaux.

La Figure 4.1 expose la représentation sémantique des deux sens de l'adjectif « cher ».

---

<sup>38</sup> <http://www.wiktionary.org/>

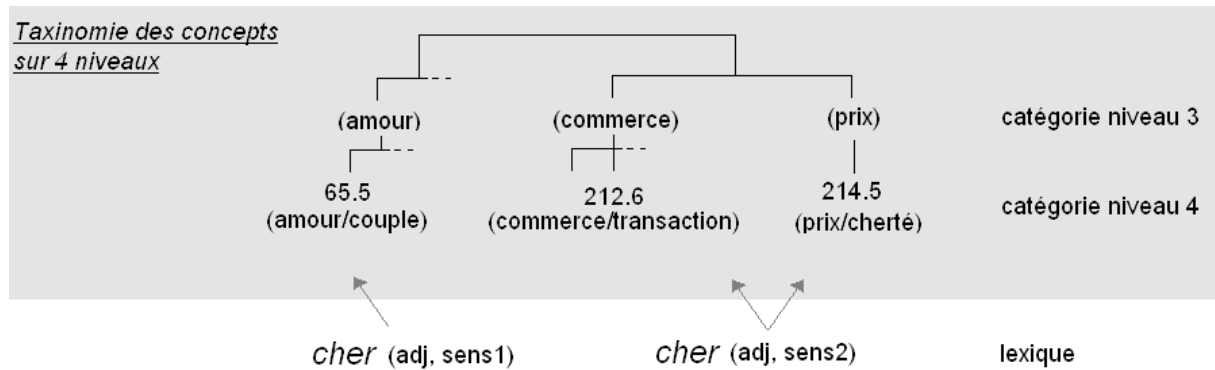


Figure 4.1 : Exemple de représentation des deux sens de l'adjectif « cher »

L'adjectif « cher » est ici modélisé comme ayant deux sens. Le sens 1 est défini par un lien vers un concept (65.5/amour/couple), tandis que le sens 2 est défini par deux liens, vers le concept (212.6/commerce/transaction) d'une part, et (214.5/prix/cherté) d'autre part. Cette taxinomie se rapproche des *synsets* de *WordNet*, à la différence qu'ici le nombre de catégories conceptuelles est borné (3387 catégories au niveau 4, 256 au niveau 3), et que les mots sont définis par sens et non par classes de regroupements.

Afin d'exploiter cette ressource, nous avons étiqueté le corpus d'évaluations (cf. présentation du corpus en section 4.3.3) avec l'analyseur syntaxique de Synapse Développement (Laurent, Nègre, and Séguéla 2009). À chaque mot est associé son lemme, sa catégorie syntaxique, et le ou les numéros de sens reconnus lors de la désambiguïsation sémantique locale. Par extension, un ensemble de catégories conceptuelles de niveau 4 est donc associé à chaque mot.

Nous avons observé que les catégories où l'on trouve des adjectifs d'opinion rassemblent des mots de sens proches et de même polarité. Elles regroupent cependant des mots d'intensités différentes, ce qui interdit tout classement des termes d'opinion par intensité en se basant sur cette ressource. L'intensité des adjectifs d'opinion n'a donc pas été prise en compte lors de l'extension du lexique présentée ci-après.

### 4.3.2. Algorithme proposé

L'algorithme repose sur la construction d'un graphe bipartite. Les sommets sont d'une part les concepts de niveau 4 de la taxinomie (sommets « concept »), et d'autre part les termes rencontrés (sommets « terme »). Un sommet « terme » est relié à tous les sommets « concept » correspondant aux concepts qui lui sont associés dans la taxinomie. Nous associons ensuite une polarité à chacun de ces sommets.

L'algorithme est ici exposé pour un lexique germe d'adjectifs d'opinion que l'on souhaite étendre à d'autres adjectifs, ce qui correspond à l'évaluation qui en sera faite en section suivante. Il peut néanmoins être aisément généralisé aux termes portant une catégorie grammaticale différente.

#### *Constitution du lexique germe*

À partir d'un petit nombre de documents annotés manuellement, on extrait une liste lemmatisée d'adjectifs germes d'opinion, ainsi que leur polarité (positif/négatif). Cette étape est critique, au

sens où un ensemble de mots germes trop restreint ou déséquilibré nuira aux résultats. Un minimum d'une vingtaine de mots germes est souhaitable pour couvrir suffisamment d'entrées concepts de la taxinomie.

### Initialisation

On initialise ensuite les sommets « concepts » du graphe, à partir des concepts de niveau 4 de la taxinomie. Pour chacun des mots du lexique germe extrait, on ajoute un sommet « terme », et on lui associe la polarité  $p$  du mot. Si à ce mot sont associés un ou plusieurs concepts, on relie ce nouveau sommet aux sommets représentant ces concepts. La Figure 4.2 montre un exemple de graphe après initialisation.

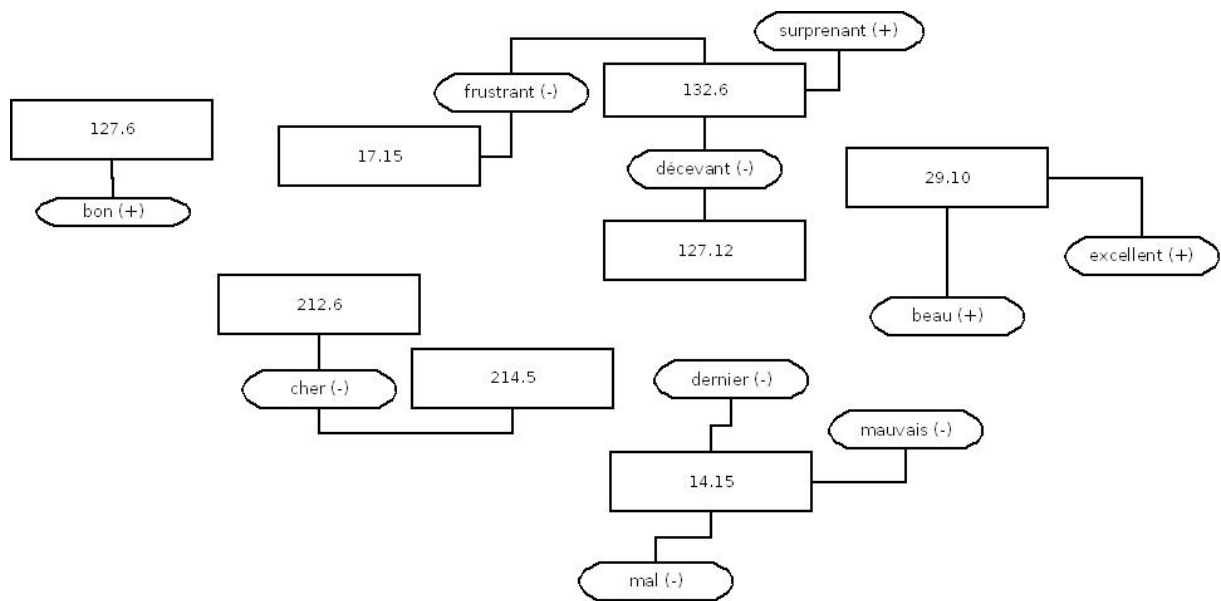


Figure 4.2 : Extrait du graphe après initialisation

### Résolution

L'objectif est de construire un arbre de décision basé sur les catégories de taxinomie des adjectifs du lexique germe. Plus formellement, on souhaite affecter à chaque nœud « concept » une polarité, et une hauteur dans l'arbre de décision (i.e. un ordre de priorité dans la prise de décision) telles que :

$$\forall \text{ terme } m, \text{ soit } T_m \text{ l'ensemble des catégories de la taxinomie qui lui sont associées}$$

$$\forall t \in T_m, \quad \{(\exists t' \in T_m, \text{ hauteur}(t') < \text{hauteur}(t)) \text{ ou } (\text{polarité}(t) = \text{polarité}(m))\}$$

$\text{hauteur}(t)$  est ici l'inverse de l'ordre de priorité dans laquelle la catégorie est prise en compte : plus celle-ci est élevée, moins la catégorie sera considérée comme discriminante pour déterminer la polarité.

Le graphe obtenu à l'étape d'initialisation se rapproche d'un graphe de CSP (*Constraint Satisfaction Problem*), où les nœuds « terme » représenteraient les contraintes à satisfaire : la polarité associée à une catégorie doit être identique à celle associée aux adjectifs de cette catégorie. On applique par conséquent un algorithme de résolution inspiré de la cohérence d'arcs (citons (Cooper et al. 2008) pour un exemple d'application de la cohérence d'arc sur un CSP).

Un nœud « concept » est dit cohérent à une étape donnée si tous ses voisins de type « terme » présente la même polarité  $p$ . À chaque étape, numérotée  $h$ , on affecte à tous les nœuds « concept » cohérents, la hauteur  $h$  et la polarité  $p$ . On élimine ensuite du graphe tous les nœuds « terme » ayant un voisin « concept » partageant leur polarité : en effet, ces nœuds sont dits « résolus » : on a trouvé un concept associé permettant de déterminer leur polarité. Dans l'éventualité où, à une étape donnée, aucun nœud « concept » ne serait affecté, on associe à un nœud « concept » quelconque une polarité arbitraire et la hauteur  $h$ . Idéalement, cette affectation minimise le nombre de contraintes violées, i.e. le nombre de nœuds « terme » reliés à une catégorie dont la polarité serait différente de celle qu'ils véhiculent.

La phase de résolution s'achève lorsque tous les nœuds « terme » ont été retirés du graphe, ou qu'une polarité a été affectée à tous les nœuds « concept » du graphe.

### *Construction de l'arbre de décision / classifieur du lexique*

On obtient ainsi une liste de nœuds correspondant aux catégories de la taxinomie, chacun associé à une hauteur  $h$  et une polarité  $p$ . En associant à chaque nœud « concept »  $N_c(h_{N_c}, p_{N_c})$  une branche de hauteur  $h_{N_c}$  menant sur une décision d'affectation de la valeur de polarité  $p_{N_c}$ , on obtient un arbre de décision permettant de classer les adjectifs rencontrés dans un texte suivant les catégories de taxinomie lui étant associées. Cet arbre de décision permet de générer, par extension, un lexique visualisable sous forme de liste : il suffit pour cela de projeter les décisions sur tous les termes répertoriés dans la catégorie. En pratique, l'arbre seul suffit néanmoins pour déterminer la polarité des termes considérés.

L'exemple suivant présente un sous-graphe de la figure 4.2, possédant une ambiguïté sur un concept : la catégorie 132.6 est ici reliée à deux termes négatifs et à un terme positif.

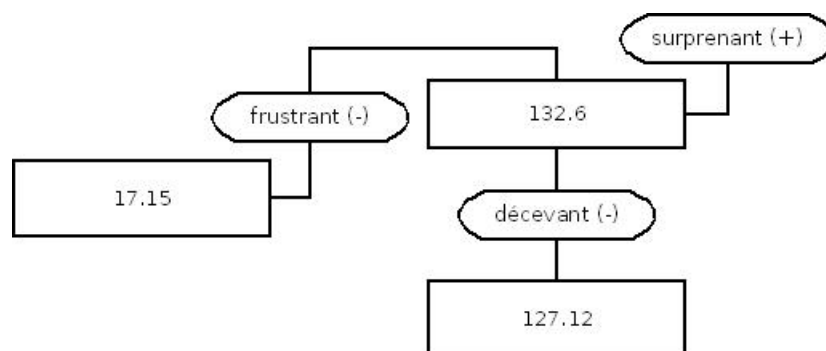
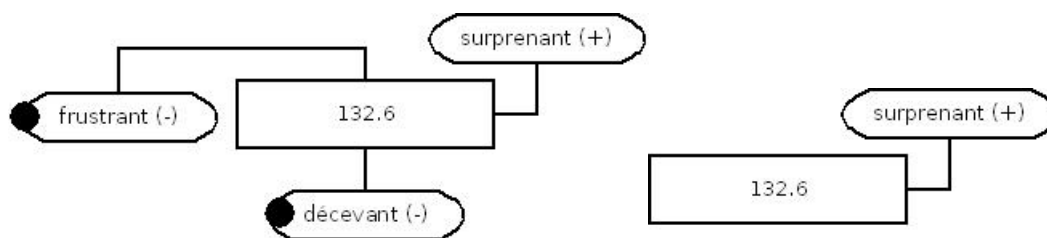


Figure 4.3 : Sous-graphe / exemple de concept non affectable directement

On ne peut résoudre ce conflit et associer une polarité à la catégorie lors de la première itération de l'algorithme. Par contre, à l'issue de celle-ci les catégories 17.15 et 127.12 se voient affecter une polarité négative : les adjectifs "frustrant" et "décevant", tous deux négatifs ont désormais un voisin "concept" partageant leur polarité, ils sont donc éliminés du graphe (figure 4.4).



Figures 4.4 (gauche) & 4.5(droite) : Détail du graphe à l'issue de la première itération (à gauche), et au début de la seconde itération (à droite)

La catégorie 132.6 se retrouve donc reliée uniquement à un terme positif (figure 4.5) : elle se voit donc affecter une polarité positive, mais à une hauteur 2. Dans l'arbre de décision résultant, rencontrer un terme de catégorie 132.6 conduit donc à une valeur positive, mais uniquement si le terme n'est pas également associé à la catégorie 17.15 ou 127.12, négatives et plus prioritaires.

La Figure 4.6 présente, sous une forme compacte, l'arbre complet correspondant au graphe de la Figure 4.2. Les flèches s'interprètent comme "*Si le terme rencontré est de catégorie <Label de la flèche> Alors affecter la polarité <Valeur de la feuille>*". Le "0" correspond à une absence d'affectation de polarité par l'algorithme : le terme est considéré comme neutre ou ne portant pas d'opinion.

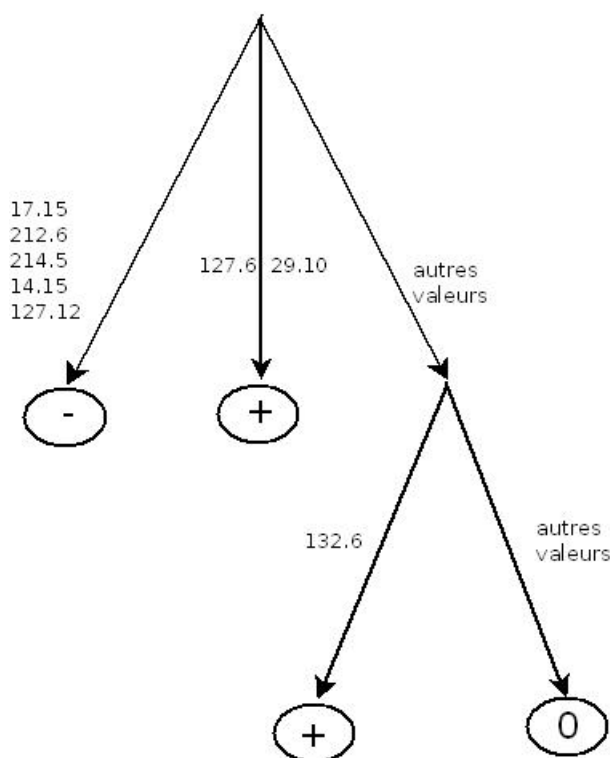


Figure 4.6 : Arbre de décision issu de la résolution du graphe Figure 4.2

#### **Optionnel : extension aux catégories proches**

Afin de permettre une extension plus importante du lexique, par exemple dans le cas où peu de mots germes sont utilisés (par exemple si peu de textes annotés sont disponibles), nous proposons une phase optionnelle d'extension aux catégories proches de la taxinomie. Deux catégories sont jugées « proches » si elles comportent au moins un terme désambiguïsé (i.e. un terme dans un

sens donné) en commun. Par exemple (cf. figure 4.1), la catégorie prix/cherté est proche de commerce/transaction, dans la mesure où elle partage le terme cher dans son sens 2. À l'aide de cette relation de proximité, on peut répercuter la polarité des catégories positives ou négatives aux catégories proches, dans la mesure où cela n'engendre pas de conflit.

Afin de pallier aux risques de faux positifs, nous avons extrait, en même temps que les mots germes, une liste d'adjectifs n'étant pas des adjectifs d'opinion. Ceci a permis de constituer une liste d'arrêt de catégories à travers lesquelles le lexique n'est pas étendu.

### 4.3.3. Évaluations

Les évaluations se sont déroulées en trois temps. Tout d'abord, nous étudierons la couverture du lexique généré ; ensuite, nous évaluerons la généricité de celui-ci, et enfin, son utilisabilité pour une tâche extrinsèque.

Le corpus utilisé pour l'évaluation est constitué de commentaires web sur des restaurants ou bars déposés en ligne sur le site Qype<sup>39</sup>. Ces commentaires sont libres, de temps à autre hors sujet, et contiennent parfois de multiples fautes d'orthographe. Afin de ne pas influencer l'annotateur, le corpus présenté ne contenait que le texte du commentaire, et aucune des informations annexes disponibles en ligne.

L'annotation a été effectuée par deux annotateurs familiers du domaine, sous l'environnement Glozz<sup>40</sup>. Cette annotation est antérieure au développement du manuel pour le projet CASOAR, évoqué au chapitre 2. La structure d'opinion annotée est néanmoins similaire sur les points concernés par cette évaluation. Le corpus est constitué de 40 documents plutôt positifs a priori, et de 20 documents plutôt négatifs a priori. Chaque commentaire est constitué en moyenne d'environ 3 phrases / 70 mots. Un manuel d'annotation a été fourni aux annotateurs afin de les guider dans la tâche : il était demandé aux annotateurs d'annoter les expressions d'opinion avec leur polarité, intensité et catégorie sémantique, ainsi que les sujets, opérateurs et émetteurs associés. Après une première annotation de 10 documents, un retour a été effectué sur le manuel afin de préciser et corriger certains points ambigus. L'annotation des éléments discursifs a été laissée en suspens pour le moment.

#### 4.3.3.1. Couverture du lexique

Le corpus utilisé a été séparé en deux groupes : un corpus d'entraînement constitué de 5 commentaires positifs et 5 commentaires négatifs, choisis préalablement pour leurs polarités tranchées, et un corpus de test, constitué des 50 commentaires restants. La liste de mots germes était constituée de 15 termes positifs, et de 11 négatifs, extraits du corpus d'entraînement. En comparant les adjectifs reconnus avec ceux annotés manuellement, nous obtenons les résultats suivants :

---

<sup>39</sup> [www.qype.fr](http://www.qype.fr)

<sup>40</sup> [www.glozz.org](http://www.glozz.org)



	Rappel	Précision	F-mesure
Sans renforcement	39%	74%	51%
Avec renforcement	66%	53%	59%

Tableau 4.1 : Couverture du lexique

Nous pouvons tout d'abord remarquer que les résultats sans l'étape optionnelle de renforcement du lexique montrent une couverture relativement faible. Ceci peut s'expliquer par le fait que les catégories étant assez fines, certains traits sémantiques approchants sont divisés en plusieurs catégories ; par exemple, les catégories « beauté » et « beauté humaine » sont distinctes. Par conséquent, des mots pourtant proches peuvent ne pas être reconnus comme tels. La phase d'extension par proximité (renforcement) permet de corriger ce défaut : les deux catégories précédentes possèdent au moins un terme en commun (« *joli* »), et le caractère positif des adjectifs de l'une se propage donc à l'autre.

Ensuite, nous pouvons observer que la précision est assez bonne pour le lexique sans renforcement. Les faux positifs relevés avant extension par proximité sont principalement dus à deux phénomènes. Tout d'abord, certains adjectifs peuvent s'avérer ambigus hors contexte. Par exemple, « *digne* » utilisé seul est clairement positif (« *Les tapas sont dignes.* »), mais utilisé dans l'expression « *digne de X* », sa polarité dépend de celle du X (« *Le poulet est digne d'un mauvais fast-food.* »). D'autre part, certains adjectifs peuvent être employés, suivant le contexte, soit comme opérateur d'intensité, soit comme expression d'opinion. Par exemple, dans les phrases :

- (1) Un super restaurant.
- (2) Un restaurant super bon.

La phrase (1) comporte une utilisation de « *super* » en temps qu'expression d'opinion, tandis que dans la phrase (2), « *super* » est utilisé comme opérateur intensifiant l'opinion porté par le terme « *bon* ». Même si dans ce dernier cas *super* et *bon* partagent la même polarité, l'annotation manuelle n'inclut pas « *super* » dans l'expression d'opinion, ce qui implique que celui-ci est détecté comme faux positif. Pour le lexique renforcé, la précision diminue, mais ceci est compensé par un bien meilleur rappel ; globalement, on observe une augmentation importante de la f-mesure.

Enfin, nous avons pu constater qu'itérer plus d'une fois la phase d'extension de proximité ne donne par contre pas de bons résultats : le rappel augmente très peu, tandis que la précision chute drastiquement.

#### 4.3.3.2. Généricité du lexique

Le lexique ainsi obtenu a été généré à partir de mots germes issus d'un corpus spécialisé – critiques de restaurants et bars – et étendu via une ressource générique, indépendante de tout domaine. On peut donc se demander dans quelle mesure ce lexique est pertinent sur de nouveaux domaines spécialisés. Pour cela, dix commentaires issus d'un domaine différent – commentaires

d'internautes sur films et séries, déposés sur le site Allociné<sup>41</sup> -- ont été annotés. Ce corpus est une sous-partie du corpus de commentaires web présenté au chapitre 2. Ces documents, de taille similaire à ceux du corpus Qype, ont été utilisés afin de générer un second jeu de mots germes. À partir de ces nouveaux termes, nous avons généré un second lexique, et comparé celui-ci au lexique obtenu précédemment.

Les deux lexiques se sont révélés relativement différents. En effet, le lexique issu de films comportait beaucoup plus de termes d'émotions fortes, notamment positives, comme « *poignant* », tandis que celui issu de restaurants comportait des termes du registre culinaire (« *délicieux* »), ainsi que des adjectifs négatifs associés à un prix (« *cher* », « *élevé* »). Statistiquement, ceci s'est traduit par une chute de moitié du rappel lors de l'annotation du corpus de test de commentaires de restaurants avec le lexique issu des critiques de films, la plupart des termes du lexique appartenant à des catégories quasi absentes du corpus traité. Cela montre que le lexique généré est plutôt spécifique au corpus : les catégories, bien que construites de manière génériques, sont identifiées comme positives ou négatives avant tout en fonction des mots germes identifiés.

#### 4.3.3.3. Utilisabilité du lexique

Les commentaires du corpus de restaurants sont, comme souvent sur les sites de critiques, associés à des étoiles (1 à 5). On peut par conséquent inférer la polarité que souhaitait a priori exprimer le rédacteur à travers sa critique : positive (4 ou 5 étoiles), négative (1 ou 2 étoiles), ou mitigée (3 étoiles). Nous avons tenté de retrouver la polarité pour les documents du corpus de tests, en comptant simplement les adjectifs d'opinion reconnus : une majorité d'adjectifs positifs (resp. négatifs) signifiant un commentaire positif (resp. négatif), et une absence de majorité un commentaire mitigé. Nous avons testé la difficulté de la tâche relativement à un lexique en calculant les scores obtenus par le lexique germe seul, ainsi que le score maximal atteignable avec le lexique « gold », issus de l'annotation manuelle complète (i.e. corpus d'entraînement + corpus de test).

	Proportion de polarité reconnue
Mots germes	45%
Liste étendue (sans extension)	65%
Liste gold	83%

Tableau 4.2 : Résultats sur tâche extrinsèque

Les résultats montrent une amélioration de 20 % de bonnes classifications avec le lexique généré. On peut également remarquer que le lexique « gold » n'obtient pas 100 %, principalement en raison de deux facteurs. Tout d'abord, l'hypothèse de conformité entre la polarité du commentaire et la polarité déduite du nombre d'étoiles n'est pas toujours vérifiée : certains internautes utilisent le commentaire uniquement afin de compléter une opinion donnée par des étoiles. Ainsi, un internaute pourra mettre 4 étoiles avec pour seul commentaire « Un peu bruyant. » : le nombre d'étoiles lui permet implicitement de signaler au lecteur que les autres

<sup>41</sup> [www.allocine.fr](http://www.allocine.fr)

aspects du restaurant étaient probablement tout à fait corrects. D'autre part, l'algorithme de classification du commentaire suivant le nombre d'adjectifs de telle ou telle polarité est relativement sommaire : il ne prend en compte ni les négations, ni la présence éventuelle d'adjectifs d'opinion portant sur des éléments hors-sujet par rapport au restaurant jugé (par exemple, les autres restaurants de la rue), ni la présence d'opinions véhiculées par des expressions d'autre nature grammaticale qu'adjectif (noms, adverbes par exemple).

#### **4.3.4. Conclusion**

Les résultats obtenus par cette méthode se sont avérés mitigés : l'extension de lexique permet effectivement d'obtenir un lexique correct de termes d'opinions, mais la précision et le rappel ne sont pas suffisants pour une utilisation dans une chaîne de traitement complète. En effet, le lexique présente une couverture assez faible ; celle-ci est améliorable dans une certaine mesure via une phase itérative, mais au détriment de la précision. Dans les deux cas, les faux positifs et/ou les silences induisent des erreurs risquant de se propager tout le long de la chaîne de traitement.

### **4.4. Génération d'un lexique de polarité via dictionnaire de synonymes**

Afin de générer une ressource dont la qualité serait plus adaptée à la chaîne de traitement, nous nous tournons vers une autre méthode de génération de lexique.

Cette méthode est dite mixte dans le sens où elle se base à la fois sur corpus, et sur une ressource linguistique pré-existante – le dictionnaire de synonymes/antonymes de Synapse Développement (Laurent and Séguéla 2003).

L'algorithme utilisé suit la proposition de (Takamura, Inui, and Okumura 2005) d'adapter la loi de calcul du spin d'un ensemble d'électrons en interaction électromagnétique. Nous présentons plus en détail cet algorithme en section 4.4.2. L'approche ici décrite introduit deux aspects originaux :

- Un score de proximité est associé à la relation entre mots, ce qui permet une amélioration de la fiabilité de la propagation. Ce score est calculé via des méthodes basées sur corpus.
- L'ajout de la désambiguïsation sémantique permet la génération d'un lexique plus fin.

Comme pour les travaux présentés en section précédente, nous nous basons exclusivement sur les adjectifs pour cette étude.

#### **4.4.1. Ressource utilisée**

Le dictionnaire de synonymes de Synapse Développement a été construit en deux étapes, comme décrit par (Laurent and Séguéla 2003).

La première étape a été la construction des liens de synonymie et d'antonymie entre termes désambiguïsés sémantiquement. Ces liens sont basés sur diverses observations sur le langage. Dans un second temps, ces relations ont été pondérées via un score de collocation. La proximité

entre les entrées a été calculée à partir des résultats donnés par le moteur de recherche Google pour une recherche de ces termes. Ces résultats bruts sont ensuite convertis en un entier allant de 1 à 255, 255 s'interprétant comme un lien sémantique très fort entre les termes, et 1 comme un lien très faible. Le tableau suivant présente un extrait de la liste des synonymes et antonymes du mot "mauvais" dans son sens 1 : "qui n'est pas bon".

	Type de lien	Score de proximité
méchant	synonymie	255
détestable	synonymie	151
abominable	synonymie	76
bon	antonymie	255
excellent	antonymie	60

Tableau 4.3 : Extrait de la liste de synonymes et antonymes de mauvais (sens 1)

Une autre particularité de cette ressource vient de son lien avec les autres ressources de Synapse Développement, et en particulier l'analyseur syntaxique. Pour chaque terme du dictionnaire de synonymes/antonymes, des relations sont données par sens. Ces sens peuvent également être déterminés dans un texte donné par l'analyseur syntaxique de Synapse Développement, via son module de désambiguïsation sémantique. Par conséquent, l'usage combiné de ces deux ressources permet de retrouver dynamiquement une liste de synonymes et/ou antonymes pour un terme donné dans un contexte donné.

D'un point de vue informatique, ce dictionnaire peut donc être vu comme un graphe ayant pour sommets les couples (mot, sens du mot), et pour arêtes les liens pondérés de synonymie et antonymie.

#### 4.4.2. Algorithme d'extension

L'algorithme d'extension de lexique se base sur un principe de propagation de la polarité, à partir d'un lexique germe, à travers le graphe du dictionnaire de synonymes. Il est inspiré du modèle du spin de l'électron. Ce modèle est utilisé en physique pour calculer la valeur du spin de chaque particule d'un groupe d'électrons en interaction. Le spin d'un électron peut prendre deux valeurs : +1/2 et -1/2 ; en cela, il peut être apparenté à une "polarité" de la particule. Pour des raisons énergétiques, deux particules en interaction forte auront tendance à avoir la même valeur de spin : ce lien peut être vu comme analogue au lien de synonymie entre deux termes.

Les équations physiques décrivant les interactions énergétiques entre les particules peuvent être résolues en appliquant la règle itérative suivante :

$$\bar{x}_i^{new} = \frac{\sum_{x_i} x_i * \exp(\beta * x_i * \sum_j \omega_{ij} \bar{x}_j^{old})}{\sum_{x_i} \exp(\beta * x_i * \sum_j \omega_{ij} \bar{x}_j^{old})}$$

où  $\bar{x}_i$  est la valeur moyenne du spin pour la particule  $i$ ,  $\omega_{ij}$  le poids du lien entre les particules  $i$  et  $j$ , et  $\beta$  un paramètre à estimer. (Takamura, Inui, and Okumura 2005) proposent une démonstration de cette formule.

Ce principe peut être étendu à un ensemble de mots subjectifs : deux termes fortement liés d'un point de vue synonymie auront tendance à partager la même polarité. De même, un lien d'antonymie peut être vu comme une interaction de valeur négative. Partant de ces observations, nous pouvons tester les équations de ce modèle pour calculer la polarité d'un ensemble de termes, en considérant  $\bar{x}_i$  comme étant la polarité du mot  $i$ , et  $\omega_{ij}$  le poids du lien entre les mots  $i$  et  $j$ .

(Takamura, Inui, and Okumura 2005) ont proposé une méthode permettant de calculer une valeur optimale pour le paramètre  $\beta$  en l'absence de données d'entraînement, via un paramètre additionnel, la magnétisation. La magnétisation  $m$  se calcule comme suit :  $m = \frac{\sum_i \bar{x}_i}{N}$ , avec  $N$  le nombre de particules en interaction. Une propriété remarquable de la magnétisation est que, lorsque  $m \approx 0$ , le spin des particules est distribué de manière aléatoire, et si  $m \neq 0$ , toutes les particules suffisamment reliées au réseau recevront la même valeur de spin. Nous souhaitons obtenir un réseau lexical polarisé localement dans le voisinage des termes germes, c'est-à-dire un réseau lexical dans lequel les termes proches d'un terme germe partagent la polarité de celui-ci. Pour cela, la valeur optimale pour le paramètre  $\beta$  est la valeur pour laquelle  $m$  s'éloigne subitement de zéro.

Pour estimer le poids du lien entre  $i$  et  $j$  (Takamura, Inui, and Okumura 2005) ont utilisé la formule suivante :  $\omega_{ij} = \pm \frac{1}{\sqrt{d(i)d(j)}}$ , avec le signe de  $\omega_{ij}$  dépendant de la nature du lien sémantique (positif pour un lien de synonymie, négatif pour un lien d'antonymie), et  $d(i)$  le nombre de liens partant de l'entrée  $i$ . L'idée derrière cette équation est que plus un mot est lié à de nombreux mots, moins ces liens sont sémantiquement forts. Comme décrit précédemment, nous disposons d'un score de proximité plus fiable entre deux mots. Par conséquent, nous avons adopté la formule suivante pour le poids du lien :  $\omega_{ij} = \pm \frac{prox_{ij}}{256}$ , avec  $prox_{ij}$  le score de proximité associé au lien dictionnaire. Le diviseur 256 garantit que ce poids soit strictement inférieur à 1.

### 4.4.3. Évaluations

Afin d'évaluer la méthode, nous l'avons instancié sur la création d'un lexique d'adjectifs d'opinion. Ce choix a été effectué pour deux raisons :

- Adéquation aux données annotées disponibles
- Adéquation au type de corpus utilisé : la majorité des annotations d'opinion relevées manuellement dans les commentaires web sont des adjectifs

Nous avons effectué deux évaluations sur notre corpus de test (cf. section 4.3.3), s'appuyant chacune sur un lexique germe différent :

- la première à partir d'un lexique germe minimal, composé de seulement deux mots : « bon » et « mauvais » (lexique germe  $G_1$ )
- la seconde en ajoutant les deux mots « positif » et « négatif » à  $G_1$  (lexique germe  $G_2$ ).

Pour chacun de ces lexiques germes, nous avons généré deux lexiques étendus :

- un lexique  $LB$  de baseline, utilisant l'algorithme de propagation classique tel que défini par (Takamura, Inui, and Okumura 2005)
- un lexique final  $LF$ , utilisant le score de pondération proposé en section précédente

Chacun de ces deux lexiques ont été comparés avec le lexique « gold », issu de l'extraction à partir du corpus annoté, afin d'obtenir un score d'accuracy. Le Tableau 4.4 présente ces résultats, ainsi que le paramètre  $\beta$  optimal prédit associé.

	$\beta$ prédit	Accuracy
$LB_1$ (baseline issue $G_1$ )	1.0	85.9 %
$LF_1$ (final issu $G_1$ )	0.3	89.4 %
$LB_2$ (baseline issue $G_2$ )	1.1	88.3 %
$LF_2$ (final issu $G_2$ )	0.2	90.1 %

Tableau 4.4 :  $\beta$  prédits et accuracy mesurées

Les résultats mettent en évidence qu'une amélioration significative est obtenue en appliquant l'algorithme décrit en section précédente, ce qui valide notre hypothèse de départ concernant l'amélioration des résultats au moyen des scores de proximité du dictionnaire de synonymes de Synapse Développement.

Nous pouvons également relever deux éléments :

- le paramètre  $\beta$  calculé est très différent entre la baseline et le lexique final : ceci est dû à une différence d'ordre de grandeur dans les poids calculés entre les deux méthodes.
- les résultats sur la baseline sont globalement meilleurs que ceux obtenus (Takamura, Inui, and Okumura 2005). Cela peut être dû au fait que la tâche est plus facile (i.e. les termes contenus dans le lexique « gold » extrait du corpus annoté sont assez fortement polarisés), ainsi qu'au fait que le dictionnaire utilisé est bien adapté au problème traité.

### Conclusion

Cette seconde méthode d'extension de lexique permet d'obtenir des résultats meilleurs que ceux obtenus par la baseline, et bien meilleurs que ceux obtenus par la précédente méthode.

Ces expérimentations peuvent être reproduites pour tous langages, à la condition de disposer d'un dictionnaire de synonymes, et de calculer pour chaque entrée un poids pour les relations, comme décrit par (Laurent and Séguéla 2003).

## 4.5. Généralisation de ces méthodes à d'autres traits sémantiques d'opinion

### 4.5.1. Intensité des expressions d'opinion

L'intensité d'une expression d'opinion fait partie des éléments essentiels pour une application de l'analyse de l'opinion. Par exemple, remonter la nuance entre un avis déposé sur le web

légèrement positif (« Ce restaurant est plutôt correct. ») et un avis extrêmement positif (« Le meilleur restaurant de la ville. ») est un élément crucial, à la fois pour le restaurateur qui cherche à connaître les avis sur son restaurant, et pour l'internaute cherchant où manger pour une occasion en particulier.

Inférer automatiquement l'intensité d'une expression d'opinion présente un challenge plus important que pour la simple polarité. En effet, introduire l'intensité fait que le nombre de classes possibles passe de deux/trois (suivant que l'on considère une valeur neutre) à entre six et sept (en se basant sur le schéma de représentation de l'opinion défini au chapitre 2). Cela induit bien évidemment une complexité importante, et un cumul des erreurs. Les algorithmes proposés pour la polarité ne peuvent s'appliquer de manière efficace, et ce chacun pour une raison différente.

Les deux sections suivantes présentes respectivement l'adaptation des techniques présentées précédemment à la génération d'un lexique d'intensités de l'extension par taxinomie, et l'adaptation de l'extension par calcul de spin, présentées en sections 4.3 et 4.4.

#### 4.5.1.1. Extension par taxinomie

L'algorithme d'extension par taxinomie (proposition 1) repose sur le fait que, lors de l'établissement de la liste de concepts, les termes opposés se sont, en général, vus attribués au moins une catégorie différente. Par contre, comme vu en section 4.3.1, les catégories regroupent des termes d'intensités différentes. Le trait d'intensité n'est donc a priori pas généralisable à l'ensemble des termes d'une catégorie.

Afin de vérifier cette hypothèse, nous avons adapté l'algorithme en utilisant les valeurs d'intensité (entre 1 et 3) au lieu des valeurs de polarité. L'algorithme reste identique, seule la condition de propagation est adaptée : un nœud concept se voit affecter une valeur d'intensité donnée si tous ses voisins "mots" ont la même intensité. Afin de traquer les cas d'affectations arbitraires, la valeur par défaut est arbitrairement fixée à 0. Le tableau suivant présente les résultats de ce test pour l'évaluation de la couverture du lexique avec intensités.

	Rappel	Précision	F-mesure
Intensité = 1	0.0	0.0	/
Intensité = 2	0.17	0.73	0.27
Intensité = 3	0.10	0.34	0.16
Total	0.15	0.53	0.23

Tableau 4.5 : Génération de lexique d'intensités via la méthode d'extension par taxinomie

Le rappel global est très faible, et la précision est insuffisante : le lexique ainsi généré n'est pas utilisable.

#### 4.5.1.2. Extension par calcul de spin

L'algorithme d'extension par calcul de spin (proposition 2, cf. section 4.4) repose quant à lui sur une formule physique visant à déterminer le signe du spin de la particule : la valeur absolue du spin d'un électron étant toujours égale à  $1/2$ .

Afin de déterminer si les valeurs de sorties brutes de l'algorithme peuvent être reliées à l'intensité de l'expression d'opinion, nous avons calculé le coefficient de corrélation de Pearson entre les intensités annotées et le coefficient de sortie de l'algorithme. Le score mesuré est de 0.35, ce qui indique l'existence d'une corrélation, mais trop faible pour être exploitable.

Ce résultat n'est pas surprenant : nous n'avons aucune garantie que la valeur donnée par l'algorithme puisse être interprétée au-delà de son signe. De plus, aucune information d'intensité n'est fournie en entrée de l'algorithme.

## 4.5.2. Catégories sémantiques fines

Pour chacune des classes ainsi définies, un travail conséquent doit être réalisé : en plus de l'adaptation de l'algorithme, il faut constituer un lexique germe. Afin de vérifier la faisabilité de la méthode, nous avons annoté les adjectifs utilisés pour les évaluations suivant les deux catégories sémantiques suivantes : "Sentiment\_Appreciation" et "Jugement" (pour leur définition, cf. chapitre 2). Ces deux catégories ont été choisies car elles étaient les plus présentes dans le corpus de test considérées.

### 4.5.2.1. Extension par taxinomie

Nous avons adapté et réitéré les expérimentations présentées en 3.5.1.1 afin de classer suivant ces catégories. Les résultats sont présentés dans le tableau suivant :

	Rappel	Précision	F-mesure
Jugement	0.25	0.94	0.40
Sentiment_appreciation	0.08	0.62	0.14
Total	0.15	0.53	0.38

Tableau 4.6 : Résultats sur les catégories

Les résultats montrent une précision plutôt correcte, mais un rappel très faible : le lexique ainsi généré est de bonne qualité, mais offre une couverture très restreinte. Il peut néanmoins être utilisé à des fins de suggestions d'enrichissement d'un lexique manuel.

### 4.5.2.2. Extension par spin

Dans un contexte multi-classe sans relation d'ordre entre les classes, l'algorithme de calcul de spin ne peut pas directement s'appliquer.

## 4.5.3. Conclusion

Nous avons pu voir que ces algorithmes peuvent parfois être étendus à d'autres traits que la polarité, tels que l'intensité et les catégories sémantiques mais avec des résultats faibles.

La taxinomie du Français de Synapse Développement n'est pas une ressource prévue pour différencier des intensités. Par conséquent, le lexique généré d'intensités s'appuyant sur cette ressource est de qualité médiocre, autant par la faible couverture que par le taux de faux positifs. L'application de l'algorithme à la reconnaissance des catégories sémantiques les plus présentes donne par contre des résultats encourageants d'un point de vue précision ; néanmoins, le faible rappel exclut une réelle utilisabilité d'un lexique de catégories construit de cette manière.



L'algorithme d'extension via calcul de spin présente de bons résultats sur la polarité, mais n'est pas généralisable à d'autres traits : le score de sortie n'est que faiblement corrélé à l'intensité, et l'algorithme en lui-même n'est pas prévu pour fonctionner sur plusieurs classes.

## 4.6. Lexique choisi pour la chaîne de traitement : lexique CASOAR

Nous avons pu voir dans les sections précédentes que si la génération de lexique donne des résultats encourageants pour un lexique de polarité, le niveau de qualité obtenu n'est pas suffisant lorsque l'on tente de généraliser ces méthodes à un lexique d'intensités et/ou de catégories sémantiques.

Partant de ce constat et dans le cadre du projet CASOAR, un lexique d'opinion comprenant polarités, intensités, et catégories sémantiques a été manuellement constitué par Farah Benamara et Yvette Yannick Mathieu. La désambiguïsation sémantique a été effectuée en partenariat avec Sophie Muller, linguiste à Synapse Développement, pour obtenir un découpage en sens compatible avec celui du module d'étiquetage syntaxique.

### 4.6.1.1. Contexte scientifique : lexiques d'opinions manuels existants

Afin de permettre l'extraction d'opinions dans un texte, plusieurs méthodes d'acquisition de ressources lexicales ont été proposées dans la littérature. Les toutes premières ressources ont été développées manuellement. Elles sont, pour la plupart, disponibles pour la communauté scientifique, et concernent la langue anglaise.

La plus connue est très certainement *General Inquirer* (Stone 1966), qui répertorie 1 914 termes positifs et 2293 négatifs. (Bradley and Lang 1999) ont, quant à eux, constitué *ANEW* (Affective Norms for English Words), qui recense 1034 mots, évalués non pas uniquement sur leur polarité, mais sur trois dimensions : plaisir affectif (*pleasant vs. unpleasant*), excitation (*calm vs. excited*), et prédominance (*submissive vs. dominant*). (Hatzivassiloglou and McKeown 1997) ont également diffusé le lexique constitué pour leur étude, qui comprend 1336 adjectifs (657 positifs et 679 négatifs). D'autres ressources comportent des informations plus fines sur les mots. Par exemple, *WordNetAffect* (Strapparava and Valitutti 2004), est une extension de *WordNet* (Miller et al. 1990; Fellbaum 1998) basée sur le modèle *OCC* de (Ortony, Clore, and Collins 1988). Cette extension ajoute aux *synsets* (ensemble de termes synonymes) un ou plusieurs labels sémantiques de catégories d'opinion. La liste des *synsets* catégorisés de cette manière a été étendue de manière semi-automatique via les relations de *thesaurus* présentes dans *WordNet*. De même, (Esuli and Sebastiani 2006a) ont également mis à leur disposition une extension de *WordNet*, *SentiWordNet*, construite à partir des proximités de gloses des *synsets*. Afin d'améliorer la couverture des lexiques d'opinion, (Ding, Liu, and Yu 2008) ont proposé une extension manuelle de lexique introduisant des termes neutre (i.e. ambigü hors contexte, positif ou négatif en contexte) dont la polarité est résolue dynamiquement, ainsi que des expressions idiomatiques (par exemple, "*This product costs an arm and a leg.*" – "*Ce produit coûte les yeux de la tête.*").

Quelques ressources, pour la plupart récentes, sont également disponibles pour d'autres langues. Par exemple, des versions localisées de ANEW existent pour l'allemand (BAWL-R – (Vö et al. 2009)) et l'espagnol (Redondo et al. 2007). À notre connaissance, peu de ressources sont disponibles pour le français. (Mathieu 2000; Mathieu 2006) propose une ressource lexicale des mots de sentiments, répartis en 41 classes sémantiquement homogènes<sup>42</sup>. Cette ressource a par la suite été adaptée par (Asher, Benamara, and Mathieu 2009) où une classification plus étendue des expressions des opinions suivant quatre catégories est proposée. Le lexique proposé dans le cadre du projet CASOAR et présenté dans cette section se situe dans la lignée de ces lexiques. Dans le cadre du projet Blogoscopie, (Daille et al. 2011) ont extraits de leur corpus 982 entrées lexicales, classées suivant les catégories de leur schéma d'annotation<sup>43</sup> (cf. section 1.1 pour la description des catégories Blogoscopie). Parmi ces entrées, 54 sont ambiguës, et un exemple de contexte est alors précisé. Cette ressource, bien que de qualité correcte, ne remplissait pas certains critères que nous jugeons importants : les entrées d'opinion évaluatives ne comportent notamment pas d'intensité.

Toutes ces ressources manuelles sont en général de bonne qualité, mais nécessitent un temps et des moyens conséquents pour être développés, voire même pour être adaptées en vue de l'utilisation sur un domaine particulier.

#### 4.6.1.2. Description du lexique

##### *Description des entrées*

Le lexique est divisé en entrées. Chaque entrée correspond à une expression d'opinion ou à un opérateur, composé d'un ou plusieurs termes.

Pour chaque entrée sont stockées des informations d'identification : lemme du terme (respectivement de chacun des termes de l'expression), nature grammaticale du terme (respectivement de chacun des termes de l'expression).

Chaque entrée peut également se décomposer en un ou plusieurs sens. Pour chaque sens sont stockées les informations suivantes :

- Numéro d'identification : permettant d'identifier le sens après passage de l'analyseur syntaxique de Synapse Développement
- Catégorisation sémantique fine (cf. chapitre 1)
- Informations relatives à la valeur d'opinion : polarité, intensité
- Dans le cas d'un opérateur, information de type de l'opérateur (modalité, négation), et éventuellement valeurs associées (degré de certitude, degré d'engagement de l'auteur – cf. chapitre 6 pour plus de détails sur ces valeurs)

Le premier exemple ci-dessous présente l'entrée lexicale « agaçant » du lexique CASOAR. Cette entrée ne contient qu'un seul sens (négatif faible, de catégorie « Sentiment-Appréciation »

---

<sup>42</sup> Il est à noter qu'un sentiment (comme la jalousie) n'est pas nécessairement une opinion.

<sup>43</sup> Lexique disponible à l'adresse : <http://www.lina.univ-nantes.fr/?Lexique-de-l-evaluation.html>

Irritation\_colère »). Les attributs *amb\_pol*, *amb\_sens*, *amb\_strength* indiquent respectivement si l'entrée est ambiguë sur sa polarité, sur la catégorie sémantique, ou sur l'intensité : dans ce cas, plusieurs balises *sens* sont associées à une balise *lexical\_entry*.

```
<lexical_entry amb_pol="false" amb_sens="false" amb_strength="false" id="ADJ_agaçant ">
  <lemme> agaçant </lemme>
  <pos> ADJ </pos>
  <sens type="sentiment_appreciation" categorie="irritation_colère" strength="1" polarity="neg">
    <synapseSens id="monosemic"/>
    <exemple> ce professeur m'agace</exemple>
  </sens>
</lexical_entry>
```

Figure 4.7 : Entrée XML pour le terme *agaçant*

Le second exemple ci-dessous présente l'entrée lexicale « noble ». Cette entrée comporte deux sens, tous deux positif faible et de catégorie « Jugement > Évaluation » : le sens « précieux », identifié par le code \$a, et le sens « élevé, beau », de code \$b.

```
<lex_entry amb_pol="false" amb_sense="false" amb_strength="false" id="ADJ_noble">
  <lemma>noble</lemma>
  <pos>ADJ</pos>
  <sense type="jugement" category="évaluation" strength="1" polarity="pos"/>
    <synapseSens id="$a" description="précieux"/>
    <synapseSens id="$b" description="élevé, beau"/>
  </sense>
</lex_entry>
```

Figure 4.8 : Entrée XML pour le terme *noble*

Un des gros avantages de ce lexique réside dans les informations de désambiguïsation sémantique présentes pour chaque entrée. Ces informations permettent, après analyse de la structure syntaxique par l'analyseur syntaxique de Synapse Développement, de retrouver en contexte si le terme rencontré est bien employé dans un sens subjectif.

### Statistiques

Le lexique CASOAR comporte 1845 entrées. Le tableau suivant présente la répartition des entrées du lexique suivant la catégorie sémantique.

Catégories	Nombre d'entrées (catégorie)	Classes	Nombre d'entrées (classe)
Reportage	62	Informer	6
		Soutenir	20
		Dire	7
		Remarquer	6
		Penser	11

		Supposer	12
Jugement	1387	Blâmer	48
		Louer	18
		Évaluation	1321
Sentiment	373	Irritation – colère	23
		Apaisement	17
		Étonnement	42
		Intérêt – passion	98
		Haine	5
		Déception	37
		Peur	31
		Offense	22
		Tristesse – Souffrance	15
		Ennui	26
		Divertissement – Joie	43
		Émotion	14
Conseil	23	Recommander	13
		Suggérer	5
		Espérer	5

Tableau 4.7 : Répartition des entrées lexicales selon la catégorie sémantique

En plus de ces entrées d'opinion, le lexique CASOAR comporte 79 émoticônes, et 176 opérateurs non catégorisés ci-dessus. Le tableau 4.8 suivant présente la répartition des entrées du lexique suivant leur catégorie syntaxique.

	Nombre d'entrées lexicales
<b>Verbes</b>	270
<b>Adjectifs</b>	632
<b>Noms</b>	296
<b>Adverbes</b>	598
<b>Interjections</b>	51
<b>Expressions</b>	178
<b>Émoticônes</b>	79
<b>Total</b>	2100

Tableau 4.8 : Répartition des entrées lexicales selon leur catégorie syntaxique

## 4.7. Conclusion

Dans ce chapitre, nous avons pu voir deux méthodes d'extension de lexique à partir d'un lexique germe.

La première méthode est une méthode basée sur dictionnaire, qui utilise la taxinomie de concepts de Synapse Développement afin de propager la polarité de termes germes. Les résultats apportés sont assez faibles en termes d'évaluation directe du lexique, mais corrects en termes d'utilisabilité.

La seconde méthode exploite l'algorithme de calcul de spin de l'électron et le score de proximité du dictionnaire de synonymes de Synapse. Les résultats obtenus, avec un lexique germe restreint, sont tout à fait honorables, avec une pertinence (accuracy) de 90 %.

Il est par contre difficile de généraliser efficacement ces méthodes à d'autres traits sémantiques. Les essais de généralisation de la première méthode montre des résultats insuffisants en matière de rappel pour l'intensité et la catégorisation sémantique, mais correct au niveau précision pour la catégorisation sémantique. La seconde méthode n'est par contre pas généralisable aux catégories sémantiques : elle repose sur un classifieur binaire basé sur l'opposition entre deux catégories. De plus, le score observé en sortie n'est pas suffisamment corrélé à l'intensité du terme pour obtenir un lexique fiable d'intensités.

Afin de soutenir l'ensemble de l'architecture, et après étude des méthodes de générations automatiques de lexique exposées dans ce chapitre, nous avons opté pour l'utilisation d'un lexique d'opinion manuellement constitué dans le cadre du projet CASOAR, tout à fait adapté à nos besoins, notamment en termes de traits sémantiques, de qualité, et de compatibilité avec les outils techniques.

Ce lexique est utilisé dans l'ensemble des modules de la chaîne de traitement, et en particulier dans le module de détection de la subjectivité au niveau d'un segment de discours, dont la description fait l'objet du chapitre suivant.



## 5. Détection de la subjectivité

---

### 5.1. Introduction

La détermination de la subjectivité et de la polarité d'un segment de texte est l'un des axes de recherche les plus dynamiques de ces dernières années : citons notamment (Pang and Lee 2004), (Riloff, Wiebe, and Phillips 2005), (Wilson, Wiebe, and Hoffmann 2009)

La première tâche généralement identifiée est de distinguer entre une proposition objective (i.e. purement factuelle) ou subjective (i.e. contenant une opinion). La classification en termes de polarité des éléments subjectifs vient ensuite, afin d'extraire les propositions positives et négatives, voire neutres, parmi celles classées comme subjectives (cf. chapitre 1 et 2).

#### 5.1.1. Contexte scientifique

Les deux tâches de classification de subjectivité et de polarité étant liées (la première constitue un filtre pour la seconde), une approche commune repose sur la polarité *a priori* des mots et expressions rencontrés dans le segment discursif. En effet, la présence de mots ou expressions appartenant à un lexique de termes subjectifs est un indice fort de subjectivité.

Néanmoins, (Polanyi and Zaenen 2006) ont montré que la polarité *a priori* seule n'est pas toujours suffisante pour une détermination fine de la subjectivité et de l'opinion, et que l'on a besoin pour cela de tenir compte du contexte.

En effet, le contexte local fournit des informations utiles de deux manières différentes. Tout d'abord, il peut être utilisé afin de désambigüiser sémantiquement un terme afin de déterminer s'il est employé dans un sens subjectif ou objectif (Akkaya, Wiebe, and Mihalcea 2009). Il peut également être utilisé afin de déterminer si un opérateur vient altérer la polarité *a priori* d'une expression d'opinion (Kennedy and Inkpen 2006), (Wilson, Wiebe, and Hoffmann 2009) : en effet, si les négations et opérateurs d'intensité ne changent pas le caractère subjectif d'une expression, certaines modalités peuvent par contre avoir un impact : par exemple, dans la phrase « Ce jeu ne devrait pas être commercialisé. », la modalité donne son caractère subjectif au segment de discours ; énoncer le caractère « commercialisé » d'un jeu n'est pas en soit une opinion.

De même, le contexte global peut être utile pour identifier des opinions implicites, et pour améliorer ainsi la reconnaissance de l'opinion globale.

À notre connaissance, peu de travaux ont entrepris de prendre en compte le discours dans le cadre d'une analyse d'opinion au niveau de la clause ou de la phrase. (Pang and Lee 2004) montrent que les phrases subjectives et objectives ont tendance à apparaître groupées. (Asher, Benamara, and Mathieu 2009) ont développé un schéma d'annotation pour une analyse d'opinion fine, prenant en compte le contexte local, et s'appuyant sur les relations discursives de la SDRT. (Taboada, Voll, and Brooke 2008) ont utilisé un *parser* de discours RTS (Rhetorical Structure Theory), et, afin de calculer l'orientation sémantique globale, ont pondéré plus fortement les *nuclei* que leurs

satellites. (Somasundaran 2010) propose un traitement discursif permettant d'améliorer la reconnaissance de polarité au niveau de la phrase et la détection de l'opinion globale. Enfin, plus récemment, (Zhou et al. 2011) proposent une méthode non-supervisée pour reconnaître des relations de discours de la RST et ainsi éliminer certaines ambiguïtés de polarité au sein d'une phrase, et (Zirn et al. 2011) se sont basés sur la présence ou l'absence de relation RST de contraste entre deux segments pour reconnaître la polarité d'un segment.

Ce travail a été réalisé en collaboration avec Farah Benamara, Vladimir Popescu, et Yannick Mathieu, et a donné lieu à la publication (Benamara et al. 2011).

### 5.1.2. Objectifs

Dans ce chapitre, nous verrons une méthode de classification automatique de segments de texte permettant de distinguer les segments subjectifs des segments objectifs. Afin d'intégrer ce travail dans une chaîne d'analyse de l'opinion basée sur le discours, et comme vu au chapitre 2, nous avons choisi comme segments les Unités Discursives Élémentaires (UDE) de la SDRT.

L'analyse au niveau de la phrase n'est pas adaptée pour effectuer une analyse de sentiment basée sur le contexte. En effet, en plus des clauses objectives, une unique phrase peut contenir une multitude de clauses porteuses d'opinion, reliées entre elles par divers relations rhétoriques. Prenons par exemple la phrase suivante :

« En dépit d'un excellent choix d'acteurs et d'images magnifiques, le réalisateur ne parvient pas à nous faire oublier l'absence totale de scénario. »

Dans cette phrase, trois opinions portant sur le film sont présentes. Les deux premières, « excellent choix » et « images magnifiques », sont en contraste avec la troisième « absence totale de scénario ».

Effectuer une classification au niveau de la clause n'est pas non plus souhaitable : plusieurs expressions d'opinion peuvent en effet être reliées entre elles par des relations rhétoriques dans une même clause. Par exemple, dans la phrase « Le film est bien mais long. », un contraste marqué relie les deux opinions positive (bien) et négative (long). Par conséquent, il est nécessaire d'effectuer la classification à un niveau plus fin : le niveau du segment.

## 5.2. Catégories de subjectivité

### 5.2.1. Aller au-delà d'une simple classification binaire

#### 5.2.1.1. Définition du problème et contexte scientifique

La détection de la subjectivité pour l'analyse de sentiments ne peut être réduite à une classification binaire objectifs vs. subjectifs. En effet, considérons l'exemple suivant, extrait de notre corpus de commentaires web de films :

(1) [ Le film n'était pas mal, ]<sub>a</sub> [ même si certaines personnes ont quitté la salle. ]<sub>b</sub>



- (2) [ Laborieux ]<sub>a</sub> [ et copier/coller de la première partie. ]<sub>b</sub>  
 (3) [ Ce film est poignant, ]<sub>a</sub> [ et les acteurs excellents. ]<sub>b</sub> [ Il restera longtemps dans votre étagère à DVD. ]<sub>c</sub>  
 (4) [ Je suppose ]<sub>a</sub> [ que la politique gouvernementale a échoué. ]<sub>b</sub>

Les segments (1.a), (2.a), (3.a), (3.b) et (4.b) sont des segments subjectifs comportant une opinion explicite. (1.b), (2.b) et (3.c) portent des opinions implicites. Enfin, (4.a) porte une marque de subjectivité, mais ne présente pas un caractère évaluatif.

(Wiebe, Wilson, and Cardie 2005) ont proposé un schéma d'annotation qui distingue entre mentions explicites d'états privés et actes de langage exprimant un état privé. (B. Liu 2010) a également observé que l'ensemble des phrases subjectives et l'ensemble des phrases portant une opinion ne peuvent pas être considérés comme égaux : au mieux, on peut considérer l'ensemble des phrases portant une opinion comme étant un sous-ensemble de l'ensemble des phrases subjectives.

Poursuivant dans la lignée de ces observations, nous proposons de distinguer entre quatre catégories de subjectivité au niveau du segment.

#### 5.2.1.2. Rappel : classification des segments

Nous reprenons dans cette section la classification des segments introduites dans le chapitre 2.

##### *Segments objectifs purs (O)*

Les segments objectifs purs (O) ne comportent pas d'opinion. Plus explicitement, ils ne comportent ni élément lexical subjectif, ni opinion implicite.

##### *Segments subjectifs évaluatifs explicites (SE)*

Les segments subjectifs explicites (SE) comportent au moins une opinion évaluative explicite, c'est-à-dire au moins une expression d'opinion. Leur polarité peut être positive, négative, ou neutre hors contexte (i.e. leur polarité peut être positive ou négative suivant le contexte).

##### *Segments subjectifs non évaluatifs (SN)*

Les segments subjectifs non évaluatifs (SN) sont des segments subjectifs qui introduisent l'opinion. Ils comportent une expression de reportage, en général un verbe d'introduction de la parole.

##### **Remarque importante :**

Les segments subjectifs non-évaluatifs ne rentrent pas dans la notion de polarité neutre hors contexte vue précédemment, puisqu'ils ne comportent pas d'opinion.

##### *Segments subjectifs implicites (SI)*

Les segments subjectifs implicites (SI) sont des segments porteurs d'opinion, mais ne contenant pas d'expression d'opinion : l'opinion, positive ou négative, est sous-entendue dans le contenu propositionnel factuel. Privés de leur contexte, il est impossible de détecter l'opinion, et ces segments deviennent objectifs.

### *Segments subjectifs mixtes (SEI)*

Les segments subjectifs mixtes comportent à la fois une opinion explicite, et une opinion implicite distincte de la précédente. Souvent, le sujet de l'opinion explicite est différent et n'a rien à voir avec le sujet de l'opinion implicite.

Dans le contexte de la phase d'apprentissage automatique, nous ne distinguons pas cette classe : ces segments seront considérés à la fois comme segments subjectifs explicites (SE) et implicites (SI). Ce choix est dû au trop faible nombre de segments SEI par rapport aux autres classes de segments.

#### **5.2.1.3. Avantages de cette classification**

Distinguer parmi ces quatre classes de segments offre plusieurs avantages sur une simple classification binaire objectifs/subjectifs.

Tout d'abord, cela nous permet de distinguer entre les expressions purement subjectives (qui appartiennent aux segments SE), et les expressions dont la subjectivité relève de l'implicite (segments SI). Ensuite, les classes définies ici peuvent être utilisées pour améliorer la classification en termes de polarité, en permettant le filtrage des segments de catégorie O ou SN. Enfin, ces classes peuvent également améliorer la reconnaissance de l'intensité de l'opinion globale : les segments SN, en renseignant le degré de certitude de l'auteur de l'opinion, peuvent jouer un rôle de filtre très important.

Des travaux récents ont entrepris de classer automatiquement les sentiments implicites. Par exemple, (Greene and Resnik 2009) ont utilisé pour cela la sémantique lexicale et la syntaxe. (Musat and Trausan-Matu 2010) ont étudié l'impact des opérateurs influençant la polarité sur la détection des sentiments implicites dans les textes économiques. Néanmoins, à notre connaissance, la distinction automatique entre segments évaluatifs et non-évaluatifs d'une part et opinions implicites et explicites d'autre part, à partir d'éléments contextuels, constitue un travail original au sein de la communauté.

## **5.3. Classifications automatiques**

### **5.3.1. Partitions binaires de l'ensemble des segments**

Au sein de notre corpus, les classes SE, O, SN, et SI sont en proportions déséquilibrées. De plus, reconnaître les membres de la classe SI est une tâche complexe, parfois même pour un juge humain. Par conséquent, nous avons décidé de définir deux partitions binaires orthogonales de l'ensemble des segments.

#### **5.3.1.1. Segments objectifs hors contexte ( $O_{HC}$ ) et subjectifs hors contexte ( $S_{HC}$ )**

La première partition est suivant leur subjectivité hors contexte, c'est-à-dire suivant le fait qu'en l'absence de contexte, il y ait présence ou non de subjectivité. Les segments subjectifs hors contexte sont les segments SE ou SN, qui sont intrinsèquement subjectifs indépendamment de

leur contexte. Les segments objectifs hors contexte, à l'inverse, ne sont pas subjectifs si on les prive de contexte.

Cette classification peut se résumer par le tableau suivant :

Objectifs hors contexte	Subjectifs hors contexte
O	SE
SI	SN

### 5.3.1.2. Segments évaluatifs ( $Eval_{Op}$ ) et non évaluatifs ( $Non\_Eval_{Op}$ )

La seconde partition distingue la présence ou non d'un caractère évaluatif, i.e. d'opinion, au segment (contexte pris en compte). Les segments SE et SI sont évaluatifs. Les segments SN annoncent un contenu évaluatif mais n'en porte pas un propre. Les segments objectifs purs (O) sont par définition non évaluatifs.

Cette classification peut se résumer par le tableau suivant :

Évaluatifs	Non évaluatifs
SE	O
SI	SN

## 5.3.2. Mise en place du classifieur

Sur le corpus *gold standard*, les partitions présentées donnent un total de 919 segments  $S_{HC}$  et 511 segments  $O_{HC}$  d'une part, et de 1083 segments  $Eval_{Op}$  et 347 segments  $Non\_Eval_{Op}$  d'autre part.

Deux classifieurs binaires sont donc construits, un pour chaque partition de l'ensemble des segments vues précédemment.

### 5.3.2.1. Traits utilisés pour la classification

Cette section présente les traits utilisés par les classifieurs automatiques binaires.

On distingue quatre catégories pour ces traits de classification, suivant la portée de l'information (locale ou contextuelle), et suivant son origine (lexicale, stylistique, syntaxique, ou discursive).

Parmi les traits locaux, on distingue les traits lexicaux, qui sont les informations issues de la reconnaissance d'un terme ou d'une expression décrite par lexique. Les traits stylistiques correspondent aux informations issues de la mise en forme du texte (ponctuation, majuscules, émoticônes). Les traits syntaxiques sont les informations issues ou déduites de l'analyse syntaxique du texte.

Par opposition à ces trois premières catégories, les traits contextuels ou discursifs comportent des informations issues de l'analyse discursive du texte, à un niveau macro discursif (place d'un

segment dans le document), ou micro discursif (marqueurs discursifs indiquant la présence de relations SDRT).

L'ensemble des traits considérés sont binaires : ils prennent soit la valeur 1 (activé), soit la valeur 0 (inactifs/désactivés). Pour plus de clarté, chaque trait est associé à un identifiant, spécifié entre crochets dans le titre de chaque sous-section.

### *Traits lexicaux*

#### Présence de termes du lexique [L1]

Ce trait binaire concerne la présence ou non, au sein de l'UDE, d'un nom, d'un adjectif, d'un adverbe de manière, d'un verbe, d'une expression, ou d'une interjection du lexique d'opinion, à l'exception des entrées signalant une modalité ou une négation.

#### **Exemple :**

(5) [ Ce restaurant est **bon**. ]

Bon est ici un adjectif d'opinion.

#### Présence de termes du lexique (désambiguïsé sémantiquement) [L2]

Ce trait reprend le précédent en prenant en compte la désambiguïstation sémantique : il est donc activé si l'UDE contient un terme du lexique, et si celui-ci est reconnu comme étant employé dans un sens subjectif.

#### **Exemples :**

(6) [ Ce restaurant est **parfait**. ]<sub>a</sub>

(7) [ Ce restaurant sert des **parfaits** au chocolat.<sup>44</sup> ]<sub>a</sub>

L'UDE (6.a) vérifie ce trait (présence de parfait dans un sens subjectif), mais pas l'UDE (7.a) : parfait désigne ici une pâtisserie, et n'est pas utilisé en temps que terme d'opinion.

#### Présence de modal [L3]

Ce trait binaire est activé si l'UDE contient un terme ou une expression présente dans le lexique et exprimant une modalité.

#### **Exemple :**

(8) [ Ce restaurant **peut** être bon. ]<sub>a</sub>

---

<sup>44</sup> [http://www.marmiton.org/recettes/recette\\_parfait-au-chocolat\\_67790.aspx](http://www.marmiton.org/recettes/recette_parfait-au-chocolat_67790.aspx)

### Traits stylistiques

#### Présence d'émoticône [L4]

Ce trait est activé si la phrase comporte un smiley (émoticône). En général, les émoticônes sont détachés à droite dans la structure de la phrase, et sont alors considérés comme un segment à part.

#### Exemple :

(9) [ Ce film est cool ]<sub>a</sub> [ :) ]<sub>b</sub>.

L'UDE (9.b) active ce trait (mais pas l'UDE (9.a)).

#### Capitalisation [L5]

Ce trait est activé lorsque l'UDE contient un mot entièrement en majuscules, à l'exception d'un acronyme connu (par exemple, DVD).

#### Exemple :

(10) [ Ce film est **FANTASTIQUE.** ]<sub>a</sub>

#### Ponctuation emphatique [L6]

Ce trait est activé si l'UDE contient une séquence de points d'exclamation/d'interrogation ("??", "!?", ...).

En pratique le patron reconnu suit l'expression régulière suivante : (!|?)(!|?)+.

#### Exemple :

(11) [ Ce film est fantastique **!!!** ]<sub>a</sub>

### Traits syntaxiques

#### Comparatifs et superlatifs relatifs [L7]

Ce trait est activé si le segment contient une structure de comparatif ou superlatif. La détection s'effectue à partir d'une série de patrons conçus manuellement, exploitant les sorties de l'analyseur syntaxique.

Les patrons utilisés ont été conçus et adaptés par rapport au corpus de développement. Ils sont disponibles, sous forme de code Myéline, en Annexe A.2.

#### Dans la portée d'un verbe de reportage ou de conseil [L8]

Ce trait est actif si l'UDE est contenue dans la portée d'un verbe de reportage ou de conseil.

En pratique, deux patrons sont utilisés :

- L'UDE se trouve dans une proposition dont le verbe principal est un verbe de reportage ou de conseil
- L'UDE se trouve dans une proposition subordonnée, et le verbe principal de la proposition principale associée est un verbe de reportage ou de conseil

### Contient un terme dans la portée d'un modal [L9]

Ce trait est actif si un des mots du segment est dans la portée d'un opérateur de modalité.

L'algorithme de détermination de la portée pour les opérateurs sera décrit dans le chapitre suivant.

### Patrons syntaxiques subjectifs [L10]

Ce trait est actif si, au sein de l'UDE, un patron syntaxique porteur de subjectivité est reconnu.

### UDE détachée à gauche [L11]

Ce trait est actif si l'UDE considérée est détachée à gauche, i.e. si l'UDE est une proposition circonstancielle de temps, ou une circonstancielle de lieu.

#### Exemples :

(12) [ Hier au soir, ]<sub>a</sub> [ un elfe a mangé une pomme. ]<sub>b</sub>

(13) [ Sur la plage ensoleillée, ]<sub>a</sub> [ un gnome bricolait. ]<sub>b</sub>

(14) [ Soyons clair, ]<sub>a</sub> [ ces fait sont inintéressants. ]<sub>b</sub>

L'UDE (12.a) active ce trait : il s'agit d'un complément circonstanciel de temps. L'UDE (13.a) active également ce trait, il s'agit d'un complément circonstanciel de lieu. Par contre, l'UDE (14.a) n'active pas ce trait.

### Traits contextuels / discursifs

Les traits contextuels sont au nombre de cinq. Ils exploitent des informations présentes dans les segments voisins immédiats de l'UDE considérée, i.e. l'UDE précédent l'UDE courante, et l'UDE suivant l'UDE courante.

### Présence à gauche de verbe de reportage ou de conseil non polarisé [D1]

Ce trait teste la présence d'un verbe de reportage ou de conseil non polarisé dans l'UDE précédente, c'est-à-dire l'existence d'une relation d'*Attribution* entre l'UDE précédente et l'UDE courante.

#### Exemples :

(15) [ Pierre **affirme** que ]<sub>a</sub> [ ce restaurant est bon. ]<sub>b</sub>

(16) [ Pierre **suggère** que ]<sub>a</sub> [ nous nous rendions dans ce restaurant. ]<sub>b</sub>

Les segments (15.b) et (16.b) voient ce trait activé en raison de la présence d'un verbe de reportage (15.a) ou de conseil non polarisé (16.a) dans le segment précédent.

### Présence d'émoticône dans l'UDE suivante : relation "Commentaire" [D2]

Ce trait teste la présence dans l'UDE suivante d'un émoticône. La présence d'un émoticône dans le segment suivant indique une relation discursive *Commentaire* entre l'UDE courante et l'UDE suivante.

#### Exemple :

[ Ce restaurant est cool ]<sub>a</sub> [ :). ]<sub>b</sub>

### UDE en relations discursive marquée avec l'UDE précédente ou suivante Eval<sub>Op</sub> [D3]

Étant donné l'ensemble de relations *ReIs* suivant :

$$ReIs = \{ Continuation, Parallèle, Contraste, Alternation \}$$

Ce trait teste le patron suivant :

- {L'UDE courante possède un marqueur de discours d'une relation de *ReIs* } ET {l'UDE précédente est de classe Eval<sub>Op</sub>}

OU

- {L'UDE suivante possède un marqueur de discours d'une relation de *ReIs* } ET {l'UDE suivante est de classe Eval<sub>Op</sub>}

Ce patron pose le problème de la détermination a priori de la classe des UDE voisines. Ce problème peut être résolu en exécutant plusieurs itérations de classification. La méthode sera détaillée dans la section 3.3.2.2 (repérage pratique des traits discursifs).

#### Exemples :

- (17) [ Ce restaurant est bon ]<sub>a</sub> [ bien que glucose, ]<sub>b</sub> [ et situé à Poing Ravier. ]<sub>c</sub>
- (18) [ Ce restaurant est bon, ]<sub>a</sub> [ bien que glucose, ]<sub>b</sub> [ mais de toutes façon chaleureux. ]<sub>c</sub>
- (19) [ Ce restaurant est bon, ]<sub>a</sub> [ glucose, ]<sub>b</sub> [ chaleureux. ]<sub>c</sub>

Pour ces exemples, supposons que les termes "bon" et "chaleureux" appartiennent au lexique de termes de sentiment. Le terme "glucose", utilisé ici comme adjectif et tiré d'un sketch des Inconnus<sup>45</sup>, n'a par contre pas nécessairement vocation à apparaître dans ce lexique : il est trop spécifique, et son interprétation n'est pas nécessairement généralisable à d'autres instances du terme.

Une pré-classification donnerait par exemple les classes suivantes pour les segments :

- Eval<sub>Op</sub> pour les segments (17.a), (18.a), (18.c), (19.a), (19.c)
- NonEval<sub>Op</sub> pour les autres segments

Pour ce trait :

- Le segment (17.b) active ce trait : "bien que" est un marqueur de contraste, et l'UDE précédente est Eval<sub>Op</sub> ;
- Le segment (18.b) active aussi ce trait, pour la même raison, mais aussi car l'UDE suivante est Eval<sub>Op</sub> et contient le marqueur de contraste "mais" ;
- En revanche, l'absence de marqueurs fait que le segment (19.b) n'active pas ce trait.

### UDE en relations discursive marquée avec les UDE précédente et suivante Eval<sub>Op</sub> [D4]

Ce trait est similaire au précédent, mais teste la présence conjointe de deux relations de discours avec les deux voisins de classe Eval<sub>Op</sub>.

---

<sup>45</sup> [http://fr.wikipedia.org/wiki/Les\\_Inconnus](http://fr.wikipedia.org/wiki/Les_Inconnus)

Le patron devient donc le suivant :

- {L'UDE courante possède un marqueur de discours d'une relation de discours} ET {l'UDE précédente est de classe Eval<sub>Op</sub>}
- OU
- {L'UDE suivante possède un marqueur de discours d'une relation de discours} ET {l'UDE suivante est de classe Eval<sub>Op</sub>}

### Exemples :

Si l'on reprend les exemples (17) à (19), on obtient pour ce trait :

- Le segment (17.b) n'active pas ce trait : "bien que" est un marqueur de contraste, et l'UDE précédente est Eval<sub>Op</sub>, mais l'UDE suivante n'est pas Eval<sub>Op</sub> ;
- Le segment (18.b) active ce trait : "bien que" est un marqueur de contraste, et l'UDE précédente est Eval<sub>Op</sub>, et l'UDE suivante est Eval<sub>Op</sub> et contient le marqueur de contraste "mais" ;
- l'absence de marqueurs fait que le segment (19.b) n'active pas ce trait.

#### Dernière UDE du document [D5]

Ce trait macro-discursif teste si l'UDE est la dernière du document. La dernière UDE du document se voit attribuer un comportement spécifique en raison d'un effet de « conclusion » observé sur le corpus de développement : les commentaires web se terminent ainsi souvent par une recommandation ou un conseil.

### 5.3.2.2. Repérage pratique des traits discursifs

#### Principe

Afin de déterminer la valeur des traits discursifs, nous avons besoin de nous appuyer sur une classification préexistante sur le caractère évaluatif des segments.

Une telle classification n'est bien sûr pas disponible pour un texte brut. Il est par contre possible d'estimer la classe d'un segment à l'aide d'une classification itérative.

Une première classification  $I_0$  est effectuée en utilisant les traits ne nécessitant pas de pré-classification. Cette classification  $I_0$  est dite de *bootstrapping*. Ensuite, pour chaque itération  $I_N$ , la classification est effectuée à partir du résultat de la classification  $I_{N-1}$ .

#### Convergence

L'idée principale derrière les deux traits discursifs considérés est de désambigüiser un segment subjectif implicite d'un segment objectif, en s'appuyant sur les caractéristiques de subjectivité des segments en relation avec lui. Sans l'utilisation de ce trait, de tels segments ne sont pas détectés. Au fur et à mesure des itérations, de nouveaux segments sont classés comme évaluatifs ; comme le nombre de segments est fini, l'algorithme devrait converger, idéalement vers une reconnaissance des segments subjectifs implicites en segments évaluatifs. Le critère utilisé pour arrêter le processus est la stabilisation de la F-mesure  $F$  définie comme suit :  $F_{K+1}$  est la F-mesure observée sur l'itération  $K+1$ , en prenant l'itération  $K$  comme référence de classification.



Le seuil de convergence est arbitrairement fixé à  $\varepsilon = 0.01$ . Le critère de convergence est donc le suivant :  $|F_{k+1}-F_k| < \varepsilon \Rightarrow F$  considérée stable

### 5.3.2.3. Répartition des traits sélectionnés par classifieur

Chacune des deux tâches de classification (classification  $O_{HC}$  vs.  $S_{HC}$  et classification  $Eval_{Op}$  vs.  $NonEval_{Op}$ ) s'appuie sur un jeu propre de traits parmi ceux vu précédemment.

#### *Premier classifieur : segments objectifs hors contexte et subjectifs hors contexte (classification S)*

La première classification concerne les propriétés hors contexte du segment : elle s'appuie donc exclusivement sur des traits locaux.

Après essais sur le corpus de développement, les traits suivants se sont révélés pertinents pour cette classification :

- L1 : présence de termes du lexique
- L2 : présence de termes du lexique (désambiguïsé)
- L4 : présence d'émoticônes
- L8 : dans la portée d'un verbe de reportage ou de conseil

La tâche traitée par ce classifieur étant assez simple, on peut remarquer que les traits utilisés concernent tous une recherche dans le lexique.

#### *Second classifieur : segments évaluatifs et non évaluatifs (classification Op)*

La tâche du second classifieur est plus complexe, et nécessite de prendre en compte les informations locales et contextuelles.

#### *Tableau résumé*

Portée	Type	Trait	S	Op
Locale	Lexical	L1 : présence de termes du lexique	√	√
		L2 : présence de termes du lexique (désambiguïsé)	√	√
		L3 : présence de modal	-	√
	Stylistique	L4 : présence d'émoticônes	√	√
		L5 : capitalisation	-	√
		L6 : ponctuation emphatique	-	√
	Syntaxique	L7 : comparatifs et superlatifs relatifs	-	√
		L8 : dans la portée d'un verbe de reportage ou de conseil	√	√
		L9 : contient un terme dans la portée d'un modal	-	√
		L10 : patrons syntaxiques	-	√
		L11 : UDE détachée à gauche	-	√
Contexte	-	D1 : présence à gauche vb de reportage/conseil non polarisé	-	√
		D2 : Présence d'émoticônes dans l'UDE suivante	-	√
		D3 : lien discursif marqué avec l'UDE préc. ou l'UDE suiv.	-	√
		D4 : lien discursif marqué avec l'UDE préc. et l'UDE suiv.	-	√
		D5 : dernière UDE du document	-	√

Tableau 5.1 : Traits utilisés pour la classification test

#### 5.3.2.4. Classification finale

Les partitions étudiées étant mutuellement indépendantes, ces deux classifieurs sont indépendants l'un de l'autre. Par conséquent, ils peuvent être lancés en parallèle. Une fois chacune des classifications effectuées, l'ensemble des UDE est divisé suivant deux partitions binaires :  $S_{HC}$  vs.  $O_{HC}$ , et  $Eval_{Op}$  vs.  $Non\_Eval_{Op}$ .

Afin de retrouver les quatre classes de segments présentées en section 3.2, il suffit d'appliquer les règles suivantes :

- Si une UDE est à la fois  $S_{HC}$  et  $Eval_{Op}$ , alors il s'agit d'une UDE de classe SE
- Si une UDE est à la fois  $S_{HC}$  et  $Non\_Eval_{Op}$ , alors il s'agit d'une UDE de classe SN
- Si une UDE est à la fois  $O_{HC}$  et  $Eval_{Op}$ , alors il s'agit d'une UDE de classe SI
- Si une UDE est à la fois  $O_{HC}$  et  $Non\_Eval_{Op}$ , alors il s'agit d'une UDE de classe O

## 5.4. Résultats

### 5.4.1. Données et statistiques d'annotations

Notre corpus est composé de 136 commentaires web de films, extraits depuis le site internet Allociné.com<sup>46</sup>. Parmi celles-ci 115 sont utilisées pour le développement (gold standard), et 21 pour le test<sup>47</sup>. Trois juges familiers du domaine ont effectués une annotation en deux temps : segmentation discursive du texte, puis classification manuelle de chacun des segments identifiés.

La phase de segmentation a été effectuée conformément au manuel d'annotation existant, compilé pendant le projet d'annotation discursive ANNODIS. Pour la classification des segments, nous avons élaboré un manuel d'annotations spécifique, guidant chaque annotateur suivant les définitions vues en section précédente. Les juges ont participé préalablement à l'annotation du corpus à un entraînement sur 10 commentaires web (non comptabilisés dans le corpus final). Ensuite, séparément, ils ont doublement annotés chaque document du corpus final.

La phase de classification a résulté en un kappa de Cohen moyen de 0.7 pour les segments SE, 0.72 pour les segments O, 0.61 pour les segments SN, et 0.54 pour les segments SI. Les deux derniers scores sont plus modérés, ce que nous interprétons comme un artefact de la longueur des textes. En effet, plus un texte est long, plus il est difficile pour le sujet humain de repérer le contexte discursif avec précision : celui-ci est parfois lointain, et la tâche d'annotation nécessite de passer plus de temps sur le texte que ne le ferait un simple lecteur. L'étude de cette hypothèse, bien qu'intéressante, est néanmoins au-delà de la portée de cette section. Ces chiffres sont néanmoins tout à fait en accord avec l'état de l'art en matière de distinction entre opinions explicites et

---

<sup>46</sup> <http://www.allocine.com>

<sup>47</sup> Les 115 documents de développement ont constitué l'annotation exploratoire pour la constitution du corpus CASOAR présenté au chapitre 2.

implicites (Toprak, Jakob, and Gurevych 2010). Afin d'établir le corpus final (*gold standard*), les cas à problèmes ont été résolus par discussion entre les annotateurs.

## 5.4.2. Annotation automatique

### 5.4.2.1. Déroulement

Plusieurs expérimentations ont été réalisées pour tester la validité de notre approche de classification de la subjectivité des segments, en particulier pour les traits contextuels.

Tout d'abord, les deux classifieurs présentés en section 3.3 ont été construits en validant les traits à partir d'une validation croisée (*10-fold cross validation*) sur le corpus de développement. Dans un second temps, les deux classifieurs ont été évalués sur les 21 documents du corpus de test, l'entraînement étant effectué sur la totalité du corpus de développement.

### 5.4.2.2. Environnement technique

#### *Type de classifieur*

Notre choix s'est porté sur un classifieur de type Machine à vecteurs de support (SVM – (Burges 1998)). En pratique, nous avons utilisé le package SVM-light<sup>48</sup>.

Dans la mesure où les espaces de vecteurs de traits sont non linéairement séparables, nous avons opté pour l'utilisation de kernels polynomiaux dans la phase d'entraînement.

#### *Environnement logiciel*

Notre classification s'appuie sur une analyse syntaxique du texte. Nous avons choisi le composant d'analyse syntaxique de Synapse Développement, appelé via un plug-in UIMA s'adaptant très bien à notre chaîne de test. Ce composant présente également l'avantage d'être couplé avec un module de désambiguïsation sémantique compatible avec notre lexique (cf. section 2 pour plus de détails sur cette jointure).

### 5.4.2.3. Résultats

Pour chaque classifieur, nous présentons les résultats en termes d'accuracy, précision, rappel et F-mesure, de la *10-fold cross validation* sur le corpus de développement, et de la classification finale sur le corpus de test.

#### *Classification S*

Pour la classification S, les traits présentés en section 3.3 sont déclinés en trois versions :

- S0 : Baseline : traits L1 (sauf adverbes de manière) + L4 + L8
- S1 : Baseline + adverbes : traits L1 (incluant adverbes de manière) + L4 + L8
- S2 : Baseline + adv. + désambiguïsation : traits L1 + L2 + L4 + L8

Le Tableau 5.2 présente les résultats obtenus.

---

<sup>48</sup> <http://svmlight.joachims.org/>

Jeux de traits (cumulatifs)	Accuracy		Précision		Rappel		F-mesure	
	Dev	Test	Dev	Test	Dev	Test	Dev	Test
S0 : Baseline	74.30	68.79	77.82	65.50	83.35	79.39	80.49	71.77
S1 : S0 + adverbes	74.68	70.00	77.85	66.18	<b>84.25</b>	<b>81.82</b>	80.92	73.17
S2 : S1 + désambiguïsation	<b>82.31</b>	<b>70.91</b>	<b>87.54</b>	<b>67.69</b>	84.08	80.00	<b>85.77</b>	<b>73.33</b>

Tableau 5.2 : Résultats pour la classification S

Nous pouvons observer que l'ajout des adverbes de manière améliore substantiellement les performances du classifieur en termes de rappel, sans nuire à sa précision.

L'ajout de la désambiguïsation sémantique induit une amélioration de la précision significative (notamment sur le corpus de développement, où l'on observe une augmentation de plus de 10 %), au prix d'une baisse du rappel. Dans les deux cas, on observe une augmentation de la F-mesure correspondante. Cette augmentation est significative pour le corpus de développement (+5 % environ), mais marginale pour le corpus de test (+0.15 % environ).

### Classification Op

Pour la classification Op, nous avons testé les combinaisons de traits suivantes :

- Op0 : Baseline : traits L1 + L3 + L4
- Op1 : Op0 + capitalisation et ponctuation : traits L1 + L3 à L6
- Op2 : Op1 + comparatifs et superlatifs : traits L1 + L3 à L7
- Op3 : Op2 + autres syntaxiques (sauf L11) : traits L1 + L3 à L10
- Op4 : Op3 + désambiguïsation sémantique : traits L1 à L10
- Op5 : Op4 + traits contextuels non marqués : traits L1 à L10 + D1 + D2 + D5
- Op6 : Op5 + traits contextuels marqués (sans itération) : traits L1 à L10 + traits contextuels
- Op7 : Op6 + EDU détachée à gauche : tous les traits
- Op8 : Op6 + itération : ajout de la phase itérative

Les combinaisons de traits Op6 et Op8 ne diffèrent que par le mode de calcul des traits contextuels marqués. Dans le premier cas, ceux-ci ne sont évalués que par l'étape de bootstrapping (cf. section 3.3.2.2). Dans le second cas, plusieurs itérations sont effectuées, jusqu'à obtention du critère de stabilité (cf. section 3.3.2.2). Le Tableau 5.3 présente les résultats obtenus :

Jeux de traits (cumulatifs)	Accuracy		Précision		Rappel		F-mesure	
	Dev	Test	Dev	Test	Dev	Test	Dev	Test
Op0 : Baseline	75.82	73.33	76.25	73.56	98.93	<b>99.59</b>	86.12	<b>84.61</b>
Op1 : + cap. + pct.	76.51	72.73	76.80	73.39	98.94	98.77	86.48	84.20
Op2 : + comp. & sup.	76.52	72.73	76.73	73.39	<b>99.13</b>	98.77	86.50	84.20
Op3 : + autres syntaxiques	76.66	72.42	76.94	73.31	98.84	98.35	<b>86.53</b>	84.00
Op4 : + désambiguïsation	77.39	73.33	81.14	78.6	91.98	87.65	86.22	82.87
Op5 : + contexte non marq.	77.32	75.15	80.35	79.49	93.31	89.30	86.35	84.10
Op6 : + contexte marq.	<b>78.35</b>	75.15	<b>83.74</b>	<b>85.15</b>	88.33	80.25	85.97	82.62
Op7 : + détach. gauche	77.18	74.55	79.15	77.89	94.71	91.36	86.23	84.08
Op8 : Op6 + itérations	77.68	<b>75.45</b>	82.78	84.32	89.02	81.89	85.79	83.08

Tableau 5.3 : Résultats pour la classification Op

Comme on peut le constater, l'ajout par rapport à la baseline de traits stylistiques améliore légèrement les résultats sur le corpus de développement, mais réduit légèrement ceux obtenus sur le corpus de test. Ceci peut être dû à un usage moins régulier des ponctuations emphatiques sur le corpus de test.

L'ajout des patrons de comparatifs et superlatifs relatifs dégrade légèrement la précision sur le corpus de développement, mais donne les meilleurs résultats observés en termes de rappel sur ce même corpus. Sur les données de test, ce trait n'a pas eu d'impact.

L'ajout des patrons syntaxiques spécifiques de subjectivité dégrade légèrement le rappel. Dans le même temps, l'accuracy et la précision augmentent, et la F-mesure atteint le plus haut score observé sur le corpus de développement. Néanmoins, cette augmentation n'est pas observée sur le corpus de test, qui voit une légère baisse des performances.

L'ajout de la désambiguïsation sémantique dégrade significativement, comme attendu, le rappel : il s'agit avant tout d'un filtre appliqué sur un trait associé à une détection. Dans le même temps, la précision est augmentée. D'un point de vue global, l'accuracy augmente légèrement sur les deux corpus ; la F-mesure augmente légèrement sur le corpus de développement, mais baisse sur le corpus de test.

L'ajout des traits contextuels ne reposant pas sur une pré-classification dégrade la précision au profit d'une augmentation significative du rappel : cela conduit, sur les deux corpus, à une augmentation de la F-mesure.

L'ajout des traits contextuels nécessitant une classification itérative donne les meilleurs résultats en termes d'accuracy et de précision, ce qui constitue une valeur ajoutée non négligeable. Néanmoins, ceci est au prix d'un rappel fortement amoindri sur chacun des deux corpus, et d'une F-mesure légèrement plus faible. Cela peut s'expliquer par le peu de marqueurs de discours dans notre corpus : en effet, les traits contextuels ne peuvent alors rendre compte que d'une partie restreinte de l'ensemble des instances de relations de discours du texte. Cela nous montre que malgré le fait que la prise en compte des informations discursives soit utile pour la détection des segments subjectifs – les segments implicites peuvent difficilement être repérés par d'autres méthodes surfaciques – le manque de couverture reste un problème à surmonter.

Lors de la classification itérative, le facteur de convergence  $\epsilon = 0.01$  est atteint en au plus 2 itérations sur le corpus de développement, et 3 itérations sur le corpus de test. Comme attendu, accuracy et précision diminuent légèrement (environ 1 %) sur les données de test, puisque les classes des UDE adjacentes ne sont pas fournies par un oracle. Néanmoins, sur les données de test, l'accuracy augmente légèrement (de moins de 0.5 %). La précision augmente légèrement sur chacun des deux corpus (environ 1 % à chaque fois), ce qui implique que les imperfections de la classification itérative sont compensées par l'intérêt qu'elle apporte.

### *Classification globale (Op + S)*

Pour la tâche finale de reconnaissance des quatre catégories de segments, nous avons sélectionné trois combinaisons des meilleurs jeux de traits obtenus :

- Meilleur jeu classifieur S + meilleur jeu Op n'exploitant pas les traits contextuels
- Meilleur jeu classifieur S + meilleur jeu Op exploitant uniquement les traits contextuels non itératifs (D1, D2, D5)
- Meilleur jeu classifieur S + meilleur jeu Op exploitant les traits contextuels itératifs

Cette sélection nous permet d'étudier l'impact des traits contextuels sur la classification finale.

Nous avons obtenu les résultats suivants, exprimés en termes d'Accuracy :

Configuration	S		SN		O		SI	
	Dev	Test	Dev	Test	Dev	Test	Dev	Test
Meilleur classifieur S / Meilleur classifieur Op non contextuel	80.83	70.60	97.58	<b>96.06</b>	79.35	<b>75.75</b>	72.64	66.06
Meilleur classifieur S / Meilleur classifieur Op contextuel, non itératif	81.38	<b>73.33</b>	97.57	92.72	<b>79.96</b>	75.54	<b>79.02</b>	<b>72.12</b>
Meilleur classifieur S / Meilleur classifieur Op contextuel, itératif	<b>81.45</b>	73.03	<b>97.64</b>	93.03	79.21	75.54	77.88	70.90

Tableau 5.4 : Résultats pour la classification finale (Accuracy)

Sur les données de test, les traits contextuels permettent une meilleure détection des segments subjectifs, explicites comme implicites. En effet, sur le corpus de développement, nous pouvons observer que l'ajout des traits contextuels augmente les performances, exceptés pour les segments de classe SN, où le score est marginalement dégradé. Sur le corpus de test, les différences sont plus marquées : l'amélioration des performances est importante pour les segments SE et SI (respectivement 2.73 % et environ 6.06 % d'augmentation), mais la chute de performances sur les segments SN s'accroît aussi (3.34 %). La reconnaissance des segments O sur le corpus de test est également atténuée (-0.21 %).

Les améliorations sont moins marquées lorsque l'on introduit les traits itératifs, ce qui est cohérent avec les performances observées sur le seul classifieur Op. Néanmoins, même avec ces traits, les performances sur la reconnaissance de segments subjectifs SE et SI restent significativement meilleures que sans les traits contextuels. Sur le corpus de développement, on peut également remarquer une légère amélioration des performances sur les classes SE et SN par rapport à la version contextuelle non itérative du classifieur.

## 5.5. Conclusion et perspectives

Ce chapitre présente une méthode permettant de déterminer la subjectivité d'un segment, à partir d'informations locales et contextuelles. L'ajout d'informations contextuelles, notamment issues du discours, permet d'améliorer significativement la classification des segments, et en particulier la détection des segments porteurs d'opinions implicites.

Afin d'améliorer cette reconnaissance, il serait intéressant d'étudier le comportement des classifieurs en se basant sur la totalité des relations discursives, y compris celles non marquées.

Pour cela, l'usage d'un corpus annoté discursivement serait nécessaire ; le corpus CASOAR dans son annotation finale, présenté au chapitre 2, serait tout à fait adapté, mais n'était malheureusement pas encore disponible au moment de ces travaux. Il pourrait également être intéressant de tester l'impact d'un changement de classifieur sur les résultats de la classification : il pourrait notamment être intéressant de tester de remplacer les deux classifieurs parallèles par un unique classifieur CRF (Lafferty, McCallum, and Pereira 2001).

La classification des segments en termes de catégories de subjectivité est une première étape vers le calcul d'une opinion globale au niveau du document. Nous verrons dans le chapitre suivant comment calculer les attributs associés à l'opinion pour les segments explicites porteurs d'opinion reconnus grâce à cette méthode.





# 6. Analyse de polarité au niveau du segment

---

## 6.1. Introduction

Dans les sections précédentes, nous nous sommes concentrés sur la notion d'opinion ponctuelle, rattachée à une expression d'opinion. Bien que cette étape soit importante, elle n'est pas suffisante pour rapporter de manière satisfaisante l'opinion véhiculée par un texte, voire par une simple phrase.

Considérons par exemple l'exemple suivant :

- (1) *Ce restaurant n'est pas très bon, et ne vaut pas l'excellente note donnée par les autres commentateurs.*
- (2) *Cette copie n'est pas très brillante.*
- (3) *Cette copie est très brillante.*

En dépit de l'utilisation exclusive de termes positifs – voire très positifs – dans l'exemple (1), la phrase employée est clairement négative. Cela est dû à l'usage d'opérateurs, en particulier ici de négations, changeant la polarité des expressions d'opinions repérées.

Néanmoins, l'effet de la négation ne s'arrête pas à ce simple renversement de polarité. En effet, considérons les exemples (2) et (3). Si l'exemple (3) est très positif (l'adjectif brillant, de base intense, est encore renforcé par l'opérateur très), son équivalent dans la portée d'une négation simple est quant à lui modérément négatif : en effet, un contexte tout à fait cohérent à la phrase (2) pourrait être le suivant : "Cette copie n'est pas très brillante. Elle n'est pas catastrophique, mais je m'attendais à mieux."

### 6.1.1. Objectifs du chapitre

Dans ce chapitre, nous verrons comment intégrer le contexte local de l'expression, afin de déterminer l'opinion au niveau du segment de phrase. Ce niveau intermédiaire, outre son adaptation à l'opinion évoquée au chapitre 2 et chapitre 5, nous permettra de faire le lien entre les différentes opinions unitaires présentes dans le texte et la structure discursive du document (cf. chapitre 7).

Dans un premier temps, nous proposons en section 6.2 une analyse linguistique des effets de chaque type d'opérateurs. Nous nous intéressons ensuite, en section 6.3, au lien entre expression d'opinion et opinion portée par le segment de discours, et plus précisément comment déterminer une valeur d'opinion portée par le segment de discours à partir des expressions d'opinion du segment et des opérateurs venant s'appliquer sur celles-ci. Nous proposons enfin en section 6.4 un modèle géométrique permettant d'appliquer en pratique ces opérations. Afin de valider les hypothèses et notre modèle, nous les confronterons à un jeu de protocoles d'évaluation

psycholinguistiques d'une part, et aux données collectées pendant la campagne CASOAR d'autre part.

Les travaux décrits dans ce chapitre ont été effectués en collaboration avec Farah Benamara, Yannick Mathieu, Vladimir Popescu, et Nicholas Asher, et ont donné lieu à la publication (Benamara et al. 2012) pour la partie analyse linguistique, et à la publication (Chardon, Benamara, Mathieu, et al. 2013) pour le modèle de calcul.

## 6.2. Analyse linguistique des effets des opérateurs

### 6.2.1. Classifications retenues

Comme évoqué au chapitre 2, nous avons choisi les catégories suivantes pour les opérateurs s'appliquant sur l'opinion :

- opérateurs de négation, dont l'action principale est de renverser la polarité d'une expression d'opinion ;
- opérateurs d'intensification, dont l'action est d'altérer l'intensité d'une expression d'opinion ;
- opérateurs de modalités, dont l'action est d'altérer la certitude ou l'intensité d'une expression d'opinion.

Les opérateurs de restriction, abordés également au chapitre 2, ne seront pas traités dans ce chapitre : nous ne disposons que d'un nombre réduit d'instances de ces opérateurs pour en étudier précisément les effets sur l'opinion.

Ces classes d'opérateurs sont néanmoins de trop haut niveau pour capturer avec précision les effets associés de l'opérateur. En effet, considérons les exemples suivants :

**Exemple a** : « Ce restaurant n'est pas bon. »

**Exemple b** : « Ce restaurant n'est jamais bon. »

**Exemple c** : « J'espère que ce restaurant est bon. »

**Exemple d** : « Ce restaurant doit être bon. »

Les phrases a et b comportent toutes deux une négation portant sur la même expression d'opinion « bon ». Néanmoins, la phrase b remonte une opinion plus intense que la phrase a.

De même, les phrases c et d comportent chacune un opérateur de modalité, mais l'effet produit n'est pas le même. Dans la phrase c, bien que l'opinion soit de base positive, l'opérateur modal boulique « espérer » réduit à néant la certitude de l'opinion. Dans la phrase d, le modal « falloir » peut être interprété de deux manières différentes : si l'on considère que celui-ci est épistémique (i.e. « Ce restaurant doit être bon, un collègue m'en a parlé en bien. »), alors le modal altère la certitude, mais sans l'annihiler ; si l'on considère que le modal est déontique (i.e. « Ce restaurant

doit être bon, sinon c'est la faillite assurée. », alors le modal vient plutôt altérer la force de l'expression.

Afin d'obtenir une classification qui rend compte au mieux des nuances d'effet entre les opérateurs, nous proposons la classification suivante, s'appuyant pour la négation sur (Godard 2013), et pour les modalités sur (Larreya 2004) et (Portner 2009) :

- Opérateurs de négation
  - o Opérateur de négation
  - o Quantifieur de négation
  - o Négation lexicale
- Opérateurs d'intensité
  - o Opérateur d'intensification
  - o Opérateur d'affaiblissement d'intensité
- Opérateurs de modalités
  - o Modalités épistémiques
  - o Modalités déontiques
  - o Modalités bouliques

La Figure 6.1 reprend cette classification.

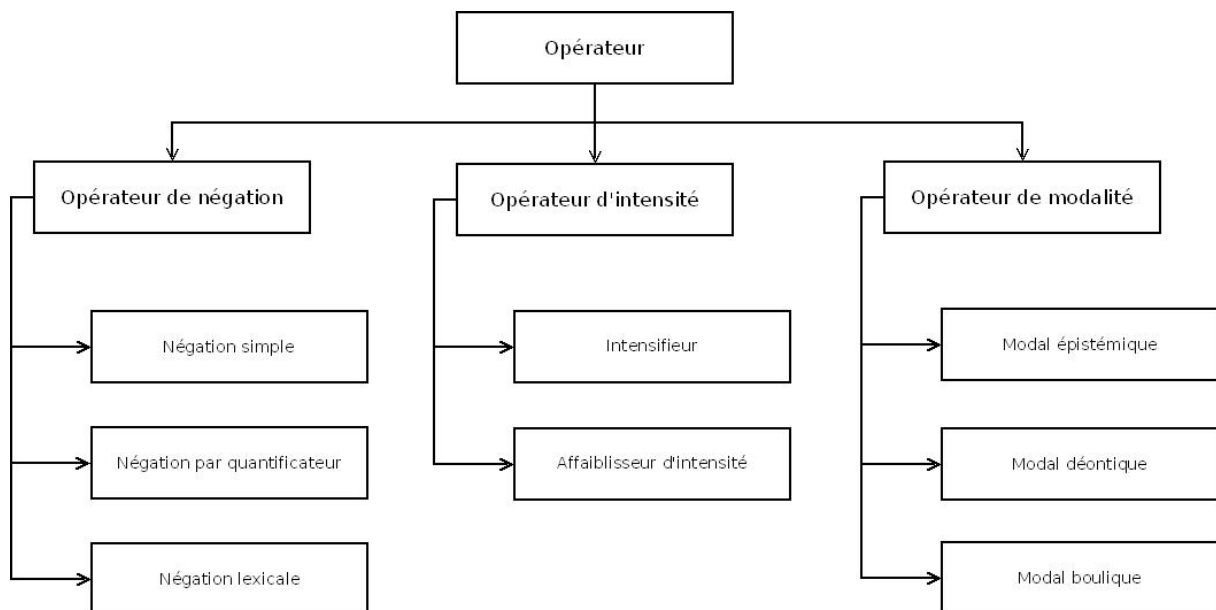


Figure 6.1 : Taxinomie retenue pour les opérateurs s'appliquant sur l'opinion

## 6.2.2. Effets qualitatifs de chaque catégorie d'opérateur

### 6.2.2.1. Opérateurs de négation

Les opérateurs de négation sont utilisés pour rejeter ou nier quelque chose. Grammaticalement, la négation peut être exprimée sous plusieurs formes :

- En utilisant un préfixe : par exemple, "mé-", "mal-" pour le français (*mécontent*, *malheureux*) ; "un-", "il-" pour l'anglais (*unhappy*, *illogical*)
- En utilisant un suffixe : par exemple, "-less" en anglais
- Via un nom ou un verbe comprenant une négation dans sa sémantique. Par exemple, en français, "éliminer".
- À l'aide d'une particule de négation, c'est-à-dire un terme ou une expression pouvant apparaître dans une phrase négative ou une question, mais pas dans son équivalent affirmatif. Par exemple, "ne ... pas", "ne ... plus" ou "ne ... jamais" en français, "any" ou "ever" en anglais.

Les deux premiers cas sont assez spécifiques et peuvent être traités simplement et efficacement via lexicque (i.e. sans explosion de la taille du lexicque). Les opérateurs de négation couvrent uniquement les deux derniers cas, où la négation est extérieure à l'expression d'opinion considérée.

#### *Contexte scientifique*

Les travaux de recherche utilisant la négation en analyse de sentiments peuvent être groupés suivant trois critères principaux :

- L'effet de la négation sur les expressions d'opinion
- Les types de négation pris en compte
- La méthode employée pour mettre à jour la polarité des expressions altérées

Pour le premier critère, la plupart des travaux considèrent que la négation a un effet de renversement de la polarité. Citons notamment (Ding, Liu, and Yu 2008), qui introduisent dans leur chaîne d'extraction des patrons de négation renversant la polarité. Néanmoins, au-delà de cet effet sur la polarité, on peut observer un impact sur l'intensité, comme le montre l'exemple suivant :

- Ce restaurant est excellent.
- Ce restaurant n'est pas excellent.

La première de ces deux phrases véhicule une opinion plus intense que la seconde. Pour tenir compte de ces effets supplémentaires, (J. Liu and Seneff 2009) traitent les négations via un modèle additif linéaire. (Taboada, Voll, and Brooke 2008; Taboada et al. 2011), proposent de modéliser l'opérateur de négation non pas comme un renversement de la polarité, mais comme un décalage d'intensité sur une échelle discrète. Le décalage proposé est d'une valeur de 4 sur une échelle de lexicque allant de -5 à 5, ce qui implique un non renversement de la polarité pour les valeurs extrêmes d'intensité. Néanmoins, la confrontation avec les données expérimentales, issues d'une annotation par Amazon Mechanical Turk, n'a pas permis de confirmer (ni d'infirmer) cette valeur.

Deux types principaux de négations ont été étudiés dans la littérature : les particules de négations (*negators*, tels que *ne...pas* en français ou *not* en anglais) et les termes de contenu incluant une négation (*content word negators*, tel que *éliminer*) (Choi and Cardie 2008). (Wilson, Wiebe, and Hoffmann 2009) distinguent, en plus des particules de négation, des modifieurs de polarité (*polarity shifter*) positifs ou négatifs, qui ne renversent la polarité d'une expression que si celle-ci est d'une polarité donnée.

Enfin, les méthodes employées pour traiter la négation peuvent être elles-mêmes classées en trois catégories (Wiegand et al. 2010) :

- Les approches surfaciques, où la négation est incluse dans un modèle « sac-de-mots » utilisé par la suite par un algorithme d'apprentissage automatique pour la classification suivant la polarité (Pang, Lee, and Vaithyanathan 2002; Ng, Dasgupta, and Arifin 2006) ; ces approches ont l'avantage d'être simple, mais ont tendance à accroître l'espace de traits avec des traits creux.
- Les approches par analyse contextuelle locale, où la négation vient altérer la polarité *a priori* des expressions (Wilson, Wiebe, and Hoffmann 2009; Taboada et al. 2011)
- Les approches utilisant la sémantique compositionnelle, où la polarité des termes constituant la phrase est agrégée (Moilanen and Pulman 2007)

### *Catégories de négation : notre approche*

Nous distinguons trois catégories de négation.

#### *Opérateur de négation (OpNeg)*

Les opérateurs de négations sont les négations simples, ne comportant pas de nuance de quantification.

#### **Exemple :**

"Ce restaurant n'est pas bon."

#### *Quantifieur de négation (OpNegQ)*

Nous appelons quantifieur de négation les particules de négation comportant une nuance de quantification.

#### **Exemple :**

"Ce restaurant n'est jamais bon."

#### *Négation lexicale (OpNegLex)*

Nous appelons négations lexicales les négations induites par des termes autres que des particules de négation, comme le verbe "manquer", ou la locution adverbiale "absence de".

#### **Exemples :**

Ce restaurant manque de bons plats.

L'absence d'une carte appétissante est à remarquer.

### *Propriétés associées*

Pour les opérateurs de négation et les cumuls d'opérateurs de négation, nous émettons les hypothèses suivantes, basées sur l'étude de nos corpus :

#### N1.a

La négation d'une expression d'opinion renverse systématiquement la polarité de cette dernière : une expression positive devient négative, et une expression négative devient positive.

**Exemple :** "Exceptionnel" est positif, "pas exceptionnel" est négatif.

#### N1.b

L'intensité d'une expression du lexique d'opinion dans la portée d'une négation est plus faible ou équivalente à celle de l'expression seule.

Par exemple : "Exceptionnel" est d'intensité forte alors que "pas exceptionnel" est d'intensité plutôt faible : dans le cas d'un film, le premier indique une opinion très positive, alors que le second correspond plutôt à une opinion légèrement négative.

#### N2

L'intensité d'une expression dans la portée d'une négation avec quantification est plus faible que celle de la même expression dans la portée d'une négation simple.

Par exemple : "jamais exceptionnel" est d'intensité plus forte que "pas exceptionnel".

#### N3

Une négation lexicale donne sensiblement le même effet qu'une négation simple.

Par exemple : "Absence de goût" et "Pas de goût".

#### N4

Le cumul d'une négation simple et d'une négation quantifiée est équivalent à l'intensification d'une négation quantifiée.

Par exemple : "Jamais plus exceptionnel." et "Jamais exceptionnel."

### 6.2.2.2. Opérateurs d'intensité

#### *Description*

La caractéristique commune des opérateurs d'intensité est d'altérer l'intensité des expressions d'opinion dans leur portée.

#### Opérateur d'intensification (OpInt+)

Ces opérateurs augmentent l'intensité d'une expression d'opinion. Il s'agit en général d'adverbes.

**Exemple :**

"Ce restaurant est très bon" véhicule une opinion d'intensité plus forte que celle exprimée par son équivalent sans opérateur "Ce restaurant est bon".

### Opérateur d'affaiblissement de l'intensité (OpInt-)

Ces opérateurs diminuent l'intensité d'une expression d'opinion.

#### Exemple :

"Ce restaurant est plutôt bon" est d'intensité moindre par rapport à "Ce restaurant est bon".

### Propriétés

Nous émettons les hypothèses suivantes concernant les opérateurs d'intensité :

#### Int1

Un opérateur d'intensité altère l'intensité d'une expression d'opinion. Cette altération peut être une augmentation ou une diminution, mais n'affecte pas la polarité de l'expression.

#### Int2

Un opérateur d'intensité a toujours un effet sur les expressions d'opinion à portée.

#### Corollaire de ces propriétés

Si l'on considère l'intensité comme une valeur bornée (au moins du côté des intensités faibles), alors l'action d'un opérateur d'intensité sur une opinion donnée dans sa portée varie nécessairement suivant l'intensité de l'opinion en question.

### 6.2.2.3. Opérateurs de modalité

La catégorie des opérateurs de modalité diffère dans sa construction des deux précédentes : elle ne se base pas sur une propriété commune, comme pour les négations et les opérateurs d'intensité, mais sur un ensemble lexical. Suivant le contexte un même terme de cet ensemble peut être interprété comme ayant plusieurs effets différents (cf. exemple 6.3.1.d).

#### Contexte scientifique

Dans le domaine de l'analyse de sentiments, la présence de modalités est parfois utilisée en tant que trait de classification pour une classification d'opinion au niveau de la phrase. Parmi les quelques travaux de recherche allant dans cette direction, nous pouvons citer (Wilson, Wiebe, and Hoffmann 2009), qui utilisent une liste de modaux, et (Kobayakawa et al. 2009) qui distinguent une taxinomie à plat de modalités (requête, recommandation, désir, colonté, jugement, ...). Les résultats avancés ne permettent pas d'estimer clairement le gain pour ce type de classification.

#### Catégories : notre approche

La classification retenue s'appuie partiellement sur (Portner, 2009), ainsi que sur (Larreya, 2004). Pour le Français, nous distinguons les trois catégories suivantes :

#### Opérateur de modalité boulique

Un modal boulique indique les souhaits ou les désirs du locuteur. Ce type de modalité est exprimé via un ensemble restreint de verbes d'espoir (par exemple *espérer*, *souhaiter*).

Par exemple, "Je souhaite que ce restaurant soit bon."

### Opérateur de modalité épistémique

Un modal épistémique indique la croyance du locuteur dans le contenu propositionnel de l'assertion. Ce type de modalité est exprimé via des adverbes de doute, de possibilité ou de nécessité (e.g. "peut-être", "décidément"), ou via les verbes "devoir", "pouvoir" et "falloir".

Par exemple,

"Ce restaurant est peut-être bon."

"Ce restaurant doit être bon [vu l'attente devant la porte]."

### Opérateur de modalité déontique

Un modal déontique indique une possibilité ou une obligation (voire une impossibilité ou une permission). La modalité déontique est exprimée via le même ensemble de verbes que la modalité épistémique : la désambiguïisation s'effectue par le contexte.

Par exemple,

Ce restaurant doit être bon [sinon, le patron va être lynché par la foule].

### *Propriétés associées*

Pour les opérateurs de modalité, nous émettons les hypothèses suivantes :

#### M1

Un opérateur de modalité boulique affaiblit grandement le degré de certitude d'une expression d'opinion dans sa portée.

Par exemple, "J'espère que ce restaurant est bon." porte une certitude très faible.

#### M2

Les modalités épistémiques altèrent le degré de certitude des expressions d'opinion situées dans leur portée.

Par exemple, dans le cas d'une interprétation épistémique, "Ce restaurant est décidément bon" est plus certain que "Ce restaurant est bon". De même, "Ce restaurant doit être bon" est moins certain que "Ce restaurant est bon"

#### M2.a

Pour les adverbes de modalité épistémiques, ce degré est altéré conformément à la certitude de l'adverbe : si celui-ci est incertain, la certitude de l'expression sera diminuée ; si au contraire l'adverbe est certain, la certitude de l'expression d'opinion est augmentée.

Par exemple, "Ce restaurant est décidément bon" est plus certain que "Ce restaurant est bon", qui est plus certain que ce restaurant est probablement bon".

#### M2.b.1

La certitude d'une expression d'opinion dans la portée d'un verbe modal épistémique est plus faible que celle de l'expression seule.



Par exemple, "Ce restaurant doit être bon" possède une certitude plus forte que "Ce restaurant peut être bon", mais ces deux phrases véhiculent une opinion de certitude moindre par rapport à leur équivalent sans modalité épistémique "Ce restaurant est bon".

### M2.b.2

La certitude d'une expression d'opinion dans la portée du modal "devoir" dans son interprétation épistémique est similaire à celle de la même expression dans la portée du modal épistémique "falloir".

Ainsi, dans leur interprétation épistémique, les phrases "Ce restaurant doit être bon [pour qu'il y ait une telle queue dehors]." et "Il faut que ce restaurant soit bon [pour qu'il y ait une telle queue dehors]." Sont de certitude équivalente.

### M3.a

Un opérateur de modalité déontique altère l'intensité d'une expression d'opinion. Cette altération dépend du verbe utilisé : ainsi "pouvoir" réduit l'intensité de l'opinion, alors que "devoir" et "falloir" l'augmentent.

Par exemple, "Ce restaurant doit être bon [pour que les clients ressortent satisfaits demain soir]." est plus intense que "Ce restaurant est bon.", qui est plus intense que "Ce restaurant peut être bon [et les clients ressortiront satisfaits demain soir]."

### M3.b

La force d'une expression d'opinion dans la portée du modal "devoir" dans son interprétation déontique est similaire à celle de la même expression dans la portée du modal "falloir" dans son interprétation déontique.

Par exemple, dans leur interprétation déontique, les phrases "Ce restaurant doit être bon [pour que les clients ressortent satisfaits demain soir]." et "Il faut que ce restaurant soit bon [pour que les clients ressortent satisfaits demain soir]." sont de certitude équivalente.

## 6.2.1. Evaluation des hypothèses linguistiques

### 6.2.1.1. Protocoles de test

Afin d'évaluer nos hypothèses linguistiques, nous avons soumis plusieurs exemples des cas considérés à un panel d'annotateurs volontaires. Ces exemples ont pris la forme de questionnaires web, construits grâce aux plateformes Google Forms<sup>49</sup> et RationalSurvey<sup>50</sup>. Cette seconde plateforme possède l'avantage d'autoriser les questions sous la forme de tableaux à plus de 5 colonnes, au détriment d'une plus grande complexité pour récupérer les données dans la version gratuite.

Les questionnaires ont été soumis à des internautes volontaires, non nécessairement du domaine du traitement du langage. Une question en introduction de chacun des questionnaires demande à l'internaute s'il est de langue maternelle française : en effet, nous voulons capter par ces

<sup>49</sup> Intégré à Google Docs, lui-même intégré à Google Drive : <http://drive.google.com>

<sup>50</sup> <http://www.rationalsurvey.com/>

questionnaires les intuitions de quelqu'un parlant parfaitement la langue et capable d'en saisir de fines subtilités : ce filtre de la langue maternelle nous a semblé une condition suffisante en regard de cet objectif.

Le principe de soumettre des internautes à un questionnaire de ce type n'est pas neuf en traitement du langage : citons notamment les travaux de (Taboada et al. 2011), proches des nôtres, qui ont utilisé la plate-forme Amazon Mechanical Turk<sup>51</sup> pour évaluer les différentes parties de leur modèle d'application des opérateurs pour l'anglais. Nous avons choisi de faire appel à des annotateurs volontaires plutôt que rémunérés pour plusieurs raisons. Tout d'abord, les annotateurs volontaires ne se sentent pas obligés de participer à l'étude, et montrent de l'intérêt pour la tâche : les annotateurs avec lesquels nous avons été en contact nous ont demandé de les tenir au courant des résultats de nos travaux. Ensuite, au contraire d'un annotateur rémunéré à la tâche faisant face à des impératifs de rentabilité (cf. notamment les travaux de (Adda et al. 2011) pour une analyse de ce problème), les annotateurs volontaires peuvent tout à fait prendre leur temps pour répondre aux questions posées. Nous pensons par conséquent que, pour cette tâche, les annotations fournies sont plus fiables que des annotations effectuées par des annotateurs rémunérés à la tâche.

En contrepartie, pour éviter une certaine lassitude ou un manque de temps des annotateurs devant une tâche trop longue, nous avons dû diviser nos questions en cinq protocoles de test distincts.

### *Protocole 1*

Le premier protocole consiste en un ensemble de six questions présentées aux sujets. Dans chaque question, une phrase porteuse d'opinion est présentée, ainsi que son équivalent comportant un opérateur de négation (par exemple : "Cet étudiant est brillant." et "Cet étudiant n'est pas brillant."). Les intensités des expressions d'opinion présentées varient d'une question à une autre sur une échelle discrète (l'échelle de notre lexique, allant de -3 à +3). Il est demandé aux annotateurs d'estimer la force de chaque phrase sur cette échelle. 81 personnes de langue maternelle française ont répondu à cette question.

### *Protocole 2*

Dans ce protocole, nous présentons huit questions aux annotateurs. Chaque question contient une paire de phrases, l'une d'elle comportant une phrase porteuse d'opinion avec une négation simple, l'autre étant son équivalent avec une négation quantifiée ou lexicale. Par exemple : "Cet étudiant n'est pas brillant." et "Cet étudiant n'est jamais brillant." Les annotateurs doivent comparer les forces des deux phrases, et donner la plus intense. 96 personnes de langue maternelle française ont répondu à cette étude.

### *Protocole 3*

Dans ce protocole, cinq questions sont proposées. Dans chacune d'elle, une phrase est présentée aux sujets : parmi les cinq phrases, quatre comportent une modalité boulique, et une ne comporte pas de modalités. Pour chaque phrase, nous demandons aux annotateurs de spécifier si le locuteur

---

<sup>51</sup> <https://www.mturk.com/mturk/welcome>

émet une opinion établie (positive ou négative). Par exemple, "J'ai vu ce film hier. J'espère que ce sera un succès.". 78 annotateurs de langue maternelle française ont répondu à ce sondage.

#### Protocole 4

Nous proposons dans ce protocole huit questions aux annotateurs. Dans chaque question, nous présentons une phrase porteuse d'opinion. La première phrase ne porte pas d'opérateur modal ; chacune des autres phrases contient une modalité épistémique d'un certain degré de certitude, soit certain, soit incertain. Les annotateurs ont ensuite été interrogés si, selon eux, la phrase était plus certaine, moins certaine, ou aussi certaine que son équivalent sans opérateur. 111 annotateurs de langue maternelle française ont répondu à ce questionnaire.

#### Protocole 5

Dans ce protocole, quatre questions sont présentées aux annotateurs. Pour chaque question, nous présentons une paire de phrases porteuses d'opinion : une sans modalité, et une comportant un modal déontique, par exemple : "Allez voir ce film, il est bien." et "Vous devriez aller voir ce film, il est bien." Nous demandons aux annotateurs de comparer la force de chacune des deux phrases de chaque paire. 78 annotateurs de langue maternelle française ont répondu à ce protocole d'évaluation.

##### 6.2.1.2. Négation

La Figure 6.2 montre les résultats obtenus pour l'évaluation des hypothèses concernant la négation (cf. section 6.2.2.1 pour la définition de chaque hypothèse).

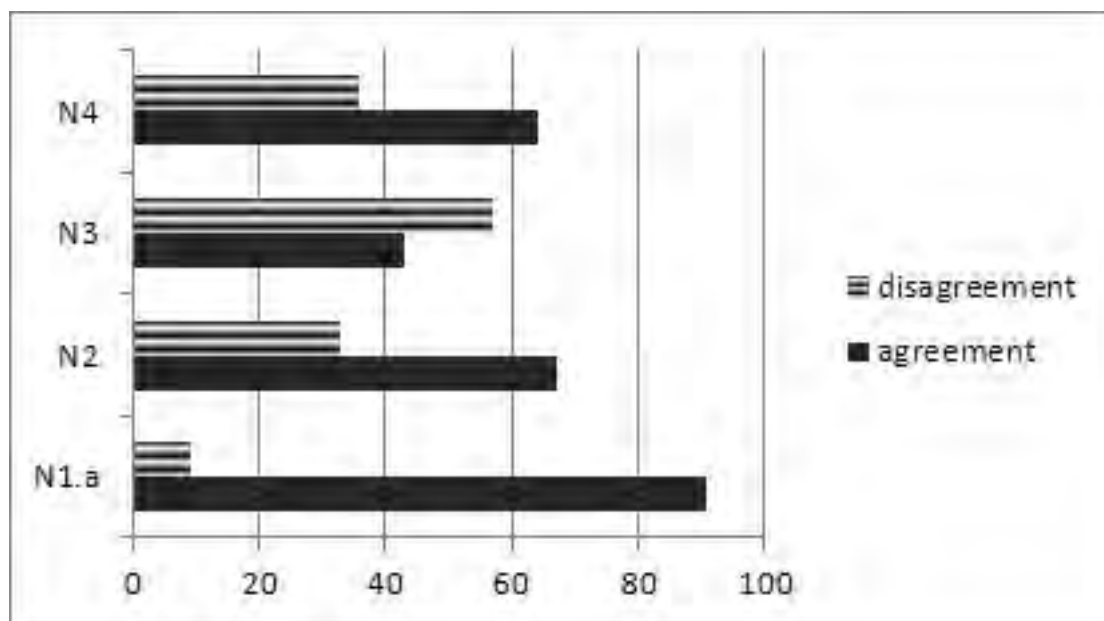


Figure 6.2 : Histogramme des résultats pour la validation empirique des hypothèses linguistiques N1.a à N4

Les quatre hypothèses formulées sont validées. Pour l'hypothèse N1.A, nous obtenons un accord de 90.7 %, en excluant les réponses correspondant au score 0 (20.37 % des réponses). Nous pouvons également remarquer que pour les opinions d'intensité entre -1 et +2, N1.a est vérifiée à 100 %. La même tendance peut être observée pour les intensités -2 et +3 (respectivement 87.8 %

et 93 % d'accord). Néanmoins, pour les opinions extrêmement négative (-3, par exemple : "Cet acteur est nullissime" vs. "Cet acteur n'est pas nullissime.") 48.8 % seulement des sujets renversent la polarité.

Le Tableau 6.1 présente la répartition des résultats pour l'hypothèse N1.b : les lignes correspondent à l'intensité donnée par les annotateurs pour la phrase sans négation, et les colonnes à celle donnée pour la phrase avec négation.

	+3	+2	+1	0	-1	-2	-3
+3	0.0 %	0.0 %	4.7 %	32.9 %	58.9 %	3.5 %	0.0 %
+2	0.0 %	0.0 %	0.0 %	4.9 %	82.0 %	13.1 %	0.0 %
+1	0.0 %	0.0 %	0.0 %	0.0 %	84.3 %	14.5 %	1.2 %
-1	0.0 %	0.0 %	62.5 %	37.5 %	0.0 %	0.0 %	0.0 %
-2	0.0 %	1.2 %	51.9 %	39.5 %	7.4 %	0.0 %	0.0 %
-3	0.0 %	1.4 %	26.4 %	43.0 %	23.6 %	5.6 %	0.0 %

Tableau 6.1 : Répartition des résultats pour l'hypothèse N1.b

Nous pouvons observer que l'hypothèse N1.b est vérifiée pour toutes les combinaisons d'intensité. Nous pouvons également remarquer qu'un pourcentage non négligeable d'annotateurs assigne 0 à la force de la négation de toutes les expressions d'opinion négatives. Ceci est particulièrement vrai pour les opinions extrêmement négatives.

L'hypothèse N2 est vérifiée à 67 %. Ce chiffre modéré vient peut être du fait que l'écart entre la force de la négation quantifiée d'une expression d'opinion et celle de la négation de l'expression est plutôt faible.

L'hypothèse N3 est vérifiée à 43 %. Ce résultat assez faible reflète le fait que les annotateurs ont en particulier considéré que l'opinion dans la portée de la négation lexicale "manque de" est moins négative que son équivalent dans la portée d'une négation simple. Cette disparité dans les résultats montre qu'une étude plus approfondie sur certaines négations lexicales doit être effectuée pour mieux comprendre leur effet sur l'opinion.

Enfin, l'hypothèse N4 concernant les cumuls de négation est vérifiée à près de 64 %. Le désaccord observé vient de la question testant l'effet du cumul de négations quantifiées. Du fait que les négations quantifiées augmentent déjà l'intensité de l'expression d'opinion, il est par conséquent possible qu'ajouter de nouvelles quantifieurs de négation n'apporte pas nécessairement plus d'intensité à l'opinion.

### 6.2.1.3. Modalités

La Figure 6.3 montre les résultats obtenus pour l'évaluation des hypothèses concernant les modalités.

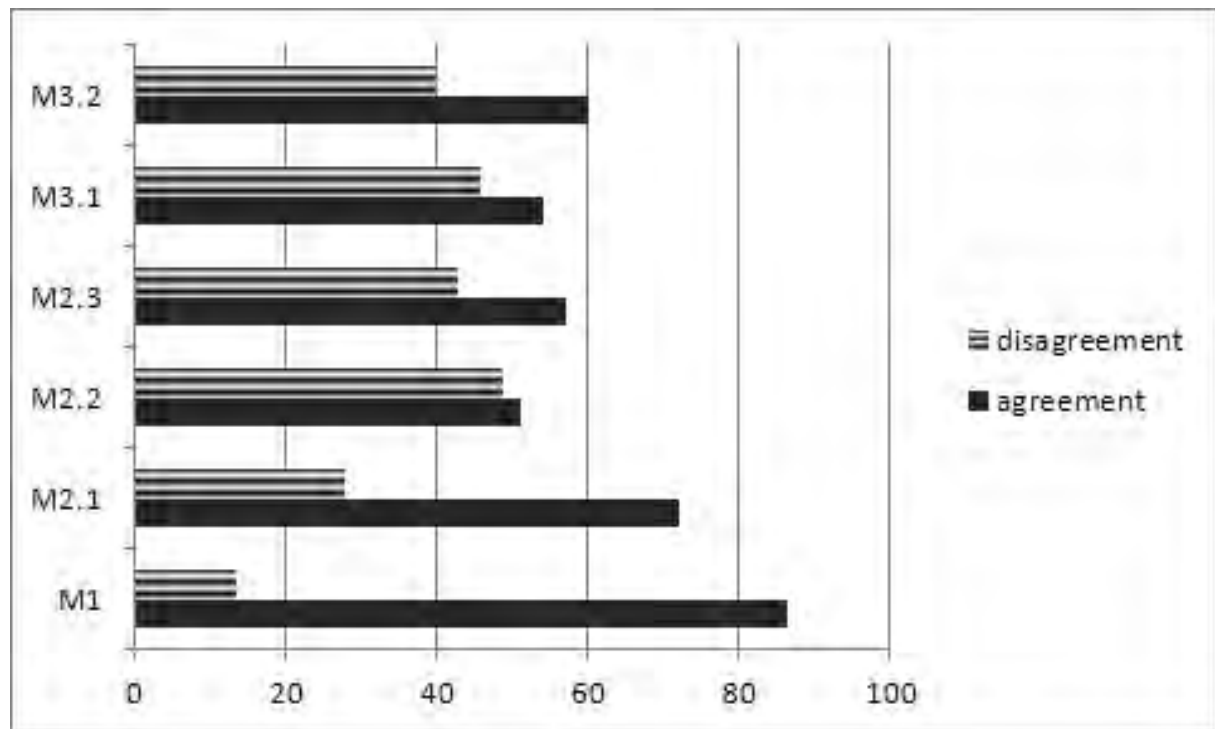


Figure 6.3 : Histogramme des résultats pour la validation empirique des hypothèses linguistiques M1 à M3.2

M1 est validée à 86.5 %. Plus spécifiquement, quand l'expression qui est dans la portée d'un modal boulique dénote un évènement, la totalité des sujets valident le fait que la phrase véhicule une opinion. Ce cas correspond en général à une opinion implicite. De plus, pour tous les sujets, la présence d'une modalité boulique annule l'opinion dans sa portée quand celle-ci dénote un état. Par conséquent, la présence de termes subjectifs dans la phrase "J'espère que ce plat sera bon." ne rend pas la phrase porteuse d'opinion.

L'hypothèse M2.1 est validée à environ 72 % pour chacun des degrés de certitude (*certain* ou *incertain*), ce qui montre que, en plus de l'intensité et de la polarité, la certitude est une caractéristique pertinente pour l'opinion. Concernant M2.2, l'hypothèse est validée par près de 79 % des sujets dans le cas du modal épistémique "pouvoir" moins certain que le modal épistémique "devoir", tandis que seulement 23 % d'entre eux considèrent que "devoir" affaiblit le degré de certitude de l'expression dans sa portée. M2.3 est validée par environ 57 % des annotateurs ; les annotateurs considèrent que les modaux "devoir" et "falloir" dans leur usage épistémique ont le même effet.

Enfin, M3.1 est validée dans une moindre mesure : 54 %. D'autre part, 62.5 % des sujets approuvent l'hypothèse que le déontique "pouvoir" réduit l'intensité de l'opinion dans sa portée. Ceci peut être expliqué par l'ambiguïté entre les interprétations déontique et épistémique de ces verbes. Concernant M3.2, les intensités des déontiques "devoir" et "falloir" sont considérées identiques à 60 %. Les 40 % restant sont équitablement répartis entre les réponses "devoir" plus fort que "falloir" et "falloir" plus fort que "devoir".

### 6.3. De l'expression d'opinion au segment subjectif

Dans cette section, nous allons voir comment passer de l'expression d'opinion, avec ses attributs a priori (polarité, intensité), à une structure d'opinion de granularité plus importante, au niveau du segment de discours subjectif, afin d'établir un lien avec la structure discursive du texte.

Pour cela, nous allons tout d'abord évoquer comment déterminer le lien entre l'expression d'opinion et les opérateurs venant s'appliquer sur elle, c'est-à-dire comment déterminer la portée d'un opérateur. Dans un second temps, nous verrons comment généraliser l'opinion d'une expression en contexte vers l'opinion d'un segment discursif.

#### 6.3.1. Lien entre expression d'opinion et opérateur : portée d'un opérateur

Le lien entre l'opérateur et l'expression d'opinion sur laquelle celui-ci s'applique n'est pas toujours direct. Dans cette section, nous proposons un algorithme simple se basant sur la segmentation discursive, permettant :

- de déterminer, pour tout couple opérateur – expression d'opinion, si l'opérateur impacte l'expression d'opinion
- dans le cas où plusieurs opérateurs impacteraient une même expression d'opinion, de déterminer dans quel ordre ceux-ci viennent s'appliquer

Cet algorithme est ensuite évalué sur les données issues du projet CASOAR (cf. chapitre 2).

##### 6.3.1.1. Association expression – opérateurs

Nous divisons ce problème en deux étapes :

- La résolution de la portée des opérateurs, permettant d'associer à chaque expression d'opinion un ensemble non ordonné  $Op_x$  d'opérateurs
- L'ordonnancement des opérateurs, permettant d'ordonner cet ensemble en une liste  $LOp_x$ , suivant l'ordre d'application de leur effet sur l'opinion

##### *Portée de l'opérateur*

L'idée principale est d'associer un opérateur à une expression d'opinion lorsque ceux-ci se trouvent dans le même segment de discours.

Nous considérons cinq variations de l'algorithme, du plus large au plus restrictif :

- Portee\_1 : un opérateur est associé à toutes les expressions d'opinion présentes dans son segment de discours, et dans tout segment de discours subordonné, directement ou indirectement, à celui-ci
- Portee\_2 : un opérateur est associé à toutes les expressions d'opinion présentes dans son segment de discours, et dans tout segment de discours directement subordonné à celui-ci

- Portee\_3 : un opérateur est associé à toutes les expressions d'opinion présentes dans son segment de discours, et, si ce segment est source d'une ou plusieurs relations d'Attribution, dans les segments cibles de ces relations
- Portee\_4 : un opérateur est associé à toutes les expressions d'opinion présentes dans son segment de discours
- Portee\_5 : un opérateur est associé à toutes les expressions d'opinion présentes dans son segment de discours après lui

### **Ordonnement des opérateurs s'appliquant sur une même expression**

À l'issue de l'étape précédente, on obtient, pour chaque expression d'opinion  $x$  l'ensemble d'opérateurs  $Op_x$  s'appliquant sur l'expression d'opinion. Il reste à déterminer dans quel ordre s'appliquent ces opérateurs, i.e. la liste ordonnée  $LOp_x$ .

Pour cela, nous pouvons définir la relation d'ordre  $>_x$  suivante :

$$\forall (op_1, op_2) \in Op_x,$$

$$op_1 >_x op_2 \text{ si } \begin{cases} \text{profondeur}(op_1, x) > \text{profondeur}(op_2, x) \\ \text{profondeur}(op_1, x) = \text{profondeur}(op_2, x) \text{ et} \\ \quad \text{distance}(op_1, x) > \text{distance}(op_2, x) \\ \text{profondeur}(op_1, x) = \text{profondeur}(op_2, x) \text{ et} \\ \quad \text{distance}(op_1, x) = \text{distance}(op_2, x) \quad \text{et} \\ \quad op_1 \text{ est situé après } x \text{ dans la phrase} \end{cases}$$

L'opérateur  $profondeur(op, x)$  désigne la profondeur dans le graphe de discours partiel (i.e. le nombre de relations minimal permettant de relier le segment contenant l'opérateur  $op$  et celui contenant l'expression d'opinion  $x$ ). L'opérateur  $distance(op, x)$  désigne la distance en nombre de mots entre l'opérateur  $op$  et l'expression d'opinion  $x$ .

La dernière condition est arbitraire : elle permet simplement de résoudre le cas, peu probable en théorie, de deux opérateurs situés de part et d'autre d'une expression d'opinion. Ainsi, la relation  $>_x$  est une relation d'ordre totale.

Les opérateurs de  $LOp_x$  seront appliqués dans l'ordre croissant issu de  $>_x$ .

#### **6.3.1.2. Évaluation de l'algorithme de détermination de portée**

Afin d'évaluer nos algorithmes d'association opérateur-expression d'opinion, nous les avons testés sur notre corpus de commentaires web, annoté lors de la campagne CASOAR (cf. chapitre 1). Ce corpus comporte 489 instances d'opérateurs portant sur une ou plusieurs expressions d'opinion, pour 397 instances d'application d'un ou plusieurs opérateurs sur une expression d'opinion (i.e. 397 liens entre un ou plusieurs opérateurs et une opinion).

L'évaluation a lieu en deux étapes. Dans un premier temps, nous avons évalué l'algorithme de détermination de portée de l'opérateur. Dans un second temps, nous avons testé l'ordonnement des opérateurs associés à l'expression d'opinion.

### Portée de l'opérateur

Le Tableau 6.2 présente, en termes de précision et de rappel, les résultats obtenus pour chacune des cinq variations proposées.

	Précision	Rappel	F-mesure
Portee_1	0.740	<b>0.976</b>	0.842
Portee_2	0.770	0.974	0.860
Portee_3	0.924	0.967	0.945
Portee_4	<b>0.938</b>	0.962	<b>0.950</b>
Portee_5	0.971	0.820	0.889

Tableau 6.2 : Évaluation de l'algorithme de portée de l'opérateur sur le corpus de commentaires web CASOAR

Comme attendu, l'algorithme « Portee\_1 » présente le meilleur rappel, et l'algorithme « Portee\_2 » la meilleure précision. Au niveau du ratio précision/rappel, on s'aperçoit par contre que les algorithmes « Portee\_1 » et « Portee\_2 » dégradent grandement la précision, qui tombe en dessous des 80 %. De même, l'algorithme « Portee\_5 » constitue un filtre trop restrictif : le rappel descend à 82 %. Les algorithmes « Portee\_3 » et « Portee\_4 » présentent quant à eux un équilibre précision – rappel bien meilleur, avec une bonne précision (respectivement 92.4 % et 93.8 %), et un très bon rappel (respectivement 96.7 % et 96.2 %). La F-mesure de l'algorithme « Portee\_4 » est enfin légèrement plus élevée que celle de « Portee\_3 ».

Le choix optimal semble donc être l'algorithme « Portee\_4 », qui présente les avantages suivants :

- Bonne précision (meilleure que Portee\_3) : ceci est particulièrement important dans le contexte d'une chaîne de traitement industrielle, afin d'éviter les sorties aberrantes.
- Très bon rappel
- Grande simplicité de mise-en-place : l'algorithme ne dépend pas de la reconnaissance de relations de discours

### Ordonnement des opérateurs

Sur l'ensemble des 397 instances d'applications d'opérateurs, 368 ont été correctement ordonnées par l'algorithme, soit 92.7 % de bonne reconnaissance.

#### 6.3.1.3. Portée d'un opérateur : conclusion et perspectives

Nous avons ici présenté un algorithme simple de détermination de la portée basé sur la segmentation du texte en unités de discours. Les résultats obtenus sur nos corpus sont plutôt bons : l'algorithme est donc utilisable dans le contexte de notre chaîne de traitement.

De meilleurs résultats sont envisageables, notamment en exploitant finement la structure grammaticale du document. Néanmoins, au vu du faible apport possible pour la tâche spécifique, ces travaux s'inscrivent au-delà du présent document.



### 6.3.2. Rappel : attributs de l'expression d'opinion en contexte, et de l'opinion au niveau du segment

Nous avons vu dans le chapitre 2 que l'on pouvait associer à une expression d'opinion plusieurs attributs *a priori*, c'est-à-dire directement héritées du lexique ayant servi à repérer ces expressions d'opinion. Ces attributs sont une polarité (positive, négative, neutre), et une intensité (faible, standard, forte).

Lorsque l'on prend en compte son contexte local, ces attributs sont modifiés. La polarité et l'intensité peuvent être altérées par des opérateurs, et une notion supplémentaire, la certitude découle directement de la présence (ou l'absence) d'opérateurs spécifiques.

Dans cette section, nous allons voir plus en détail chacun de ces attributs.

#### 6.3.2.1. Attributs de l'expression en contexte

##### *Polarité en contexte*

La polarité en contexte local est fonction de la polarité *a priori* de l'expression, et des opérateurs s'appliquant sur celle-ci. Comme la polarité *a priori*, la polarité en contexte a une valeur qui peut être :

- Positive : « Ce n'est pas mauvais. » (opérateur « ne ... pas » venant renverser la polarité négative *a priori*)
- Négative : « Ce n'est pas bon. » (opérateur « ne ... pas » venant renverser la polarité positive *a priori*)
- Neutre : « C'est intéressant. » (pas d'opérateur, la polarité reste ambiguë si l'on se limite au contexte local)

##### *Intensité en contexte*

L'intensité en contexte locale est fonction de l'intensité *a priori* de l'expression, et des opérateurs venant s'appliquer sur l'expression.

##### *Certitude associée à l'expression*

La certitude associée à une expression d'opinion est directement déduite du ou des opérateurs s'appliquant sur l'expression d'opinion. Cette valeur peut être :

- Faible (possible) : la certitude de l'expression est affaiblie par un opérateur :  
« Ce restaurant est potentiellement bon. »
- Standard : aucun opérateur ne vient altérer la certitude de l'opinion :  
« Ce restaurant est bon. »
- Forte : la certitude est renforcée par un opérateur :  
« Ce restaurant est décidément bon. »

### 6.3.2.2. Attributs de l'opinion au niveau du segment

Au niveau du segment, l'opinion revêt les mêmes attributs qu'au niveau de l'expression en contexte, à savoir :

- une polarité
- une intensité
- une valeur de certitude

### 6.3.3. Généralisation : remonter l'opinion au segment

L'objectif est ici, à partir de l'opinion de l'expression en contexte, de déterminer l'opinion du segment englobant cette opinion.

#### 6.3.3.1. Sujet de l'opinion d'un segment

Comme vu au chapitre 2, toute opinion porte sur un sujet. Celui-ci peut être classé en différentes catégories par rapport au sujet principal du document :

- Sujet principal du document
- Sujet partiel : désigne une partie ou un aspect du sujet principal
- Sujet hors-sujet : désigne un sujet d'opinion différent du sujet principal, et sans rapport direct avec celui-ci

Afin de déterminer le type de sujet d'un segment subjectif, il nous faut distinguer plusieurs cas : le cas d'un segment explicite vs. implicite tout d'abord, et, pour les segments explicites, le cas d'un segment comportant une ou plusieurs expressions d'opinion.

#### *Segment explicite comportant une unique expression d'opinion*

Ce cas est le plus simple à traiter : le sujet de l'opinion sélectionné pour le segment est celui de l'unique expression contenue dans le segment.

#### *Segment explicite comportant plusieurs expressions d'opinion*

Il peut arriver qu'un segment de discours possède plusieurs opinions distinctes. Par construction de la structure du segment de discours CASOAR (cf. chapitre 1), ce cas est plutôt rare. Il doit néanmoins être pris en compte.

Dans ce cas, nous considérons que le sujet de l'opinion attaché au segment est le suivant :

- Sujet principal : si au moins une des expressions portent sur le sujet principal du document
- Sujet partiel : si aucune opinion ne porte sur le sujet principal, et qu'au moins une des expressions est attachée à un sujet partiel
- Hors-sujet : sinon

L'idée derrière cette sélection est de favoriser l'opinion la plus en rapport avec le sujet du document, et donc la plus liée à l'opinion globale sur le sujet du document.

### *Segment implicite*

Le cas d'un segment subjectif implicite est un peu particulier : en effet, un tel segment ne comporte par définition pas d'expression d'opinion à laquelle on puisse rattacher un sujet. Néanmoins et également par définition, un segment implicite présente typiquement des faits, mais dont la présentation et le contexte induisent une opinion. La présentation de ces faits étant le fait du rédacteur, i.e. à l'émetteur principal, nous pouvons lui associer l'opinion induite par ce type de segment.

### *Cas particulier des segments mixtes : explicite et implicite*

Dans le cas d'un segment mixte, ce cas étant assez rare et difficile à reconnaître via nos méthodes de classification (la méthode du chapitre 4 ne repère pas cette classe très minoritaire) nous le traitons comme un segment explicite traditionnel.

#### **6.3.3.2. Attributs de l'opinion**

Nous distinguons là encore plusieurs cas, suivant le type de subjectivité du segment et suivant son contenu en termes d'expressions d'opinion.

### *Segment explicite comportant une unique expression d'opinion*

Dans ce cas, le segment hérite des attributs portés par son unique expression d'opinion.

### *Segment explicite comportant plusieurs expressions d'opinion*

Nous distinguons ici deux sous-cas, suivant que les expressions d'opinion portent sur un même sujet, ou sur deux sujets de types différents.

#### *Les opinions portent sur un même sujet*

Dans ce cas, il est nécessaire d'avoir recours à une méthode de fusion afin de remonter une opinion unique au niveau du segment. La méthode la plus simple est la moyenne des opinions.

#### **Remarque :**

Certaines des heuristiques de calcul appliquées au niveau du discours pourraient également être appliquées ici. Néanmoins, les segments avec plusieurs opinions comportent typiquement, par construction, deux expressions d'opinion, le plus souvent cohérentes entre elles.

#### *Les opinions portent sur plusieurs types de sujet*

Dans ce cas et en cohérence avec les règles sur le type de sujet, nous ne gardons que les opinions portant sur le type de sujet du segment. Il suffit ensuite de se reporter à l'une des sections précédentes, suivant qu'il reste une ou plusieurs expressions portant sur ce sujet.

### *Segment implicite*

Les segments porteurs d'une opinion implicite ne comportent par définition pas d'expression d'opinion, et ne rentrent donc pas dans le cadre de cette étude.

### 6.3.4. Opérations élémentaires sur l'opinion

L'objectif de cette section est de rattacher chacune des catégories d'opérateur à un ensemble ordonné d'opérations à appliquer. Pour cela, nous distinguons l'ensemble d'opérations élémentaires suivant :

- Opération élémentaire de négation
- Opérations élémentaires d'intensification et d'affaiblissement d'intensité
- Opérations élémentaires de renforcement, d'affaiblissement, et d'annulation de certitude

#### *Opération élémentaire de négation (neg)*

Il s'agit ici de l'opération qui consiste à appliquer une négation simple (par exemple, "ne ... pas") sur une expression d'opinion.

#### *Opérations élémentaires d'intensification et d'affaiblissement de l'intensité (int+, int-)*

Ces opérations consistent à augmenter (int+) ou diminuer (int-) l'intensité d'une expression d'opinion.

#### *Opérations élémentaires de renforcement, d'affaiblissement, et d'annulation de la certitude (cert+, cert-, cert0)*

Ces opérations consistent à définir le niveau de certitude, lorsque celui-ci est explicité par le lecteur. L'opération élémentaire d'affaiblissement de la certitude fixe l'opinion sur la valeur de certitude *affaiblie*. L'opération élémentaire de renforcement de la certitude fixe l'opinion sur la valeur de certitude *renforcée*. L'opérateur d'annulation de la certitude fixe quant à lui la certitude à 0, c'est-à-dire que l'opinion est annulée (ce qui était *a priori* une opinion est en fait une supposition sans fondement, ou un simple espoir).

### 6.3.5. De l'opérateur à l'opération élémentaire

Dans cette section, nous verrons comment passer d'une liste ordonnée d'opérateurs, par exemple issue de l'algorithme proposé en section 6.3.1 à une liste ordonnée d'opérations élémentaires.

À l'issue des étapes précédentes, nous avons, pour une expression d'opinion donnée, une liste ordonnée  $L_{\text{opx}}$  d'opérateurs l'impactant. Le principal problème de cette liste est la non compositionnalité des opérateurs la constituant. En effet, en raison des effets spécifiques des cumuls d'opérateurs :

Soit  $x$  une opinion portée par une expression du texte,

Soit  $y$  l'opinion résultant de l'application de l'opérateur  $op1$  sur  $x$ ,

Alors, il est envisageable que  $op2(op1(x)) \neq op2(y)$

Pour pallier à ce problème, une solution est de transformer la liste d'opérateurs en liste équivalente d'opération élémentaires. Pour cela, nous proposons dans le Tableau 6.3 une série de règles de transformation en accord avec les hypothèses formulées précédemment. Les règles sont classées par ordre de priorité dans leur application :

Opérateurs reconnus	Opérations élémentaires associées
Neg_Quant(Neg_Simple) / Neg_Simple(Neg_Quant)	Int+(Int+(Neg))
Neg_Quant(Neg_Quant)	Int+(Int+(Neg))
Neg_Quant(Neg_Lex)	Int+
Neg_Simple(Neg_Lex)	∅
Mod_E+(Mod_E-) / Mod_E-(Mod_E+)	Cert-
Neg_*(Mod_E+) <sup>52</sup>	Cert-
Neg_*(Mod_E-)	Cert+
Mod_B	Cert0 (Opinion annulée <sup>53</sup> )
Mod_E+	Cert+
Mod_E-	Cert-
Mod_D+	Int+
Mod_D-	Int-
Intensif+	Int+
Intensif-	Int-
Neg_Simple	Neg
Neg_Quant	Int+(Neg)
Neg_Lex	Neg

Tableau 6.3 : Règles de passage des opérateurs vers les opérations élémentaires

Les compositions d'opérateurs non spécifiées dans ces règles donnent une composition des opérations élémentaires associées. Par exemple, Neg\_Simple(Intensif+) – *Ce restaurant n'est pas très bon.* – donne la composition d'opérations élémentaires Neg(Int+).

#### 6.4. Modèle parabolique pour le calcul local

Le modèle que nous proposons se base sur la projection des opinions relevées dans le texte sur une parabole contenue dans un plan pour calculer l'opinion au niveau d'un segment de discours.

Soit  $T$  un segment discursif subjectif contenant au moins une expression d'opinion  $exp$  portant sur un sujet  $S_{exp}$ . L'expression d'opinion est caractérisée par un score d'opinion *a priori*  $score_{exp} = pol_{exp} * int_{exp}$ . Ici,  $pol_{exp}$  désigne la polarité *a priori* de l'expression, égale à -1 ou +1, et  $int_{exp}$  est l'intensité *a priori* de l'expression, appartenant à l'ensemble  $\mathbb{Z} \cap [1; 3]$ . Ces deux valeurs sont en général fournies par le module de repérage suivant lexicale, décrit en section 3.2.4. L'expression  $exp$  peut être altérée par un ensemble d'opérations élémentaires  $OP = \{op_i, pour i = 1..n\}$ . L'effet cumulatif de ces opérations sur l'expression  $exp$  est représenté par la formule  $op_1(op_2(...op_n(exp)))$ , où l'ordre des opérateurs reflète leur portée sur  $exp$  :  $op_1$  est ici l'opérateur le plus distant, qui inclut dans sa portée les autres opérateurs sur l'opinion.

L'objectif de cette section est de déterminer la valeur en contexte  $score_T$ , correspondant à l'application des opérateurs  $OP$  sur l'expression  $exp$ . Pour cela, nous proposons un modèle

<sup>52</sup> Neg\_\* désigne l'un des trois types de négation

<sup>53</sup> En pratique, la certitude de l'opinion est réduite à 0.

parabolique où une expression  $exp$  est représentée par un point  $E$  d'une parabole de foyer  $F$  et de sommet  $O$ , tel que  $E \neq O$ <sup>54</sup>. La Figure 6.4 illustre cette parabole.

L'angle  $\theta_E = \widehat{OFE} \in ]0; \pi[$  mesure la valeur d'opinion de l'expression : le signe correspond à la polarité de l'expression, tandis que la valeur absolue de l'angle correspond à l'intensité de l'expression.

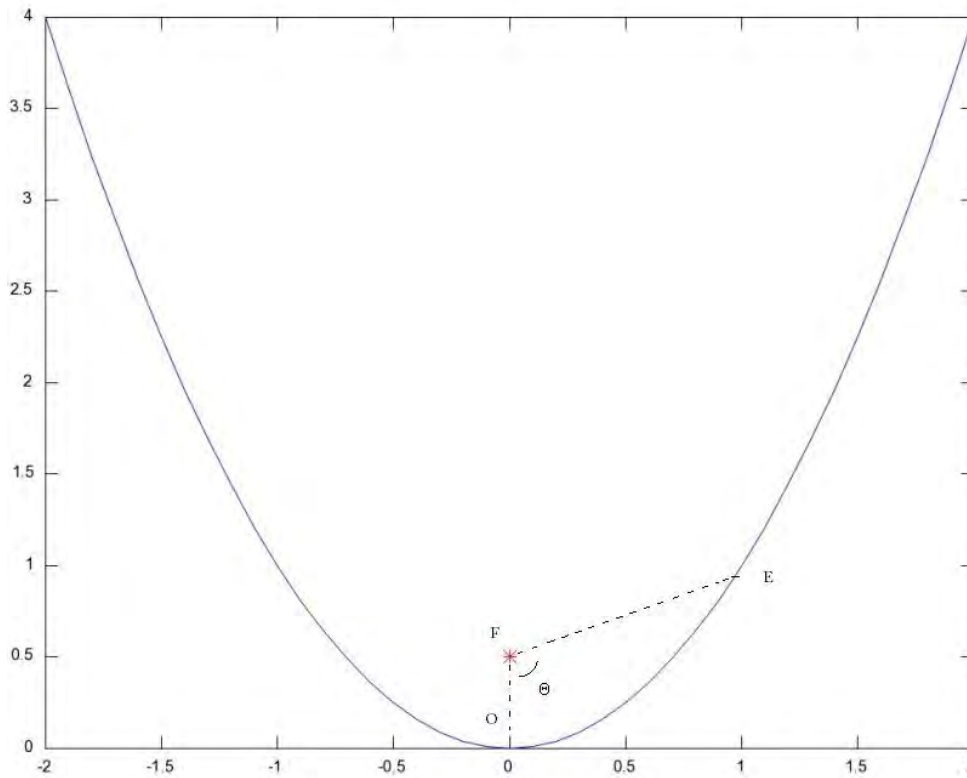


Figure 6.4 : Placement du point  $E$  sur la parabole

La parabole présentée en figure 6.4 est comprise dans un ensemble de trois paraboles de même foyer, mais de pentes différentes (cf. Figure 6.5). Ces pentes correspondent à des degrés de certitude différents. Par convention, nous établissons la valeur de référence  $p_0$ , correspondant à un degré de certitude "standard".  $p_1 > p_0$  désigne alors le degré de certitude "certain", et  $p_2 < p_0$  le degré "incertain". Une expression non modifiée par un opérateur est de certitude standard : le point  $E$  correspondant à cette expression est situé sur la parabole de paramètre  $p_0$ . Une expression modifiée par un opérateur de certitude appartiendra à la parabole  $p_1$  ou  $p_2$ , suivant que l'opérateur accroît ou diminue la certitude accordée à l'opinion.

<sup>54</sup> Le sommet  $O$  de la parabole représente les expressions non porteuses d'opinion, qui ne sont pas couvertes par ce modèle. Par conséquent, le point  $E$  ne peut être situé en  $O$ .

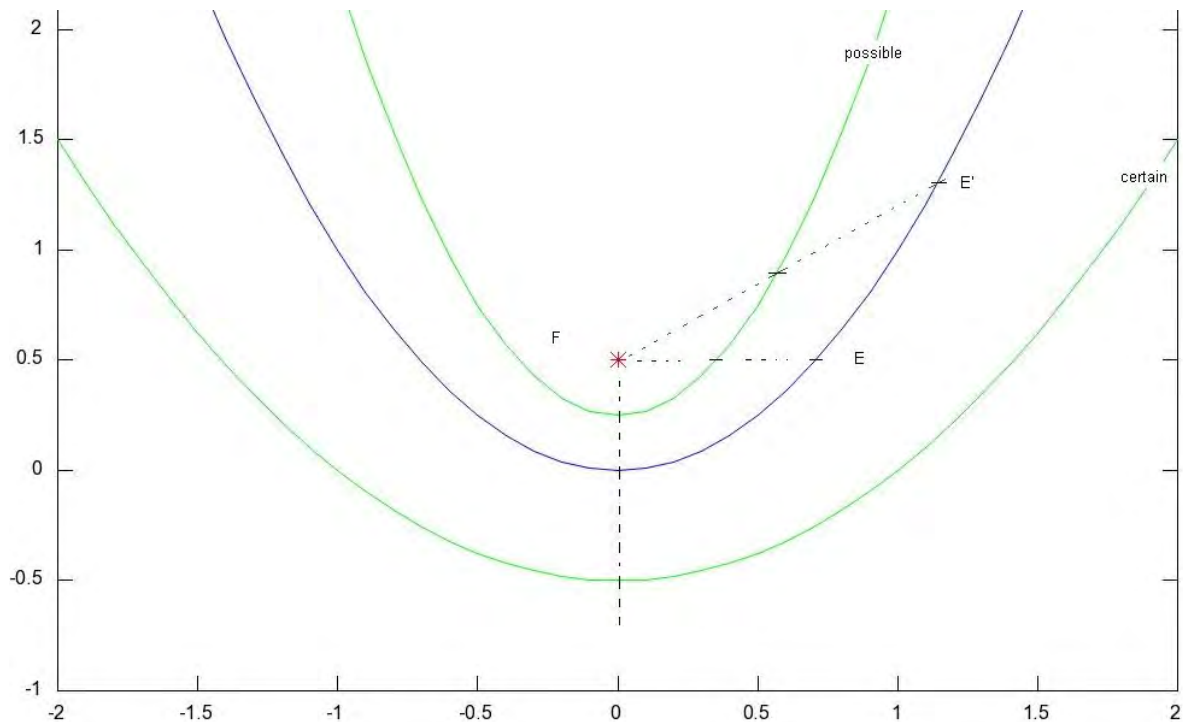


Figure 6.5 : Modèle parabolique étendu

Le modèle que nous proposons possède les caractéristiques suivantes :

- modèle multi-échelle : nous proposons un algorithme d'adaptation à l'échelle d'un lexique donné en section 6.4.1 ;
- modèle prenant en compte la certitude ;
- modèle multilingue : bien que présenté pour le Français, ce modèle est adaptable à une autre langue, moyennant un lexique d'opérateurs adapté.

Dans cette section, nous présentons le modèle en plusieurs étapes. Tout d'abord, nous introduisons l'algorithme de normalisation permettant d'adapter l'échelle du lexique pour les entrées du modèle. Dans un second temps, nous introduisons le modèle parabolique restreint, sans prise en compte de la certitude (i.e. parabole  $p_0$  uniquement). Enfin, nous présentons dans une troisième partie l'extension du modèle parabolique pour la certitude (i.e. modèle présentant les trois paraboles de même foyer).

#### 6.4.1. Adaptation par rapport à un lexique donné

Afin d'adapter le modèle à une échelle donnée, nous proposons un algorithme de mise à l'échelle permettant d'adapter les valeurs des sorties. Pour cela, nous proposons une normalisation de l'échelle du lexique vers une échelle d'intensité comprise entre 0 et  $\pi$ .

Soit un lexique possédant une échelle d'intensité finie  $I = [1; MAX] \cap \mathbb{Z}$ . Soit  $w_0$  un terme du lexique, de polarité  $p_{w_0}$  et d'intensité  $i_{w_0}$ , tel que, dans la portée d'une négation simple, les valeurs de polarité et d'intensité de  $neg(w_0)$  deviennent  $p_{neg(w_0)} = -p_{w_0}$  et  $i_{neg(w_0)} = i_{w_0}$ . Les termes "bon" en français et "good" en anglais sont de bons exemples de telles expressions. Une telle expression  $w_0$  est dite pivot du lexique.

Par convention, nous associons à ce terme pivot l'intensité normalisée  $norm(i_{w_0}) = \pi/2$ . Pour tout autre terme  $w$  du lexique, nous appliquons la formule :

$$\left\{ \begin{array}{l} \text{si } i_w < i_{w_0}, norm(i_w) = \frac{\pi * i_w}{2 * i_{w_0}} \\ \text{si } i_w = i_{w_0}, norm(i_w) = \pi/2 \\ \text{si } i_w > i_{w_0}, norm(i_w) = \frac{\pi * (i_w - 1)}{2 * (MAX + 1 - i_{w_0})} + 1 \end{array} \right.$$

Cette formule permet de conserver l'ordre des valeurs d'intensité, ainsi qu'un espacement régulier des valeurs d'intensité avant et après la valeur pivot.

Le point final  $E$  correspondant à l'expression  $exp$  de polarité  $p_{exp}$  et d'intensité  $i_{exp}$  est donc placé sur la parabole de certitude standard (i.e. de paramètre  $p_0$ ), au point d'angle  $\theta_E = p_{exp} * norm(i_{exp})$ .

### Exemples :

A. Considérons le lexique présenté au chapitre 2. Ce lexique présente une échelle d'intensité sur trois valeurs  $[1; 3] \cap \mathbb{Z}$ . L'intensité du terme pivot "bon" est 1, nous avons donc les valeurs suivantes d'intensité normalisées :

$i_w$	$norm(i_w)$	exemple
1	$\pi/2$	bon
2	$2\pi/3$	brillant
3	$5\pi/6$	excellent

Tableau 6.4 : Intensité normalisée pour le lexique de l'exemple A

B. Considérons maintenant un lexique portant sur une échelle d'intensité sur cinq valeurs  $[1; 5] \cap \mathbb{Z}$ , et de terme pivot "bon" de valeur d'opinion +3. Nous obtenons dans ce cas les valeurs d'intensité normalisées :

$i_w$	$norm(i_w)$
1	$\pi/6$
2	$\pi/3$
3	$\pi/2$
4	$2\pi/3$
5	$5\pi/6$

Tableau 6.5 : Intensité normalisée pour le lexique de l'exemple B

### Remarque : fonction inverse :

Il est possible de revenir aux valeurs correspondant à l'échelle du lexique via la fonction de normalisation inverse  $norm^{-1}$  :



$$\left\{ \begin{array}{l} \text{si } \text{norm}(i_w) < \pi/2, i_w = \frac{2 * i_{w_0} * \text{norm}(i_w)}{\pi} \\ \text{si } \text{norm}(i_w) = \pi/2, i_w = i_{w_0} \\ \text{si } \text{norm}(i_w) > \pi/2, i_w = \frac{2 * (\text{norm}(i_w) - 1) * (\text{MAX} + 1 - i_{w_0})}{\pi} + 1 \end{array} \right.$$

## 6.4.2. Modèle parabolique restreint

### 6.4.2.1. Opération élémentaire de négation

Nous définissons l'opération élémentaire de négation comme étant un décalage de  $\pi$  le long de la parabole, comme l'illustre la figure 6.6.

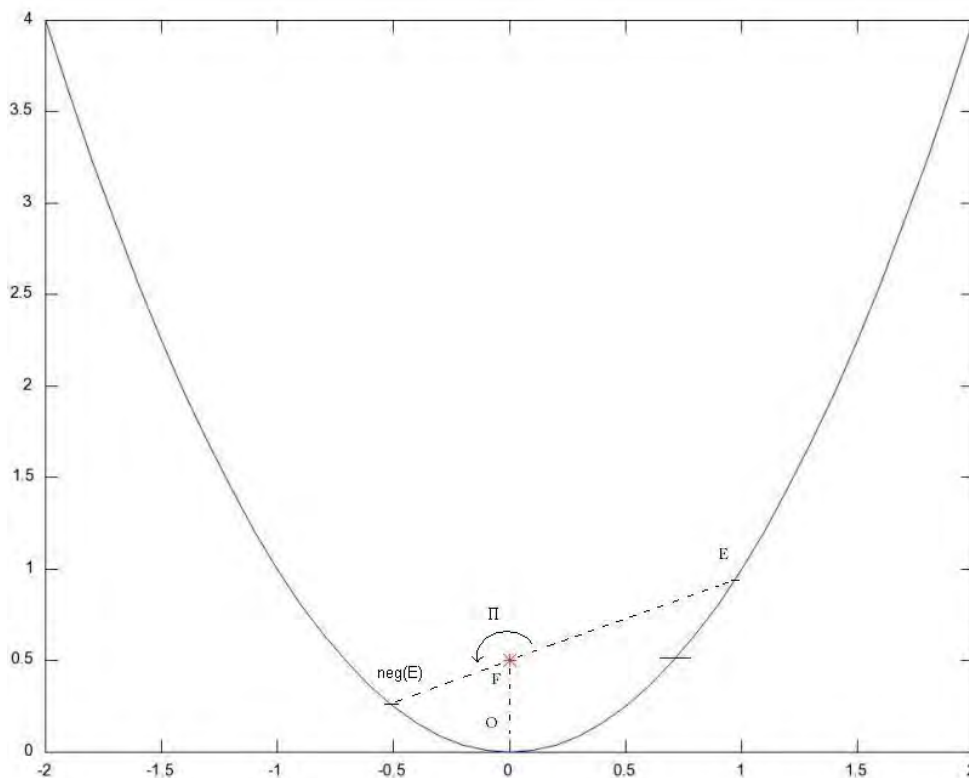


Figure 6.6 : Application de l'opération élémentaire de négation

La formule ainsi appliquée permet de garantir les propriétés observées en section précédente :

$$\text{neg}(\theta) = \begin{cases} (\theta + \pi), & \text{si } \theta < 0 \\ (\theta - \pi), & \text{si } \theta > 0 \end{cases}$$

**Exemples** : plaçons-nous dans le contexte du lexique CASOAR (cf. chapitre 2)

Le terme pivot "bon" est associé à la valeur d'opinion  $\text{score}_{\text{bon}} = +1$ . Suivant le Tableau 6.4, la valeur normalisée associée est  $\text{norm}(\text{score}_{\text{bon}}) = \pi/2$ . Par conséquent, pour la négation simple "pas bon", nous obtenons le score  $\text{norm}(\text{score}_{\text{neg}(\text{bon})}) = -\pi/2$ .

Le terme "excellent" est quant à lui associé à la valeur d'opinion  $score_{excellent} = +3$ . La valeur normalisée associée est donc  $norm(score_{excellent}) = 5\pi/6$ , et, pour la négation simple "pas excellent", nous obtenons le score  $norm(score_{neg(excellent)}) = -\pi/6$ .

La figure 6.7 illustre ces deux exemples.

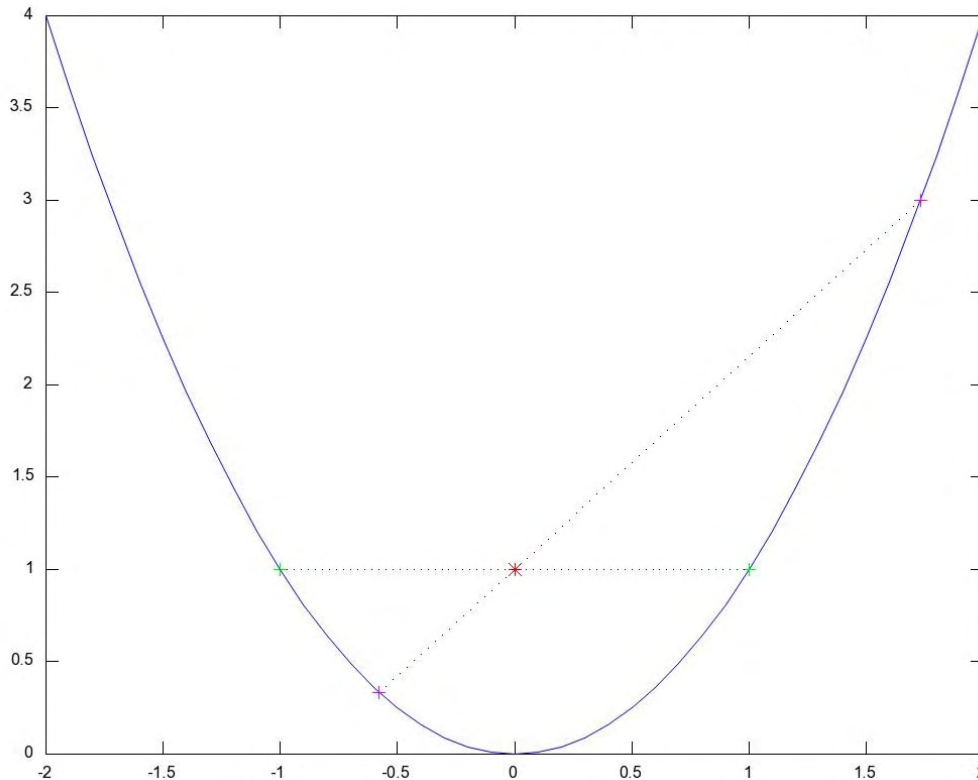


Figure 6.7 : Application de l'opération de négation sur l'exemple

Nous obtenons ainsi  $|norm(score_{neg(excellent)})| < |norm(score_{neg(bon)})|$  : l'intensité de l'expression négative "pas bon" (sur la figure en vert) est plus forte que celle de "pas excellent" (sur la figure en magenta), ce qui est en accord avec nos observations.

#### 6.4.2.2. Opérations élémentaires d'altération de l'intensité

Nous définissons l'opération d'intensification comme étant un décalage sur la parabole courante. Cette opération doit répondre à plusieurs contraintes :

- Cohérence : le résultat ne doit pas être une valeur interdite (0 ou  $\pi$ )
- Conservation de l'orientation sémantique : le résultat de l'opération doit conserver la même polarité que la valeur d'opinion initiale
- Existence sémantique : le résultat de l'opération ne doit pas être identique à la valeur initiale

Afin de préserver ces propriétés, nous proposons les formules suivantes :

$$\begin{cases} \text{int}_+(\theta) = \begin{cases} |\theta|/\theta * 2 * \mu * |\theta| & \text{si } |\theta| \leq \pi/3 \\ |\theta|/\theta * (\pi/2 + \frac{1}{2} * \mu * |\theta|) & \text{si } |\theta| > \pi/3 \end{cases} \\ \text{int}_-(\theta) = \pi - \text{int}_+(\pi - \theta) \end{cases}$$

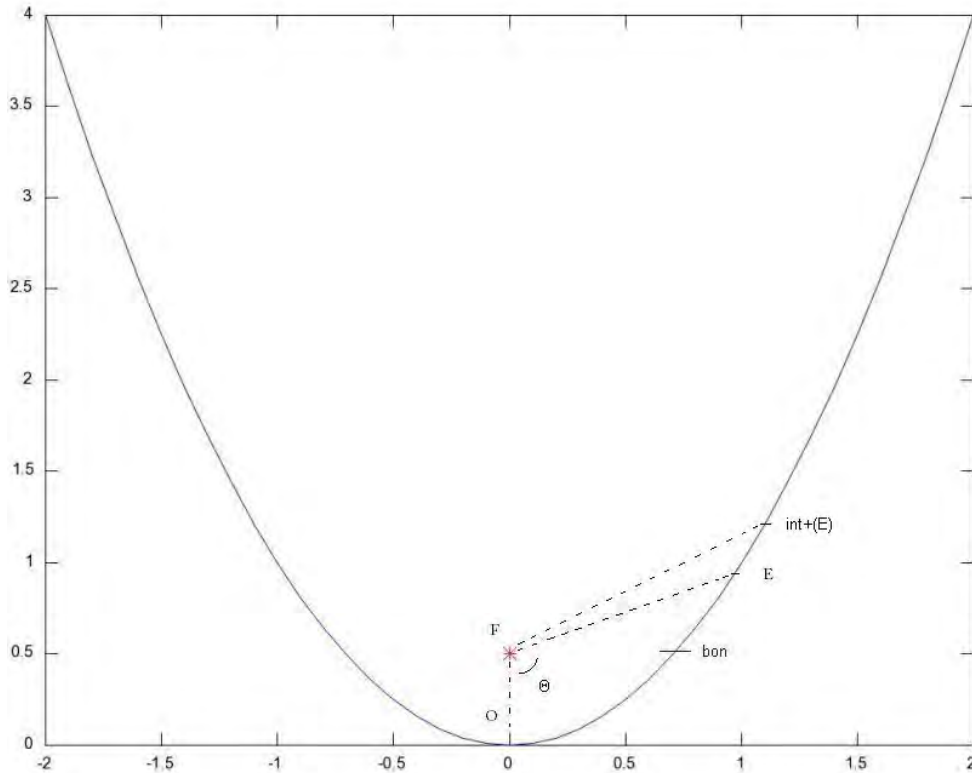


Figure 6.8 : application de l'opération élémentaire d'intensification

Cette formule conserve les propriétés mentionnées précédemment grâce à un effet aux limites atténué : l'intensification d'une valeur d'intensité déjà forte aura moins d'impact que l'intensification d'une valeur d'intensité standard.

**Exemple :** plaçons-nous dans le contexte du lexique CASOAR (cf. chapitre 2)

Le terme pivot "bon" est associé à la valeur d'opinion  $score_{bon} = +1$  et une valeur normalisée associée  $norm(score_{bon}) = \pi/2$ .

Les opérations d'intensification et d'affaiblissement d'intensité donnent respectivement les résultats suivants :

$$\begin{cases} \text{int}_+(norm(score_{bon})) = \text{int}_+\left(\frac{\pi}{2}\right) = \frac{\pi}{2} + \frac{\pi}{4} = \frac{3\pi}{4} \\ \text{int}_-(norm(score_{bon})) = \pi - \text{int}_+(\pi - norm(score_{bon})) = \pi - \text{int}_+\left(\frac{\pi}{2}\right) = \frac{\pi}{4} \end{cases}$$

Le terme "excellent" est quant à lui associé à la valeur d'opinion  $score_{excellent} = +3$ , soit une valeur normalisée associée de  $norm(score_{excellent}) = 5\pi/6$ .

Les opérations d'intensification et d'affaiblissement d'intensité donnent respectivement les résultats suivants :

$$\begin{cases} int_+(norm(score_{excellent})) = int_+\left(\frac{5\pi}{6}\right) = \frac{\pi}{2} + \frac{5\pi}{12} = \frac{11\pi}{12} \\ int_-(norm(score_{excellent})) = \pi - int_+(\pi - norm(score_{excellent})) = \pi - int_+\left(\frac{\pi}{6}\right) = \pi - \frac{\pi}{3} = \frac{2\pi}{3} \end{cases}$$

#### 6.4.2.3. Positionnement des travaux par rapport à l'état de l'art

Concernant l'opération de négation, nos travaux s'inscrivent dans le même esprit que (Taboada et al. 2011) : nous considérons la négation comme un opérateur décalant l'opinion d'une valeur fixe d'intensité plutôt que comme un opérateur renversant la polarité. Nos travaux diffèrent néanmoins sur la valeur de ce décalage. En effet, notre opération élémentaire de négation renverse systématiquement la polarité, ce qui est plus en accord avec les observations relevées en section 6.2. À l'inverse, le décalage proposé par (Taboada et al. 2011) est inférieur à la valeur maximale de leur échelle de lexique, ce qui implique une conservation de la polarité pour les valeurs extrêmes d'intensité. L'opération de négation dans le cadre du modèle parabolique permet ainsi de mieux rendre compte des observations linguistiques énoncées en section 6.2, et constitue une opération graphique simple de rotation correspondant bien à l'idée intuitive de renversement correspondant à la négation.

Concernant l'opération d'intensification (respectivement d'affaiblissement d'intensité), l'opération que nous proposons satisfait bien les propriétés attendues d'un opérateur de renforcement (resp. d'affaiblissement) de l'intensité énoncées en section 6.2. L'atténuation de l'effet aux limites permet ainsi, outre la cohérence mathématique de l'opérateur, de rendre compte de l'effet des cumuls d'opérateurs d'intensité tels qu'énoncés précédemment.

Enfin, le modèle parabolique permet d'introduire une représentation visuelle de l'opinion intéressante dans une perspective industrielle.

#### 6.4.3. Modèle parabolique avec certitude

Comme nous avons pu le voir au chapitre 1 et en section précédente, certains opérateurs, en particulier des opérateurs modaux, peuvent altérer le niveau de certitude d'une opinion. Afin de rendre compte de cette altération, nous proposons une extension au modèle parabolique restreint vu en section précédente.

La parabole considérée en section précédente est caractérisée par un paramètre  $p$ , facteur de la distance entre le sommet de la parabole et le foyer  $F$ . En fixant le foyer  $F$  et en faisant varier le paramètre  $p$ , nous obtenons une série de paraboles imbriquées, de foyer commun  $F$ . Nous proposons de représenter le niveau de certitude de l'opinion par cette distance entre le foyer de la parabole et le foyer (i.e. par le paramètre  $p$ ).

Notre modèle comporte trois degrés de certitude correspondant à trois paraboles de foyer commun  $F$ . Le degré *standard* ( $p_{std} = 1$ ) correspond aux expressions d'opinion dont la certitude n'est pas altérée ; les opérations vues précédemment s'appliquent sur cette parabole. Le degré

*renforcé* ( $p_{renf} = 0.5$ ) correspond aux expressions d'opinion dont la certitude est renforcée par un opérateur ; la parabole correspondante est plus basse et plus large que celle du degré standard. Enfin, le degré de certitude *affaibli* ( $p_{aff} = 2$ ) correspond aux expressions dont la certitude est affaiblie par un opérateur ; la parabole associée est plus haute et plus resserrée que celle correspondant à la certitude standard. La Figure 6.9 illustre ces trois paraboles. Les traits en pointillés illustrent l'équivalent en certitude *affaibli* des projections d'opinion de certitude standard  $E$  et  $E'$ .

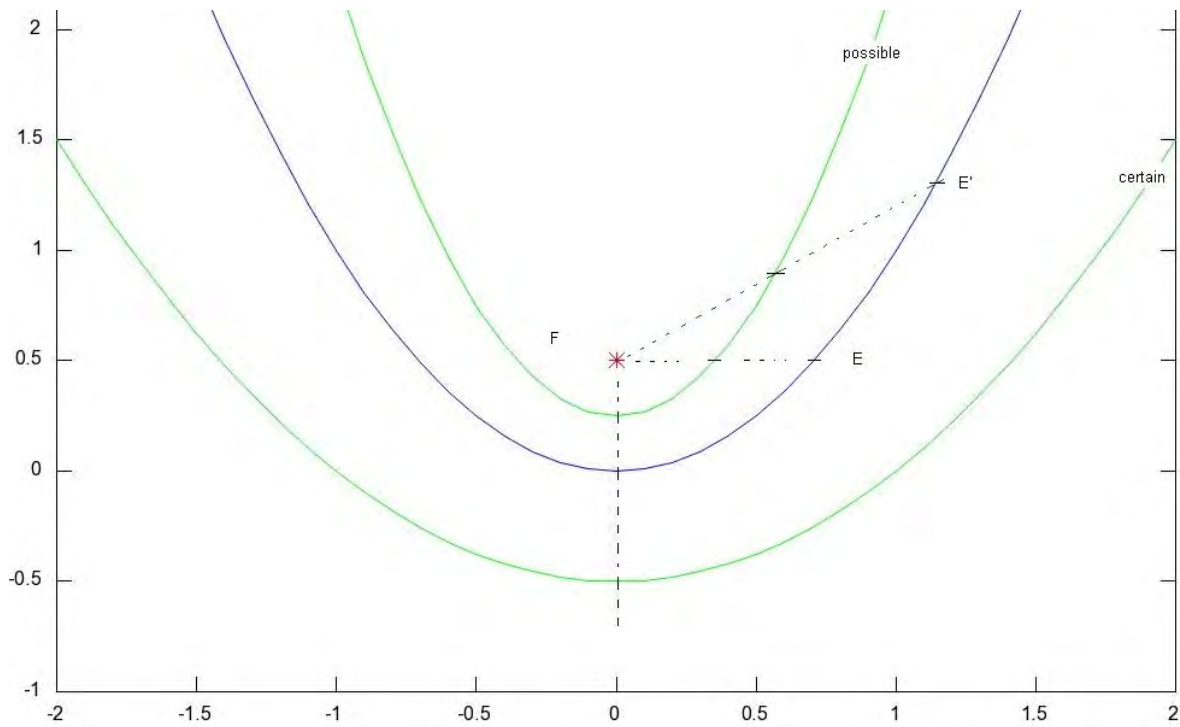


Figure 6.9 : Les trois paraboles du modèle parabolique avec certitude

#### 6.4.3.1. Traitement des modaux "de certitude" (épistémiques)

L'opération élémentaire de renforcement (Cert+) ou d'affaiblissement (Cert-) de certitude consiste en un saut de certitude. Graphiquement, cette opération est appliquée en changeant la parabole sur laquelle l'opinion est basée.

Considérons par exemple les deux phrases suivantes :

- (1) Ce restaurant est bon.
- (2) Ce restaurant est définitivement bon.

L'opinion de la phrase (1) est projetée sur la Figure 6.10 au point  $E$ . L'opinion de certitude renforcée de la phrase (2) est portée par le point  $E'$ , qui partage avec  $E$  le même angle  $\theta_E = \theta_{E'} = \theta$ .

La figure suivante illustre cet exemple :

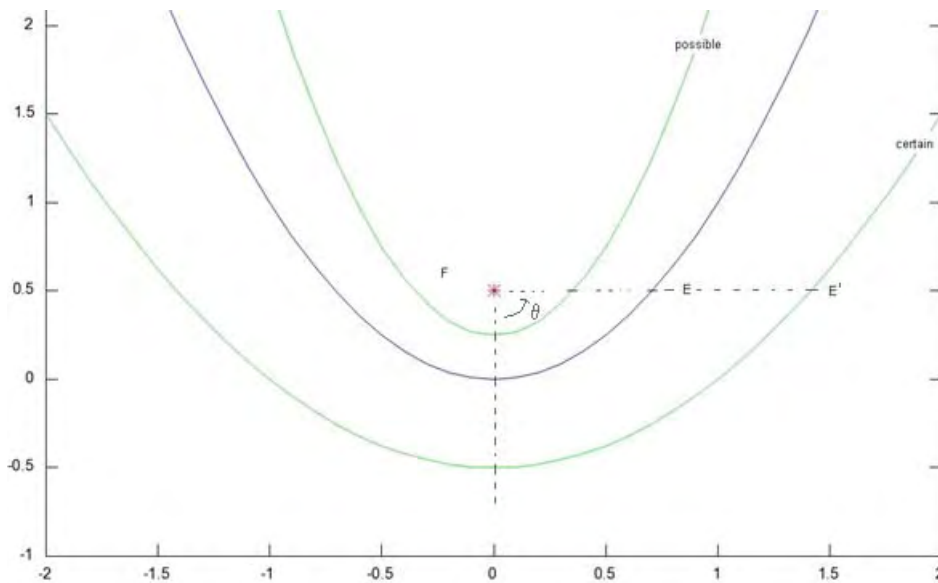


Figure 6.10 : Exemple d'application d'un opérateur renforçant la certitude

La formule d'application de l'opération est la suivante :

$$\begin{cases} \text{cert}_+(\theta, p) = (\theta, p_{\text{renf}}) \\ \text{cert}_-(\theta, p) = (\theta, p_{\text{aff}}) \end{cases}$$

#### 6.4.3.2. Extension négation pour la certitude

Dans le cadre d'une expression  $e$  dont la certitude n'est pas standard, la certitude de l'opinion est également altérée par l'opération, comme nous avons pu le voir en section précédente. Nous modélisons cette opération comme un renversement du paramètre  $p$ .

La formule d'application de l'opération est la suivante :

$$\text{neg}(\theta, p) = \begin{cases} (\theta + \pi, p), \text{ si } \theta < 0 \\ (\theta - \pi, p), \text{ si } \theta > 0 \end{cases}$$

Considérons par exemple les deux phrases suivantes :

- (1) Je suis sûr que ce restaurant est bon.
- (2) Je ne suis pas sûr que ce restaurant soit bon.

L'opinion de la phrase (1) est projetée sur la Figure 6.11 suivante au point  $E$ . L'opinion de certitude renforcée de la phrase (2) est portée par le point  $E'$ , avec  $\theta_{E'} = -\theta_E$  et  $p_{E'} = 1/p_E$ .

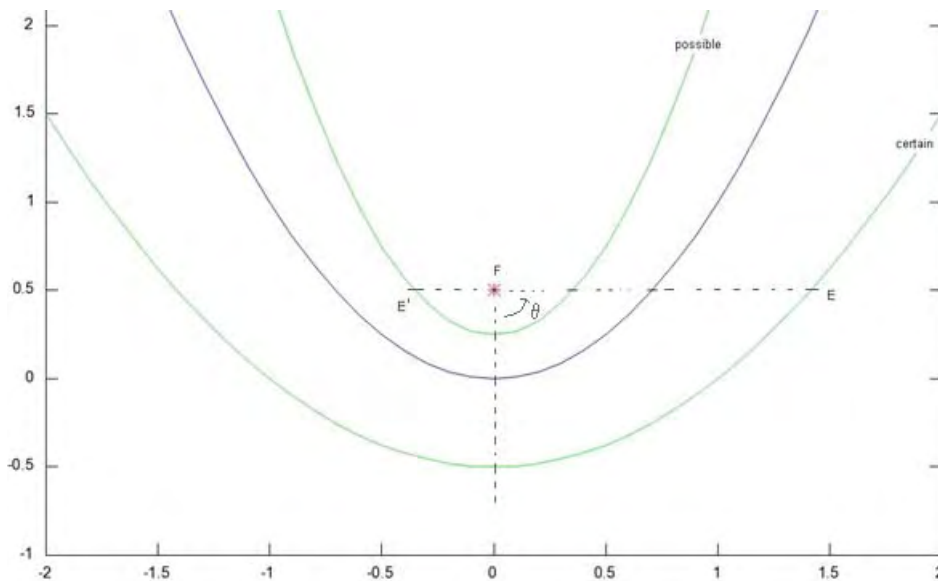


Figure 6.11 : Exemple d'application d'une opération de négation sur une expression de certitude altérée

#### 6.4.4. Évaluation du modèle de calcul

Dans cette section, nous confrontons notre modèle d'une part aux données obtenues lors de nos protocoles de tests, et d'autre part aux données annotées dans le cadre du projet CASOAR (cf. chapitre 1). Nous comparons également notre modèle à d'autres modèles concurrents de détermination de l'intensité et de la polarité d'autre part.

##### 6.4.4.1. Modèle parabolique simple

Nous évaluons tout d'abord la partie négation et intensifieurs du modèle parabolique.

Pour cela, nous comparons les résultats obtenus avec deux autres modèles :

- un modèle de baseline, implémentant un simple renversement de polarité, et un décalage d'intensité unitaire pour les intensifications (modèle *Baseline*)
- le modèle proposé par (Taboada et al. 2011), implémentant un décalage d'intensité pour la négation, et un facteur multiplicatif pour l'intensification (modèle *Taboada et al.*). Le paramétrage de ce modèle est présenté ci-après.

##### *Paramétrage du modèle Taboada et al.*

Le modèle de (Taboada et al. 2011) se base sur une échelle différente de notre annotation, sur 11 points, de -5 à +5. Afin de pouvoir comparer les résultats obtenus pour cette tâche, nous avons effectué une conversion linéaire d'échelle, présentée en extension dans les tableaux suivants.

Valeur échelle [-5;+5]	Valeur dans $\mathbb{R}$ équivalente	Valeur arrondie : échelle [-3;+3]
-5	-3	-3
-4	-12/5	-2
-3	-9/5	-2
-2	-6/5	-1
-1	-3/5	-1
0	0	0
+1	+3/5	+1
+2	+6/5	+1
+3	+9/5	+2
+4	+12/5	+2
+5	+3	+3

Tableau 6.6 : Tableau de passage de l'échelle  $[-5; +5] \cap \mathbb{Z}$  vers l'échelle des données  $[-3; +3] \cap \mathbb{Z}$

Valeur échelle [-3;+3]	Valeur dans $\mathbb{R}$ équivalente	Valeur arrondie : échelle [-5;+5]
-3	-5	-5
-2	-10/3	-3
-1	-5/3	-2
0	0	0
+1	+5/3	+2
+2	+10/3	+3
+3	+5	+5

Tableau 6.7 : Tableau de passage de l'échelle  $[-3; +3] \cap \mathbb{Z}$  vers l'échelle des données  $[-5; +5] \cap \mathbb{Z}$

Le modèle *Taboada et al.* comprend également des paramètres de force des intensifieurs, compris entre 0.5 et 0.9 pour un opérateur affaiblissant l'intensité, et entre 1.1 et 2.0 pour un opérateur augmentant l'intensité. Notre lexique ne distinguant pas cette nuance, nous avons lancé l'évaluation sur l'ensemble des combinaisons de valeurs possibles, en prenant un pas de 0.1 entre chaque valeur. Le Tableau 6.8 expose les résultats de ce test de paramètres. Les lignes représentent les différentes valeurs du facteur multiplicatif pour une opération Int-, et les colonnes les différentes valeurs du facteur multiplicatif pour une opération Int+.

	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9
0.5	0.365	0.365	0.390	0.390	0.564	0.647	0.647	0.647	0.647
0.6	0.348	0.348	0.373	0.373	0.547	0.607	0.607	0.607	0.607
0.7	0.348	0.348	0.373	0.373	0.547	0.607	0.607	0.607	0.607
0.8	0.343	0.343	0.368	0.368	0.542	0.602	0.602	0.602	0.602
0.9	0.335	0.335	0.360	0.360	0.534	0.594	0.594	0.594	0.594

Tableau 6.8 : Résultats (*f*-mesure) du modèle *Taboada et al.* sur nos données pour les différents paramètres d'intensification

Les résultats obtenus sont meilleurs pour les valeurs fortes (respectivement faibles) du facteur multiplicatif pour Int+ (resp. pour Int-), ce qui correspond à des opérateurs ayant un effet important sur l'intensité de l'opinion. Les meilleurs résultats sont obtenus pour un facteur



multiplicatif Int- de 0.5, et un facteur multiplicatif pour Int+ entre 1.6 et 1.9. Arbitrairement, nous choisissons parmi ces valeurs possibles le couple (0.5, 1.7). Le tableau précédent présente de nombreuses cases de valeurs égales. Ceci est dû à un effet de seuil : les nuances de paramétrage peuvent être gommées par la mise à l'échelle des données.

### *Validation du modèle par protocoles linguistiques*

Afin de valider notre modèle, nous avons tout d'abord utilisé les protocoles 1 et 2 présentés en section 6.2.

Pour mémoire, le premier protocole consiste en deux parties. Tout d'abord, un ensemble de six questions est présenté aux sujets. Dans chaque question, une phrase porteuse d'opinion est présentée, ainsi que son équivalent comportant un opérateur de négation (par exemple : "Cet étudiant est brillant." et "Cet étudiant n'est pas brillant." Il est demandé aux annotateurs d'estimer la force de chaque phrase sur une échelle allant de -3 (très négatif) à +3 (très positif). Dans le second protocole, nous présentons huit questions aux annotateurs. Chaque question contient une paire de phrases, l'une d'elle comportant une phrase porteuse d'opinion avec une négation simple, l'autre étant son équivalent avec une négation quantifiée ou lexicale. Par exemple : "Cet étudiant n'est pas brillant." et "Cet étudiant n'est jamais brillant." Les annotateurs doivent comparer les forces des deux phrases, et donner la plus intense.

L'évaluation face aux données du protocole 1 nous permet de mesurer à quel degré les modèles testés reflètent l'intuition des annotateurs, tandis que celle du protocole 2 nous permet de mesurer l'accord entre l'ordre d'intensité donné par les annotateurs et celui donné par le modèle.

Le Tableau 6.9 présente l'accord entre les réponses données par les annotateurs et les résultats des modèles.

<b>Modèle</b>	<b>Accord – protocole 1</b>	<b>Accord – protocole 2</b>
Baseline	27.03 %	5.80 %
Taboada et al.	47.77 %	73.04 %
Modèle parabolique	61.43 %	73.04 %

*Tableau 6.9 : Résultats des différents modèles sur les annotations CASOAR*

Nous pouvons tout d'abord remarquer que les modèles 'Taboada et al.' et 'paraoblique' se comportent tous deux bien face à l'évaluation du protocole 2 : ils permettent de prédire correctement une bonne part des données. À l'inverse, le modèle baseline ne parvient presque jamais à prédire l'ordre donné par les annotateurs. Concernant l'évaluation sur le protocole 1, nous pouvons constater que l'accord maximal est obtenu pour le modèle parabolique. Le modèle baseline obtient là encore un résultat faible. Enfin, le modèle 'Taboada et al.' obtient un score intermédiaire, bien meilleur que la baseline, mais inférieur à celui de notre modèle. La principale explication à cette différence réside dans le fait que le modèle 'Taboada et al.' ne renverse pas systématiquement la polarité.

### Validation du modèle sur les données CASOAR

Nous évaluons maintenant notre modèle sur les données du corpus de commentaires web CASOAR. Le Tableau 6.10 présente les résultats des trois modèles, après paramétrage optimal du modèle *Taboada et al.*

Modèle	Accord avec les annotations
Baseline	59.5 %
Taboada et al.	64.7 %
Modèle parabolique	68.8 %

Tableau 6.10 : Résultats des différents modèles sur les annotations CASOAR

Nous pouvons observer que les meilleurs résultats sont obtenus par le modèle parabolique. La moins bonne performance du modèle *Taboada et al.* sur cette tâche s'explique par plusieurs facteurs :

- L'adaptation à l'échelle des données : la mise à l'échelle du modèle vers les données de la tâche engendre nécessairement une perte de finesse
- Le non respect de l'hypothèse N1 (renversement systématique de la polarité par la négation) : les données CASOAR vérifient en grande majorité cette hypothèse. Or, pour les valeurs d'intensité extrêmes, le modèle *Taboada et al.* n'assure pas ce renversement.
- Le passage à une intensification constante, qui peut avoir perdu certaines nuances entre différents opérateurs d'intensité

#### 6.4.4.2. Modèle parabolique étendu

Nous évaluons maintenant la partie reconnaissance de certitude du modèle. Ceci étant une spécificité de nos travaux, cette tâche n'est pas traitée par les modèles de la littérature.

Afin de procéder à cette évaluation, nous avons vérifié, en termes de précision et de rappel, la reconnaissance des segments identifiés dans le corpus comme étant de certitude renforcée ou affaiblie. Le Tableau 6.11 présente les résultats obtenus :

	Certitude renforcée	Certitude affaiblie	Moyenne
Précision	92.3 %	85.3 %	88.8 %
Rappel	89.3 %	87.8 %	88.5 %
F-Mesure	90.4 %	86.5 %	88.6 %

Tableau 6.11 : résultats de l'évaluation de la prise en compte des opérateurs de certitude

Ces résultats montrent une reconnaissance adéquate des valeurs de certitude au niveau du segment par le modèle. Néanmoins, ceux-ci doivent être relativisés par le nombre restreint d'instance de modalité de certitude s'appliquant sur l'opinion dans les corpus.

Concernant les cas d'erreurs, nous avons trouvés deux sources d'imprécision dans la valeur retournée :

- Certaines modalités viennent s'appliquer directement sur le sujet de l'opinion, mais pas sur les adjectifs et adverbess – expression d'opinion et opérateur – directement associés à celle-ci. Par exemple, dans la phrase "On doute du message très discutabile.", le verbe, opérateur de doute, ne vient pas s'appliquer sur toutes les caractéristiques du message.
- La non prise en compte de la relation de correction : le segment correcteur peut venir invalider un premier opérateur de certitude et le remplacer par un autre. Ainsi, dans l'exemple : "Ce restaurant est peut-être bon. Non, il l'est à coup sûr.", l'opérateur d'affaiblissement de certitude "peut-être" s'applique bien sur l'expression bon, mais son effet est invalidé.

## 6.5. Conclusion

Nous avons présenté dans ce chapitre une étude linguistique sur les effets sémantiques fins des différents opérateurs portant sur l'opinion d'une part, et un modèle computationnel d'application de ces opérateurs d'autre part.

L'étude linguistique a permis de mettre en évidence les variations de comportements entre les différentes catégories de négation (simple, lexicale, quantifiée), et les différentes interprétations des opérateurs de modalité (épistémiques, déontiques, bouliques). Elle a également permis de mettre en lumière l'effet de plusieurs cumuls d'opérateurs, en particulier les cumuls de négation, et de modaux de certitude. Enfin, cette étude a également permis de souligner la pertinence de la certitude comme caractéristique de l'opinion en contexte.

Le modèle géométrique présenté prend en compte la certitude de l'opinion, ainsi que la plupart des effets fins relevés lors de l'étude linguistique. Il permet également de définir une distance entre deux opinions, éventuellement exploitable dans le contexte d'une application industrielle : ainsi, un moteur de recherche dans des commentaires web pourrait, en se basant sur cette distance, proposer des suggestions issues d'avis positifs de personnes dont les opinions renseignées sur le site sont proches de celle de l'utilisateur. Ce modèle se comporte bien vis-à-vis de nos données annotées, et permet une reconnaissance précise de l'opinion émise.

Ces travaux peuvent être approfondis suivant deux pistes : tout d'abord, une étude complémentaire de certains cumuls, en particulier concernant les intensités, permettrait d'affiner encore les hypothèses et par conséquent le modèle. Ensuite, l'annotation de données supplémentaires concernant les opérateurs de modalité, plus rares que les négations et les intensifieurs, permettrait également d'étudier plus en profondeur le comportement de ces opérateurs complexes.



# 7. Calcul de l'opinion au niveau d'un document

---

## 7.1. Introduction

Dans les chapitres précédents, nous avons abordé les différentes étapes permettant de remonter au niveau du segment élémentaire de discours les informations relatives à l'opinion, en particulier une valeur de subjectivité (via classification – cf. chapitre 5), une polarité, et une intensité (via règles d'application d'opérateurs – cf. chapitre 6).

Ces informations peuvent avoir un intérêt à être affichées telles qu'elles, notamment à des fins illustratives. Néanmoins, l'objectif principal de notre chaîne de traitement est de calculer une opinion globale sur le sujet principal du document. Pour cela nous proposons dans ce chapitre une étude des méthodes permettant d'effectuer cette tâche. Ces travaux ont fait l'objet de la publication (Chardon, Benamara, Popescu, et al. 2013).

### 7.1.1. Objectifs et principe de fusion : définition de la tâche

Le principe général de la fusion est le suivant : à partir d'un ensemble d'opinions unitaires, il s'agit de remonter une information globale, résumé de l'opinion de tout le document sur le sujet principal (*main topic*). Dans notre chaîne de traitement, le niveau de l'opinion unitaire choisi est celui du segment discursif.

L'opinion, au niveau du segment discursif, comporte théoriquement les attributs suivants, obtenus à partir des expressions d'opinion repérées, des opérateurs influant sur l'opinion, ou des opinions implicites relevées :

- Le type de subjectivité du segment : est-il porteur d'opinion (*SE, SI, SEI*) ? Est-il objectif (*O*) ?
- La polarité du segment : la polarité des expressions d'opinion corrigée par application des opérateurs (en particulier des négations)
- L'intensité du segment : l'intensité des expressions d'opinion, corrigée par application des opérateurs (en particulier des intensifieurs et des négations)
- Le sujet concerné par l'opinion du segment, et plus particulièrement son type : s'agit-il du sujet principal du document (*MT*), d'une partie de ce sujet (*PT*), ou d'un sujet totalement distinct (*OT*) ?

À cela, nous pouvons ajouter également les informations issues de la structure discursive :

- La présence d'une relation entre deux segments
- Le type de la relation (*Coordonnante* ou *Subordonnante*)
- La sémantique de la relation (*Continuation, Contraste, Élaboration, ...*)

L'objectif est donc d'établir un algorithme de détermination de l'opinion globale, basé sur ces éléments.

**Remarque :**

Par souci de concision dans certaines figures, le sujet de l'opinion pourra parfois être qualifié par le terme emprunté à l'anglais *topic*. De même, l'émetteur sera parfois appelé *holder*.

### 7.1.2. Notre approche

Nous distinguons deux étapes principales dans le processus de fusion :

- Une étape de constitution de l'ensemble d'opinions pertinentes, qui peut s'effectuer :
  - o Par filtrage du graphe de discours : dans cette étape, certaines opinions (associées aux segments du graphe), jugées non pertinentes, sont élaguées afin de restreindre les opinions prises en compte.
  - o Par règles : dans cette étape, des règles sont associées aux relations discursives afin de fusionner l'information d'opinion des segments source et cible
- Une étape de calcul numérique : application d'une heuristique de calcul ou de sélection sur l'ensemble issu de l'étape précédente, afin de calculer une valeur d'opinion unique pour le graphe

Pour la première étape (constitution de l'ensemble d'opinions), la principale nuance entre une étape de filtrage pure et une remontée par règles est que, dans le second cas, l'ensemble des opinions remontées n'est plus nécessairement un sous-ensemble des opinions présentes dans le graphe. En effet, il est dans ce cas possible d'associer à un type de relation discursive un résultat qui ne soit ni l'opinion portée par le segment source, ni celle remontée par le segment cible.

Nous souhaitons garder l'étape d'application d'heuristique numérique dans le cas d'une première étape par règles : en effet, cela nous permet de disposer de plus de souplesse pour l'établissement des règles. Ainsi, le résultat de nos règles peut-être un ensemble de valeurs d'opinion, dont la résolution peut s'effectuer soit plus en amont dans le graphe, soit via une heuristique de calcul.

La figure 7.1 ci-dessous illustre ce processus.

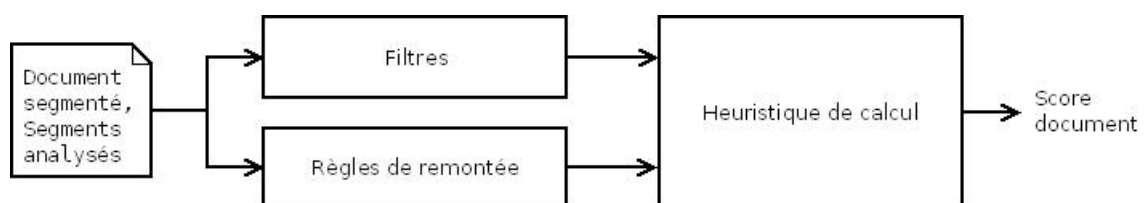


Figure 7.1 : Déroulement général du processus de fusion

Ce chapitre s'articule en cinq sections. La section 7.2 présente un cas d'étude, extrait de notre corpus de commentaire web, qui permettra d'illustrer les méthodes présentées dans ce chapitre. La section 7.3 décrit les différents filtres testés, ainsi qu'un processus d'élagage préservant la structure discursive. Nous proposons en section 7.4 un jeu de règles de remontée de l'opinion

basées sur le discours, et pouvant se substituer aux filtres. La section 7.5 expose les différentes heuristiques de calcul testées. Enfin, nous présentons et analysons les résultats obtenus dans la section 7.6.

## 7.2. Cas d'étude

Le document suivant, très positif, est extrait de notre corpus de commentaires web – il sera considéré comme exemple *fil rouge* tout au long de ce chapitre. Ce commentaire s'est vu attribué manuellement la note de +3, ce qui constitue la note maximale sur notre échelle (i.e. entre -3 et +3). Le document a été retranscrit tel qu'écrit par l'internaute. Le découpage en crochets numérotés correspond à la segmentation discursive du texte, effectuée telle que présentée en section 3.2.2.

[Que dire ?]\_1 [Magnifique,]\_2 [déroutant,]\_3 [passionnant ?]\_4 [Une belle]\_5 [et grande réussite de la mise en scène]\_6 [qui nous entraîne dans la vie de B. Dylan.]\_7 [La performance de chacun des acteurs est exceptionnelles...]\_8 [Certains spectateurs ont quitté la salle.]\_9 [Peut-être est-ce à cause du côté trop léché de chaque séquence]\_10 [ou une méconnaissance de l'artiste.]\_11 [Moi je suis resté... SCOTCHE... !]\_12

Le Tableau 7.1 reprend les valeurs d'opinions annotées (*gold*) pour chacun des segments.

Segment	Subjectivité	Valeur d'opinion	Sujet associé <sup>55</sup>
1	SN	∅	∅
2	SE	+3	mt (implicite)
3	SE	+2	mt (implicite)
4	SE	+3	mt (implicite)
5	SE	+2	pt (mise-en-scène)
6	SE	+3	pt (mise-en-scène)
7	O	∅	∅
8	SE	+3	pt (acteurs)
9	SI	-2	mt (implicite)
10	SEI	-2	pt (mise-en-scène)
11	SE	-1	pt (l'artiste)
12	SE	+3	mt (implicite)

Tableau 7.1 : Valeurs de subjectivité et d'opinion pour chacun des segments

Enfin, la structure rhétorique du document a été annotée de la façon suivante (les segments complexes sont indiqués par la liste entre crochets de leurs constituants) :

- Continuation(2,3)

<sup>55</sup> Le sujet associé est soit lexicalisé au sein d'un segment, soit implicite et inféré automatiquement. La section 6.5 décrit plus en détails l'obtention des données à partir de l'annotation.

- Continuation(3,4)
- Elaboration(1,[2,3,4])
- Continuation(5,6)
- E-Elab(6,7)
- Elaboration(1,[5,6,7])
- Elaboration(1,8)
- Alternative(10,11)
- Explication(9,[10,11])
- Elaboration(1,[9,10,11])
- Contrast([9,10,11],12)

Cette structure peut être représentée graphiquement par le graphe suivant :

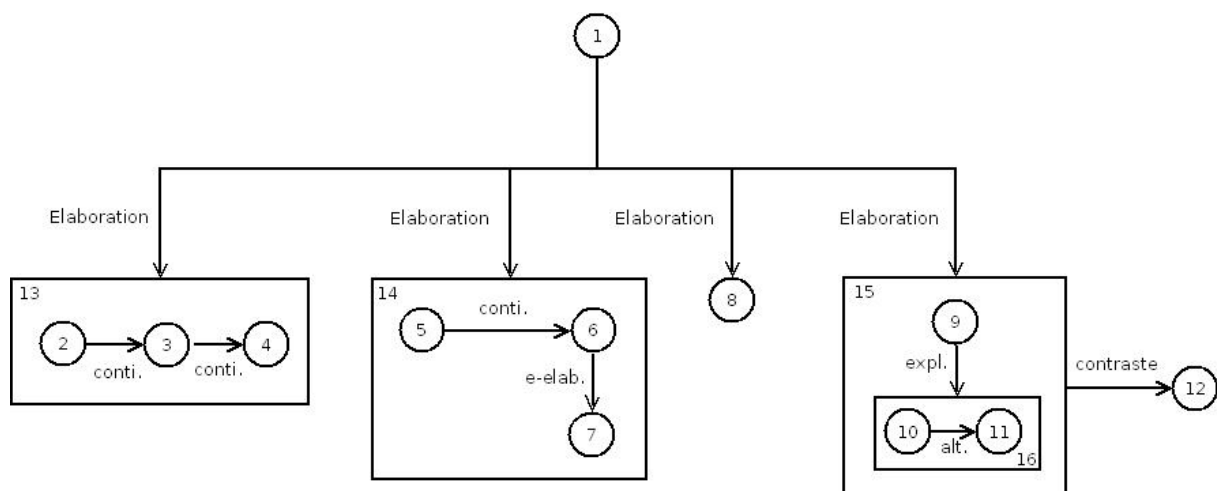


Figure 7.2 : Graphe discursif correspondant à l'exemple cas d'étude

### 7.3. Stratégies de filtrages

Les stratégies d'élagage consistent à appliquer, préalablement à l'utilisation d'une heuristique de calcul, un ou plusieurs filtres permettant d'ignorer certaines parties du document.

Nous définissons trois types de filtres utilisables pour sélectionner les segments d'entrée :

- filtres par type de sujet de l'opinion : décrits en section 7.3.1.
- filtres par catégorie de subjectivité du segment : décrits en section 7.3.2.
- filtres discursifs, basés sur le graphe de discours : décrits en section 7.3.3.

Un quatrième filtre, par émetteur de l'opinion, pourrait être envisagé. Néanmoins, nos données comportent globalement peu d'instances d'émetteur explicite de l'opinion, et ne nous permettent pas d'avoir des statistiques pertinentes sur cet élément.

Enfin, nous discuterons en section 7.3.4 du processus pratique d'application des filtres : caractère connexe du graphe résultant (section 7.3.4.1), ordre d'application (section 7.3.4.2), et possibilités de relaxation des contraintes (section 7.3.4.3).



### 7.3.1. Filtres par type de sujet de l'opinion

Les filtres par types de sujet de l'opinion consistent à filtrer suivant le type du sujet de l'opinion relativement au sujet principal. De ce fait, nous testons les quatre filtres suivant :

- *keep\_all* : pas de filtrage sur les sujets de l'opinion (filtre désactivé).
- *keep\_mt\_pt* : filtre ne conservant que les segments concernant le *topic* principal, soit directement (comme *keep\_mt*), soit indirectement (comme *keep\_pt*). Ce filtre élimine en pratique uniquement les opinions portant sur un sujet reconnu comme *autre*, i.e. sans rapport avec le sujet principal.
- *keep\_mt* : filtre ne conservant que les segments concernant directement le *topic* principal.
- *keep\_pt* : filtre ne conservant que les segments concernant une partie du *topic* principal (mais pas directement le *topic* principal).

#### Exemple cas d'étude :

Le tableau suivant présente les ensembles de segments et d'opinions résultants de l'application de chacun des filtres par sujet de l'opinion.

Filtre par topic employé	Ensemble de segments résultant	Ensemble d'opinions résultant
<i>keep_all</i>	{1; 2; 3; 4; 5; 6; 7; 8; 9; 10; 11; 12}	{+3; +2; +3; +2; +3; +3; -2; -2; -1; +3}
<i>keep_mt_pt</i>	{2; 3; 4; 5; 6; 8; 9; 10; 11; 12}	{+3; +2; +3; +2; +3; +3; -2; -2; -1; +3}
<i>keep_mt</i>	{2; 3; 4; 9; 12}	{+3; +2; +3; -2; +3}
<i>keep_pt</i>	{5; 6; 8; 10; 11}	{+2; +3; +3; -2; -1}

Tableau 7.2 : Application de chaque filtre par sujet de l'opinion sur l'exemple cas d'étude (cf. section 7.2)

### 7.3.2. Filtres par subjectivité

Les filtres par type de subjectivité consistent à filtrer les segments implicites (SI), explicites (SE), ou mixtes (SEI). Cinq filtres sont testés :

- *keep\_all* : tous les segments subjectifs sont pris en compte
- *keep\_se* : seuls les segments explicites sont pris en compte
- *keep\_si* : seuls les segments implicites sont pris en compte
- *keep\_se\_sei* : seuls les segments explicites ou mixtes sont pris en compte
- *keep\_si\_sei* : seuls les segments implicites ou mixtes sont pris en compte

#### Rappel :

Les segments non porteurs d'opinion, objectifs (O) et subjectifs non évaluatifs (SN), sont dans tous les cas ignorés pour le calcul de l'opinion globale. En effet, ceux-ci ne portent pas intrinsèquement d'opinion évaluative.

**Exemple cas d'étude :**

Le tableau suivant présente les ensembles de segments et d'opinions résultants de l'application de chacun des filtres par type de subjectivité.

Filtre par subjectivité employé	Ensemble de segments résultant	Ensemble d'opinions résultant
keep_all	{1; 2; 3; 4; 5; 6; 7; 8; 9; 10; 11; 12}	{+3; +2; +3; +2; +3; +3; -2; -2; -1; +3}
keep_se	{2; 3; 4; 5; 6; 8; 9; 11; 12}	{+3; +2; +3; +2; +3; +3; -2; -1}
keep_se_sei	{2; 3; 4; 5; 6; 8; 9; 10; 11; 12}	{+3; +2; +3; +2; +3; +3; -2; -2; -1; +3; +3}
keep_si	{9}	{-2}
keep_si_sei	{9; 10}	{-2; -2}

Tableau 7.3 : Application de chaque filtre par sujet de l'opinion sur l'exemple cas d'étude

### 7.3.3. Filtres par discours

Les filtres par discours consistent à raisonner sur le graphe de discours afin de proposer un élagage. Nous proposons cinq filtres :

- *keep\_all* : tous les segments sont pris en compte.
- *keep\_top* : seuls les constituants (EDU ou CDU) du haut du graphe (i.e. n'étant ni subordonné à un autre segment, ni coordonné à un segment étant lui-même subordonné à un autre segment) sont gardés. Les segments simples (EDU) appartenant à une CDU du haut du graphe sont tous gardés.
- *keep\_top\_recurisif* : seuls les segments du haut du graphe (i.e. n'étant ni subordonné à un autre segment, ni coordonné à un segment étant lui-même subordonné à un autre segment) sont gardés. Les segments simples (EDU) appartenant à une CDU du haut du graphe ne sont gardés que s'ils ne sont pas subordonnés à un autre segment à l'intérieur de la CDU : la règle de filtrage est ainsi appliquée récursivement pour chaque sous-graphe constitué par une CDU.
- *keep\_sub\_src* : seuls les segments source d'au moins une relation discursive subordonnante sont conservés. L'idée derrière ce filtre est que si le rédacteur explique, élabore, commente... un segment, ce dernier possède plus d'importance que les autres.
- *keep\_sub\_src\_notDest* : seuls les segments source d'au moins une relation discursive subordonnante sont conservés. L'idée derrière ce filtre est la même que la précédente, à ceci près que cette version filtre les subordinations en série (e.g. l'élaboration d'un commentaire par exemple).

L'application des filtres présentés ci-dessus sont extrêmement restrictifs dans le cadre de l'exemple cas d'étude (cf. section 7.2) : pour chacun de ces filtres, seul le segment racine 1 – non porteur d'opinion – est conservé. Par conséquent, l'application de ces filtres est illustrée par un exemple spécifique, présenté ci-après.

**Exemple spécifique** : application des filtres *keep\_top*, *keep\_top\_recurif*, *keep\_sub\_src*, et *keep\_sub\_src\_notDest* :

La figure 7.3 représente un graphe discursif dont les relations ont été simplifiées pour ne garder que leur caractère coordonnant ou subordonnant.

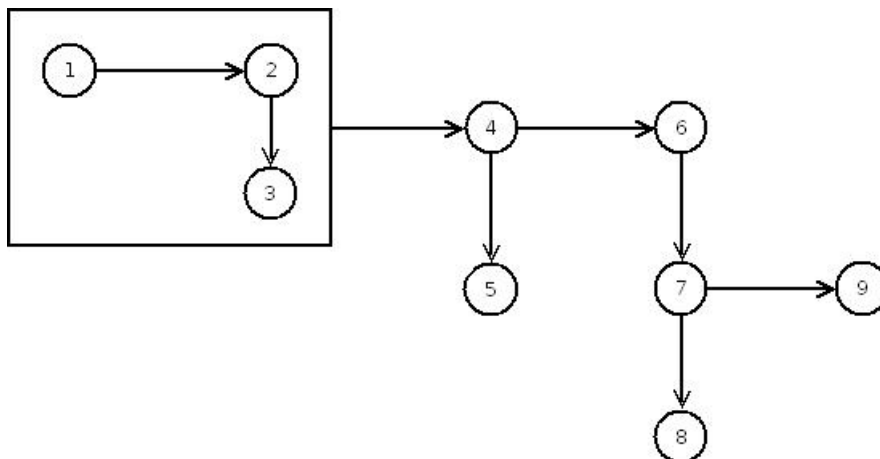


Figure 7.3 : Exemple simplifié de graphe de discours

L'application du filtre *keep\_top* conduit au graphe de discours suivant. Les segments filtrés sont grisés.

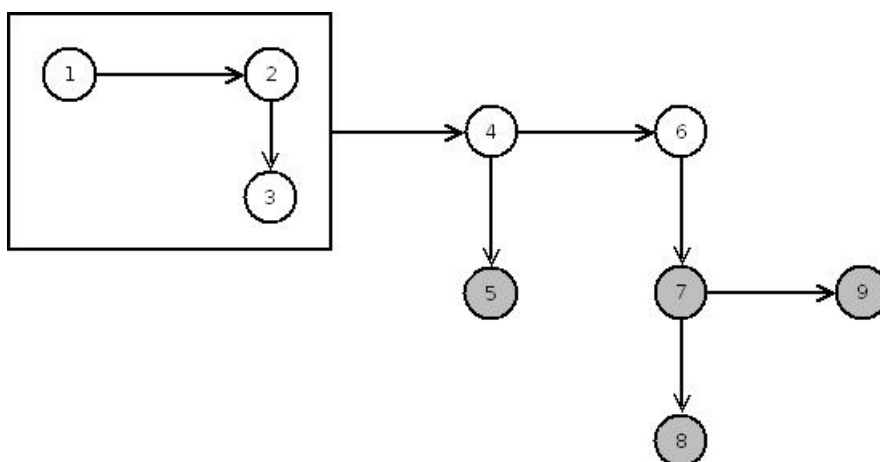


Figure 7.4 : Graphe après application du filtre *keep\_top*

De même, la figure 7.5 montre l'application du filtre *keep\_top\_recurif* sur le graphe de discours : le filtre est appliqué récursivement à l'intérieur de la CDU, éliminant le segment 3.

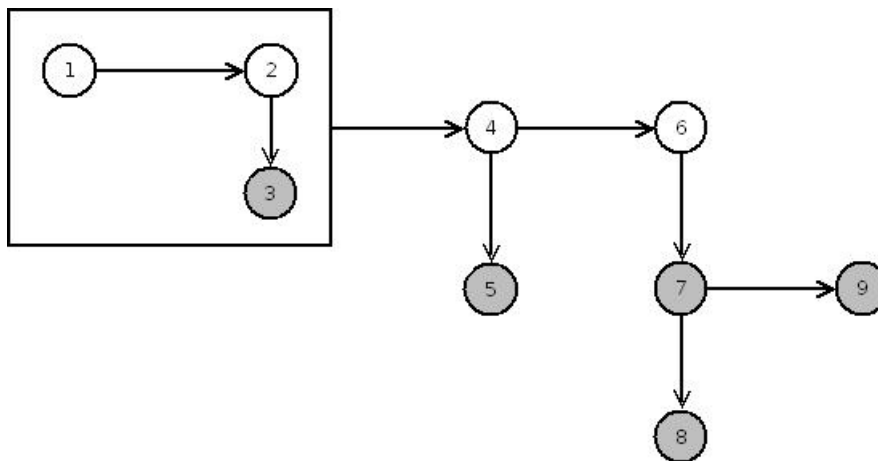


Figure 7.5 : Graphe après application du filtre *keep\_top\_recuratif*

La figure 7.6 illustre l'application du filtre *keep\_sub\_src* sur le graphe.

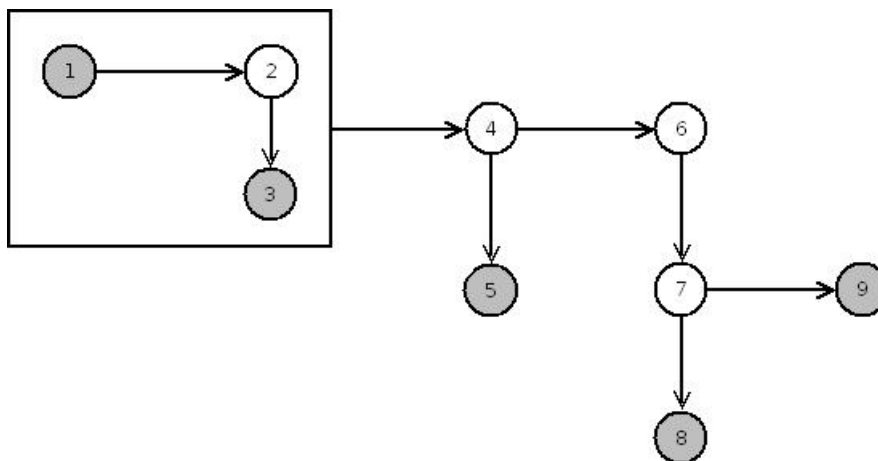


Figure 7.6 : Graphe après application du filtre *keep\_sub\_src*

Enfin, la figure suivante montre le graphe de discours après application du filtre *keep\_sub\_src\_not\_dest*. Le nœud 7 est ici filtré, car bien que subordonnant du nœud 8, il est subordonné au nœud 6.

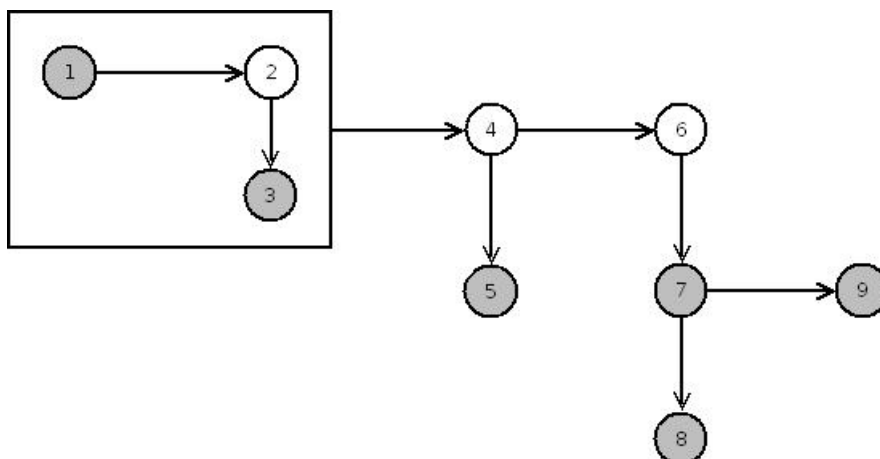


Figure 7.7 : Graphe après application du filtre *keep\_sub\_src\_not\_dest*

### 7.3.4. Processus d'application des filtres

Nous avons proposé trois filtres applicables sur un graphe de discours permettant de restreindre l'ensemble de segments utilisé par l'heuristique de calcul de l'opinion globale, basés respectivement sur le sujet de l'opinion associée à un segment, sur le type de subjectivité d'un segment, et sur la structure discursive reliant un segment à ses voisins.

Se posent maintenant trois questions en vue d'appliquer ces filtres :

- Que doit-on faire des segments filtrés ? Doit-on supprimer les nœuds du graphe, où doit-on essayer de garder le graphe connexe ?
- L'ordre d'application des filtres a-t-il une importance ? Si oui, y-a-t-il un ordre d'application préférable ?
- Que faire dans le cas où les filtres seraient trop restrictifs sur un document donné, ne laissant passer aucun segment porteur d'opinion ?

#### 7.3.4.1. Graphe résultant : le garder connexe ?

Jusqu'à présent, nous n'avons pas détaillé ce qu'il advenait des nœuds du graphe filtrés. Sachant que l'objectif principal est que ces nœuds n'interviennent plus dans le calcul de l'opinion globale, nous distinguons plusieurs options qui s'offrent à nous.

L'option la plus simple (*ilots*) est de retirer simplement ces nœuds du graphe, ainsi que les relations les concernant. Cette méthode, possède l'inconvénient de diviser le graphe en potentiellement plusieurs sous-graphes non connexes : les heuristiques discursives ne pourront alors pas s'appliquer.

Une autre option (*objectivation*) est de considérer chacun des segments filtrés comme objectifs. Cette option permet de conserver la structure du graphe intacte.

Enfin, une troisième option (*élagage\_physique*) est de retirer non seulement les nœuds du graphe filtrés, mais également les nœuds dépendant directement de celui-ci. Cette option est

relativement forte, et on peut observer des graphes résultant vides ou sans subjectivité. Cette option ne garantit pas le caractère connexe du graphe final.

### Exemple cas d'étude :

La figure suivante représente le graphe résultant de l'application de l'option d'élagage *ilots* sur le cas d'étude (cf. section 7.2). On obtient quatre sous-graphes isolés.

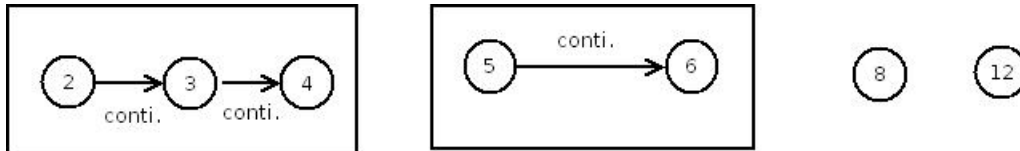


Figure 7.8 : Graphe de discours de l'exemple cas d'étude avec élagage par règles, option *ilots*.

La figure suivante représente le graphe résultant de l'application de l'option d'élagage *objectivation*. Les segments non porteurs d'opinion ou considérés comme tels après filtrage sont grisés.

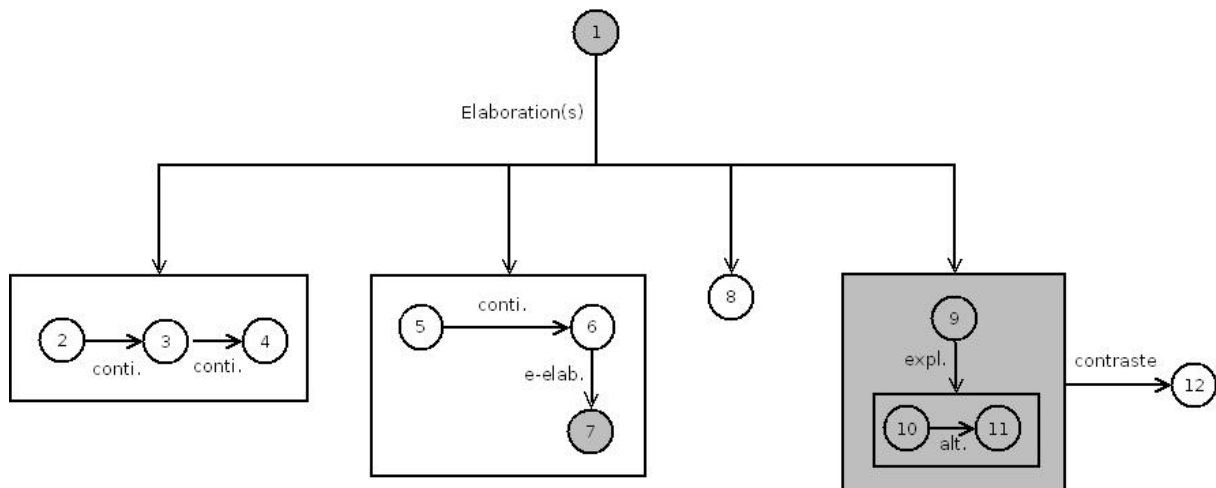


Figure 7.9 : Graphe de discours de l'exemple cas d'étude avec élagage par règles, option *objectivation*.

Enfin, la figure suivante présente le graphe résultant de l'application de l'option *élagage\_physique*. On obtient trois sous-graphes isolés. Le segment 12 est éliminé avec la CDU le raccrochant.

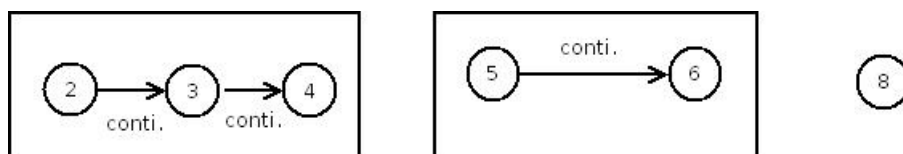


Figure 7.10 : Graphe de discours de l'exemple cas d'étude avec élagage par règles, option *élagage\_physique*.

Afin de pouvoir tester la totalité des combinaisons de filtres et d'heuristique de calcul, nous avons opté pour un filtrage par *objectivation*, qui permet facilement de maintenir la cohérence du graphe de discours

#### 7.3.4.2. Ordre d'application des filtres

Le mode de filtrage choisi en section précédente, n'altère ni la structure discursive, ni le rattachement d'un segment à au sujet de l'opinion. Par contre, il altère le type de subjectivité des segments. Afin d'appliquer les filtres et puisque la structure discursive est conservée, nous pouvons soit raisonner sur le graphe initial, soit sur le graphe altéré par le filtrage.

Si l'on choisit de raisonner sur le graphe initial, l'ordre d'application des filtres n'a pas d'importance. Quelque soit celui effectué en pratique, le résultat sera identique.

Si par contre on choisit d'effectuer le filtrage à partir du graphe en cours de filtrage, l'ordre d'application peut influencer sur le résultat. Le filtre par sujet de l'opinion n'est pas influencé par la subjectivité du segment, et peut donc être appliqué en n'importe quelle position. Le filtre par subjectivité dépend bien sûr du type de subjectivité du segment. Néanmoins, cette altération ne touche que les segments déjà filtrés par un éventuel filtre amont ; en pratique, ce filtre peut donc également être appliqué en n'importe quelle position. Par contre, le type de subjectivité d'un segment altère la décision prise par les filtres discursifs (en particulier le filtre *keep\_by\_rules*) sur les segments voisins, filtrés ou non. Ce filtre doit donc être appliqué en priorité afin de pouvoir raisonner sur le document non altéré.

#### 7.3.4.3. Relaxation des filtres

Certaines combinaisons de filtres peuvent être très restrictives. Par conséquent, il n'est pas exclu que l'application des filtres sur un document renvoie un ensemble de segments vide, ou composé uniquement de segments objectifs.

Pour pallier à ce cas, nous proposons une méthode de relaxation des filtres. Si un jeu de filtres aboutit à un document sans segment subjectif, alors :

- On tente de relaxer le filtre par sujet de l'opinion : les filtres *keep\_mt* et *keep\_pt* sont assouplis en *keep\_mt\_pt*, et *keep\_mt\_pt* est assoupli en *keep\_all* (cf. Figure 7.11 : graphe (1)).
- Si cela n'est pas possible ou si le filtre n'aboutit pas à un ensemble de segments subjectifs non vide, on tente de relaxer le filtre par subjectivité : les filtres *keep\_se* et *keep\_si* sont relaxés en filtre *keep\_all* (cf. Figure 7.11 : graphe (2)).
- Enfin, si les relaxations précédentes ont échouées, on tente de relaxer le filtre discursif : *keep\_top*, *keep\_top\_recuratif*, *keep\_sub\_src* et *keep\_sub\_src\_nonDest* sont relaxés en *keep\_all* (cf. Figure 7.11 : graphe (3)).

La figure suivante représente les règles de relaxation des filtres, pour le filtre par sujet de l'opinion (1), pour le filtre par subjectivité (2), et pour le filtre par structure discursive (3).

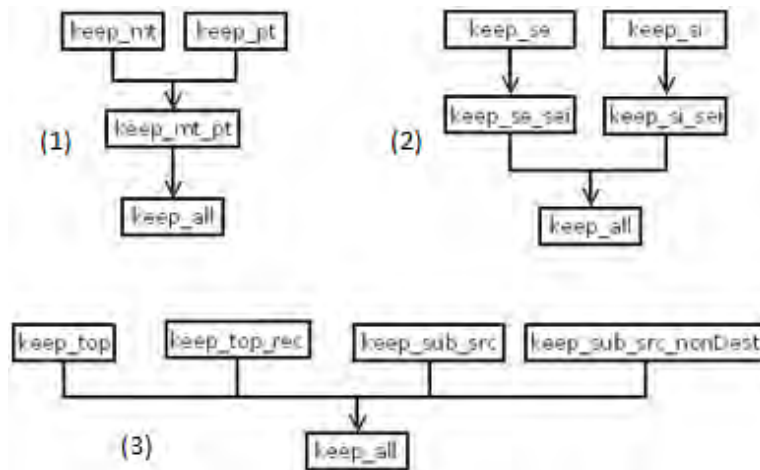


Figure 7.11 : Règles de relaxation des filtres : la flèche s'interprète comme "se relaxe en"

#### 7.3.4.4. Conclusion : procédure globale d'application des filtres

La figure suivante 7.12, détail de la figure 7.1, résume le processus et l'ordre d'application des filtres.

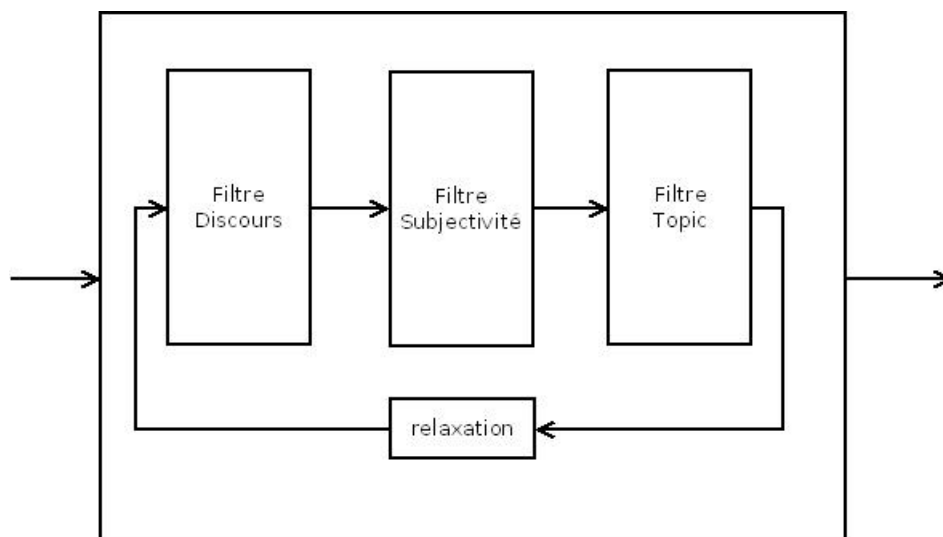


Figure 7.12 : Processus d'application des filtres

### 7.4. Remontée de l'opinion guidée par la structure discursive

L'application de filtres basés sur le sujet de l'opinion, la subjectivité du segment, ou le type de relation reliant ce segment au reste du graphe est une solution simple permettant de restreindre le nombre de segments pertinents sur lequel appliquer une heuristique de calcul. Néanmoins, prises dans leur ensemble, ces informations peuvent être utilisées de manière plus précise pour déterminer un ensemble d'opinions représentant de manière pertinente le document.

Dans cette section, nous présentons une approche guidée par la structure discursive permettant de remonter un tel ensemble, par simplification récursive des CDU, et parcours en profondeur du graphe de discours. Cette approche se base sur des règles spécifiques à chaque relation, décrites



en section 7.4.3. En préalable, nous présentons une formalisation de la tâche effectuée (section 7.4.1), et décrivons l'algorithme d'application des règles (section 7.4.2).

### 7.4.1. Tâche effectuée : formalisation

Le principe de cet algorithme est la remontée d'un segment élémentaire de discours équivalent au segment complexe considéré, le graphe complet du document étant considéré comme un cas particulier de segment complexe.

Formellement, étant donné un graphe de discours  $Gr$ , l'objectif est de remonter un triplet unique  $\langle Top, Sub, Val_{Set} \rangle$ , où  $Top$  est le sujet de l'opinion,  $Sub$  le type de subjectivité du segment, et  $Val_{Set}$  un ensemble de valeurs d'opinion (polarité et intensité) correspondant au contenu sémantique du segment. Cette sous-tâche de la fusion d'opinion peut donc se substituer au filtrage vu précédemment.

### 7.4.2. Algorithme de parcours du graphe

#### Traitement préalable

Le graphe de discours est préalablement normalisé, en extériorisant les liens sortant des CDU. Formellement, toute relation  $R(n1, n2)$  où  $n1 \subset C$ , avec  $C$  une CDU et  $n2 \not\subset C$ , est transformée en  $R(C, n2)$ .

#### Algorithme de remontée de l'opinion

Cette section décrit, en langage algorithmique, le processus de parcours du graphe et de remontée de l'opinion. L'algorithme consiste à associer à chaque CDU, en partant de la CDU la plus interne, un triplet de valeurs associées  $Valeurs = (Subjectivité, Topic, Score)$ .

Pour cela, nous pouvons définir les fonctions `Associer_valeurs`, qui permettent respectivement d'affecter un triplet de valeurs  $v$  à une CDU  $c$ , ou de mettre à jour les valeurs associées à un nœud du graphe  $n$ . Un nœud du graphe peut être soit un segment simple, soit une CDU imbriquée précédemment résolue.

```
Associer_valeurs(CDU c, Valeurs v)
Associer_valeurs(Nœud n, Valeurs v)
```

La fonction `Valeurs(Nœud n)` permet d'accéder au triplet de valeurs présentement associé au nœud  $n$ .

Une CDU est dite résolue lorsqu'elle se voit affectée ces valeurs, i.e. lorsque toutes les relations internes à la CDU ont été traitées. Le graphe complet est traité comme un cas particulier de CDU, englobant tout le graphe. Pour chaque CDU  $c$  non résolue du graphe, on applique la fonction `Solve_CDU(c)` :

```
Function Solve_CDU(CDU c) {
    Tant que (( $\exists$  CDU c') && (c'  $\subset$  c)) {
        Solve_CDU(c');
    }
}
```

```

    Nœud courant := nœud le plus à gauche dans les relations internes;
    Solve_node(courant);
    Associer_valeurs(c, Valeurs(courant));
}

```

La fonction `Solve_node` consiste à appliquer les règles sur les relations reliant le nœud courant et les nœuds suivant dans le graphe, en appliquant un parcours en profondeur.

```

Solve_node(Nœud courant){
    Tant que (∃ Relation R && R subordonnante && R(courant, autre)){
        Solve_node(autre);
        val = Appliquer_regle(R, courant, autre);
        Associer_valeurs(courant, val);
    }
    Si (∃ Relation R && R coordonnante && R(courant, autre)){
        Relation Rcoo = selection_coo(courant, autre);
        val = Appliquer_regle(R, courant, autre);
        Associer_valeur(autre, val);
    }
}

```

La fonction `Appliquer_regle` applique, pour la relation  $R$  reliant les nœuds source et cible donnés, la règle associée et retourne les valeurs fusionnées. La section suivante décrit, pour chaque relation  $R$  et suivant les valeurs des nœuds source et cible reliés, le triplet de valeur (Subjectivité, Topic, Score) retournés.

La SDRT spécifie qu'un nouveau nœud ne peut être rattaché au graphe que sur un nœud situé sur la frontière droite de celui-ci. En pratique, cela interdit d'avoir plusieurs relations coordonnantes partageant le même nœud source, mais pas la même destination, d'où l'unicité annoncée et exploitée dans l'algorithme pour le nœud de destination des relations coordonnantes.

La fonction `selection_coo` permet, dans le cas où l'on aurait plusieurs relations coordonnantes issues d'un même nœud courant, de sélectionner une relation dite *principale*, dont la sémantique est plus forte que celle des autres relations, relativement à la notion d'opinion. L'ordre de priorité des relations est donné dans le tableau suivant : plus l'indice est élevé, plus la relation est prioritaire.

Relation	Priorité associée
Conditionnelle	8
Contraste	7
Résultat	6
But	5
Parallèle	4
Narration / flashback	3
Alternative	2
Continuation	1

Tableau 7.4 : Priorités des relations coordonnantes

### 7.4.3. Règles discursives de remontée de l'opinion

Chaque règle associée à une relation  $R(a, b)$  reliant deux nœuds  $a$  (source) et  $b$  (cible) un triplet de valeurs  $\langle Top, Sub, Val_{set} \rangle$ .

Ce triplet de valeurs est calculé à partir :

- Du type de  $R$  ;
- Du triplet de valeurs  $\langle topic(a), subj(a), score(a) \rangle$  associé à  $a$  ;
- Du triplet de valeurs  $\langle topic(b), subj(b), score(b) \rangle$  associé à  $b$ .

Les règles suivent donc un principe de compositionnalité.

**Remarque :** La subjectivité  $S$  désigne une subjectivité indéterminée  $SE$  ou  $SI$ .

#### 7.4.3.1. Relation de Continuation

##### Sémantique de la relation

La relation de Continuation est une relation coordonnante à la sémantique relativement faible, qui indique que le segment  $b$  continue le rôle rhétorique du segment  $a$ . Elle relie par contre deux segments partageant un même sujet, et en général, le caractère subjectif ou objectif.

Par définition, nous avons donc les propriétés suivantes :

$$\begin{aligned} topic(a) &= topic(b) = topic \\ subj(a) &= subj(b) = subj \end{aligned}$$

En l'absence de sémantique forte, il n'est pas possible de privilégier l'opinion portée par la partie gauche ou droite de la relation à ce niveau du discours.

##### Règle associée

Pour le triplet résultant  $\langle Top, Sub, Val_{set} \rangle$ , nous avons :

- $Top = topic$  : le type de sujet commun
- $Sub = subj$  : la subjectivité commune
- $Val_{set} = score(a) \cup score(b)$  : les opinions portées ont même importance

##### Exemples

- Considérons la portion de discours suivante, extraite d'un commentaire de film :  
(1) [Les seconds rôles sont bons,]<sub>a</sub> [la musique du générique est super.]<sub>b</sub>

Le graphe de discours associé est le suivant :

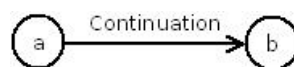


Figure 7.13 : Arbre discursif associé à l'exemple 1

Le tableau suivant présente les triplets de valeurs associés à chacun des deux segments :

	<b>1.a</b>	<b>1.b</b>
<b>topic</b>	pt (les seconds rôles)	pt (la musique du générique)
<b>subj</b>	S (SE)	S (SE)
<b>score</b>	+2	+3

Tableau 7.5 : Triplets de valeurs associés aux segments de l'exemple 1

Le triplet résultant  $\langle Top, Sub, Val_{Set} \rangle$  est donc :  $\langle pt, S, \{+2; +3\} \rangle$ .

- Considérons la portion de discours suivante, extraite également d'un commentaire de film :  
(2) [Les dialogues n'ont aucun intérêt]<sub>a</sub> [et l'intrigue est quasi-inexistante.]<sub>b</sub>

Le graphe discursif associé est le suivant :

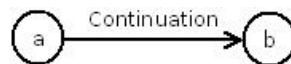


Figure 7.14 : Graphe discursif associé à l'exemple 2

Le tableau suivant présente les triplets de valeurs associés à chacun des deux segments :

	<b>2.a</b>	<b>2.b</b>
<b>topic</b>	pt (les dialogues)	pt (l'intrigue)
<b>subj</b>	S (SE)	S (SE)
<b>score</b>	-2	-2

Tableau 7.6 : Triplets de valeurs associés aux segments de l'exemple 2

Le triplet résultant  $\langle Top, Sub, Val_{Set} \rangle$  est donc :  $\langle pt, S, \{-2; -2\} \rangle$ . On peut remarquer que dans ce cas, on conserve la double valeur -2 dans l'ensemble  $Val_{Set}$ . Ceci est souhaitable : en effet, une continuation supplémentaire pourrait introduire une nouvelle valeur dans l'ensemble, potentiellement différente.

#### 7.4.3.2. Relations d'Elaboration et d'Explication

##### Sémantique des relations

La relation d'Elaboration relie deux segments décrivant chacun un état ou chacun un événement. Le segment cible (b) décrit un sous-état ou un sous-événement qui permet de préciser comment l'événement ou l'état décrit dans le segment a s'est produit. La relation d'Elaboration inclut également les cas d'exemplification, de reformulation et de paraphrase.

La relation d'explication relie quant à elle deux segments dont le second (b) explique pourquoi quelque chose s'est produit dans le segment a, ou pourquoi un agent a effectué une action dans le segment a.

Comme proposé par (Asher, Benamara, and Mathieu 2009), nous regroupons ces deux relations, dont les effets en termes d'opinion globale sont similaires. Dans les deux cas, tout ou partie du contenu du segment a est détaillé par le segment b.

### Règle associée

Pour le triplet résultant  $\langle Top, Sub, Val_{Set} \rangle$ , nous avons :

- $Top = topic(a)$  : le sujet de a est le sujet principal du bloc, il est élaboré ou expliqué dans le nœud b
- $Sub$  :
  - o  $Sub = S$  si  $subj(a) = S$  (SE ou SI) ou  $subj(b) = S$  (SE ou SI)
  - o  $Sub = O$  sinon
- $Val_{Set}$ :
  - o  $Val_{Set} = score(a)$  si  $subj(a) = S$  : l'opinion de a est élaborée ou expliquée par b, b ne fait que détailler et n'apporte a priori pas d'élément nouveau
  - o  $Val_{Set} = score(b)$  si  $subj(a) = O$

### Exemples

- Considérons la portion de discours suivante, extraite d'un commentaire de série TV :  
 (3) [Cela devient très vite gonflant,]<sub>a</sub> [Brooke, Haley et Nathan sont tellement plus intéressant que leur histoire.]<sub>b</sub>

Le graphe de discours associé est le suivant :

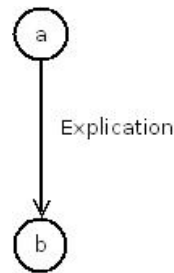


Figure 7.15 : Graphe discursif associé à l'exemple 3

Le tableau suivant présente les triplets de valeurs associés à chacun des deux segments :

	3.a	3.b
<b>topic</b>	mt (Cela)	pt (Brooke, Haley, Nathan)
<b>subj</b>	S (SE)	S (SE)
<b>score</b>	-3	+2

Tableau 7.7 : Triplets de valeurs associés aux segments de l'exemple 3

Le triplet résultant  $\langle Top, Sub, Val_{Set} \rangle$  est donc :  $\langle mt, S, \{-3\} \rangle$ . L'opinion positive portée sur les personnages (*Brooke, Haley et Nathan*) n'est pas gardée : on comprend en effet que cet élément n'a pas réussi à changer l'opinion du rédacteur sur la série TV.

- Considérons la portion de discours suivante, extraite d'un commentaire de film :  
(4) [Ce film est absolument somptueux.]<sub>a</sub> [C'est un vrai drame psychologique.]<sub>b</sub>

La structure discursive associée est la suivante :

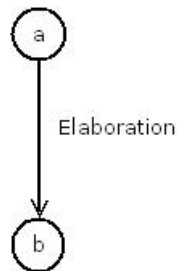


Figure 7.16 : Graphe discursif associé à l'exemple 4

Le tableau suivant présente les triplets de valeurs associés à chacun des deux segments :

	4.a	4.b
<b>topic</b>	mt (Ce film)	mt (C')
<b>subj</b>	S (SE)	S (SE)
<b>score</b>	+3	+2

Tableau 7.8 : Triplets de valeurs associés aux segments de l'exemple 4

Le triplet résultant  $\langle Top, Sub, Val\_Set \rangle$  est donc :  $\langle mt, S, \{+3\} \rangle$ . L'opinion, portée dans ce contexte par l'expression *vraie*, vient en support de l'opinion portée sur le film, mais n'apporte rien de neuf en termes de score résumé de l'opinion sur le sujet principal.

#### 7.4.3.3. Relation de Commentaire

##### Sémantique de la relation

La relation de Commentaire indique que le segment cible donne un point de vue d'un agent ou de l'auteur sur ce qui est décrit dans le segment source (a). Ce point de vue peut être exprimé via des modaux ou via des adjectifs évaluatifs, par exemple.

Par définition, nous avons donc la propriété  $topic(a) = topic(b) = topic$  : le segment b exprime un point de vue sur le sujet du segment a.

Dans le cas où les deux segments sont porteurs d'opinion, le commentaire vient soit en renforcement de l'opinion, soit en opposition. Dans ce dernier cas, les deux opinions ne peuvent pas être endossées en même temps par un locuteur cohérent : l'une d'entre elles dépend nécessairement d'un émetteur différent.

##### Règle associée

Pour le triplet résultant  $\langle Top, Sub, Val_{set} \rangle$ , nous avons :

- $Top = topic$  : le sujet commun aux deux segments
- $Sub = subj(b)$  : par définition, la subjectivité est ici portée par le segment cible

- $Val_{Set}$  :
  - $Val_{Set} = score(b)$  si  $subj(a) = 0$
  - $Val_{Set} = score(a)$  si  $subj(b) = 0$
  - $Val_{Set} = score(a) \setminus \{score(a_n)\} \cup \{Int_+(score(a_n))\}$  si  $score(a_n) * score(b) > 0$  : si les deux segments sont de même polarité, le commentaire vient renforcer l'opinion du segment commenté (mais pas celles éventuellement conservées depuis des relations coordonnantes précédentes)
  - $Val_{Set} = score(a) \setminus \{score(a_n)\}$  si  $score(a_n) * score(b) < 0$  : si la polarité est inconsistante entre les deux segments, nous ignorons les deux : le locuteur n'endosse probablement que l'un d'entre eux.

### Exemples

- Considérons la portion de discours suivante, extraite d'un commentaire de film :  
(5) [Un générique et sa musique excellent,]<sub>a</sub> [qui met la pêche.]<sub>b</sub>

Le graphe discursif associé est le suivant :

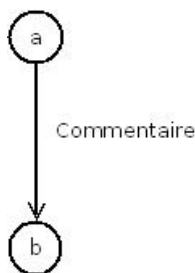


Figure 7.17 : Graphe discursif associé à l'exemple 5

Le tableau suivant présente les triplets de valeurs associés à chacun des deux segments :

	5.a	5.b
<b>topic</b>	pt (générique)	pt (générique)
<b>subj</b>	S (SE)	S (SE)
<b>score</b>	+3	+2

Tableau 7.9 : Triplets de valeurs associés aux segments de l'exemple 5

Le triplet résultant  $\langle Top, Sub, Val_{Set} \rangle$  est donc :  $\langle pt, S, \{+3\} \rangle (Int_+(+3) = +3$ , car il s'agit ici du maximum de l'échelle). L'opinion, portée dans ce contexte par l'expression *vraie*, vient en support de l'opinion portée sur le film, mais n'apporte rien de neuf en termes de score résumé de l'opinion sur le sujet principal.

- Considérons les portions de discours suivantes, dans le contexte d'un commentaire de film :  
(6) [Pierre a adoré.]<sub>a</sub> [Je ne comprends pas pourquoi.]<sub>b</sub>  
(7) [J'ai adoré.]<sub>a</sub> [Pierre ne comprends pas pourquoi.]<sub>b</sub>

La structure rhétorique associée est, dans les deux cas, la suivante :

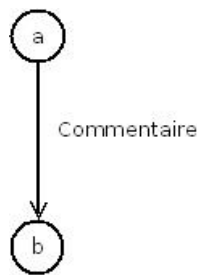


Figure 7.18 : Arbre discursif associé aux exemples 6 et 7

Le tableau suivant présente les triplets de valeurs associés à chacun des segments :

	6.a	6.b	7.a	7.b
<b>topic</b>	mt	mt	mt	mt
<b>subj</b>	S (SE)	S (SI)	S (SE)	S (SI)
<b>score</b>	+3	-2	+3	-2

Tableau 7.10 : Triplets de valeurs associés aux segments des exemples 6 et 7

Le triplet résultant  $\langle Top, Sub, Val\_Set \rangle$  est ici, dans les deux cas :  $\langle Ind., S, \emptyset \rangle$ . De manière symétrique, l'une des deux opinions de chacun des exemples n'est pas endossée par le locuteur, et l'orientation de la relation ne fournit pas un indice suffisant pour décider du segment contenant l'opinion pertinente pour notre tâche.

#### 7.4.3.4. Relation de Contraste

##### Sémantique de la relation

La relation de Contraste est une relation où les deux segments sont mis en opposition sémantique, soit au moyen d'un marqueur explicite de contraste (e.g. *mais, par contre, cependant*), soit implicitement.

Nous avons pu observer sur nos corpus qu'en ce qui concerne les opinions émises dans chacun des segments, la relation de Contraste n'est pas symétrique. En effet, considérons l'exemple suivant, dans le contexte d'un commentaire de film :

(8) [L'idée est originale,]<sub>a</sub> [mais il y a quelques passages creux]<sub>b</sub>

(9) [Il y a quelques passages creux,]<sub>a</sub> [mais l'idée est originale.]<sub>b</sub>

La phrase (8) donne un effet d'avertissement : le lecteur ne doit pas se fier uniquement sur l'idée annoncée, le film comporte des passages creux. À l'inverse (9) rassure le lecteur : même si on lui a annoncé des passages creux (sans doute dans un commentaire précédent, à moins que ce ne soit une caractéristique récurrente des films du scénariste), l'idée est originale et vaut le coup.

Partant de cette observation, nous proposons ici une règle favorisant l'opinion portée par la partie droite de la relation.



### Règle associée

Pour le triplet résultant  $\langle Top, Sub, Val_{set} \rangle$ , nous avons :

- *Top* :
  - $Top = topic(a) = topic(b)$  : si  $topic(a) = topic(b)$  (e.g. "Le film est bien, mais un peu long")
  - $Top = mt$  : sinon et si  $topic(a) = mt$  ou  $topic(b) = mt$  (e.g. "Le film est bien, dommage que les acteurs ne jouent pas bien." ; "J'ai apprécié le film, même si les fauteuils de la salle étaient très sales.")
  - $Top = pt$  : sinon et si  $topic(a) = pt$  ou  $topic(b) = pt$  (e.g. "Le scénario était quand même pas mal, même si celui d'Avatar était plutôt léger.", dans le cadre d'une critique d'un film de James Cameron.)
- *Sub* :
  - $Sub = S$  si  $subj(a) = S$  (SE ou SI) ou  $subj(b) = S$  (SE ou SI)
  - $Sub = O$  sinon
- *Val\_Set* :
  - Si  $topic(a) = topic(b) = mt$  ou  $topic(a) = topic(b) = pt$ 
    - $Val\_Set = score(b)$  si  $Max_{Int}(score(a)) \leq Max_{Int}(score(b))$  :
    - $Val\_Set = Int\_ (score(a))$  sinon
  - Si  $topic(a) = mt$ ,  $Val\_Set = Int\_ (score(a))$
  - Si  $topic(b) = mt$ ,  $Val_{set} = score(b)$

$Max_{Int}$  est ici une fonction renvoyant, pour un ensemble de valeurs d'opinion  $E$ , l'intensité maximale  $|E_i|$ , où  $E_i$  est une opinion de l'ensemble telle que  $\forall j E_j \in E, |E_j| \leq |E_i|$ .

### Exemples

- Considérons de nouveau la portion de discours suivante, extraite d'un commentaire de film:

(10) [L'idée est originale,]<sub>a</sub> [mais il y a quelques passages creux.]<sub>b</sub>

Le graphe associé est le suivant :

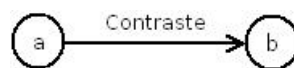


Figure 7.19 : Arbre discursif associé à l'exemple 10

Le tableau suivant présente les triplets de valeurs associés à chacun des deux segments :

	10.a	10.b
<b>topic</b>	pt (l'idée)	pt ([des] passages)
<b>subj</b>	S (SE)	S (SE)
<b>score</b>	+2	-1

Tableau 7.11 : Triplets de valeurs associés aux segments de l'exemple 10

Les segments partagent le même type de sujet  $pt$ , et le segment cible 10.b est d'intensité plus faible que le segment source 10.a. Le triplet résultant  $\langle Top, Sub, Val\_Set \rangle$  est donc :  $\langle pt, S, \{-1\} \rangle$ .

- Considérons la portion de discours suivante, extraite d'un commentaire de film :  
(11) [OK, c'est pas très crédible,]<sub>a</sub> [mais c'est un excellent moment !]<sub>b</sub>

Le graphe associé est le suivant :

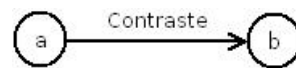


Figure 7.20 : Graphe discursif associé à l'exemple 11

Le tableau suivant présente les triplets de valeurs associés à chacun des deux segments :

	11.a	11.b
<b>topic</b>	mt	mt
<b>subj</b>	S (SE)	S (SE)
<b>score</b>	-1	+3

Tableau 7.12 : Triplets de valeurs associés aux segments de l'exemple 11

Les segments partagent ici le même type de sujet  $mt$ . Le segment source 11.a est légèrement négatif, mais le segment cible est fortement positif, ce qui masque l'aspect négatif soulevé. Le triplet résultant  $\langle Top, Sub, Val\_Set \rangle$  est donc :  $\langle mt, S, \{+3\} \rangle$ .

#### 7.4.3.5. Relation d'Elaboration d'entités (e-Elaboration)

##### Sémantique de la relation

La relation d'Elaboration d'entité lie deux segments dont le second (b) précise une propriété d'une des entités impliquées dans le premier segment (a).

##### Règle associée

Pour le triplet résultant  $\langle Top, Sub, Val\_Set \rangle$ , nous avons :

- $Top = topic(a)$  : le sujet de a est le sujet principal du bloc, en tant qu'entité, il est élaboré dans b
- $Sub$  :
  - o  $Sub = S$  si  $subj(a) = S$  (SE ou SI) ou  $subj(b) = S$  (SE ou SI)
  - o  $Sub = O$  sinon
- $Val\_Set$ :
  - o  $Val\_Set = score(b)$  si  $subj(a) = O$
  - o  $Val\_Set = score(a)$  si  $subj(b) = O$
  - o  $Val\_Set = Int_+(score(a))$  si  $score(a) * score(b) > 0$
  - o  $Val\_Set = Int_-(score(a))$  si  $score(a) * score(b) \leq 0$

### Exemple

Dans le contexte d'un commentaire de série TV :

(12) [Bon scénario,]<sub>a</sub> [qui se renouvèle continuellement.]<sub>b</sub>

Le graphe discursif associé est le suivant :

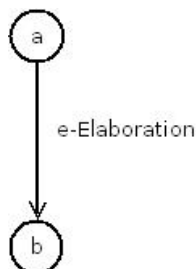


Figure 7.21 : Graphe discursif associé à l'exemple 12

Le tableau suivant présente les triplets de valeurs associés à chacun des deux segments :

	12.a	12.b
<b>topic</b>	pt (scénario)	pt (scénario)
<b>subj</b>	S (SE)	S (SI)
<b>score</b>	+2	+2

Tableau 7.13 : Triplets de valeurs associés aux segments de l'exemple 12

Les deux segments sont ici de même polarité positive. Le *renouvellement continu* du scénario vient renforcer l'expression d'opinion présente dans le segment 12.a. Le triplet résultant  $\langle Top, Sub, Val_{Set} \rangle$  est donc :  $\langle mt, S, \{+3\} \rangle$  ( $Int_+(+2) = +3$ ).

#### 7.4.3.6. Relation de Résultat

##### Sémantique de la relation

La relation de Résultat est une relation coordonnante reliant deux segments portant deux éventualités (i.e. événements ou états), dont le second, porté par le segment cible b, résulte ou est la conséquence du premier, porté par le segment source a.

La notion de Résultat, comme le Contraste, est une relation coordonnante asymétrique. Néanmoins, suivant le sujet de l'opinion pour chacun des segments, l'accent n'est pas nécessairement mis sur le segment cible. En effet, considérons l'exemple suivant, dans le contexte d'un commentaire de film :

(13) [Le film est excellent,]<sub>a</sub> [ce qui met en valeur le très bon jeu d'actrice de Nathalie Portman.]<sub>b</sub>

##### Règle associée

Pour le triplet résultant  $\langle Top, Sub, Val_{Set} \rangle$ , nous avons :

- *Top* :
  - o  $Top = topic(b)$  si  $topic(b) \neq ot$

- $Top = topic(a)$  sinon
- $Sub$  :
  - $Sub = S$  si  $subj(a) = S$  (SE ou SI) ou  $subj(b) = S$  (SE ou SI)
  - $Sub = O$  sinon
- $Val_{Set}$ :
  - Si  $topic(a) = topic(b)$ 
    - $Val_{Set} = score(b)$  si  $subj(b) = S$
    - $Val_{Set} = score(a)$  sinon
  - Si  $topic(b) = mt$  et  $topic(a) = pt$ ,  $Val_{Set} = score(b)$
  - Si  $topic(a) = mt$  et  $topic(b) = pt$ ,  $Val_{Set} = score(a)$
  - Si  $topic(b) \neq ot$ ,  $Val_{Set} = score(b)$
  - Sinon,  $Val_{Set} = score(a)$

### Exemple

Considérons la portion de discours suivante, extraite également d'un commentaire de série TV :

(14) [Je ne m'en souviens plus bien,]<sub>a</sub> [alors beaucoup de nostalgie !]<sub>b</sub>

Le graphe de discours associé est le suivant :

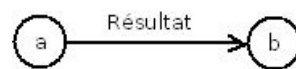


Figure 7.22 : Graphe discursif associé à l'exemple 14

Le tableau suivant présente les triplets de valeurs associés à chacun des deux segments, tel qu'annoté manuellement :

	14.a	14.b
<b>topic</b>	mt	mt
<b>subj</b>	S (SI)	S (SE)
<b>score</b>	+2	+3

Tableau 7.14 : Triplets de valeurs associés aux segments de l'exemple 14

Le segment 14.b (cible) étant subjectif, le triplet résultant  $\langle Top, Sub, Val_{Set} \rangle$  est le suivant :  $\langle pt, S, \{+3\} \rangle$ . On peut remarquer que la relation de Résultat permet également de désambiguïser la polarité de l'opinion portée par 14.a. Au final, l'information nécessaire à la compréhension est contenue dans la relation de Résultat et l'opinion du segment 14.b, ce qui est tout à fait consistant avec le score d'opinion remonté par la règle.

#### 7.4.3.7. Relation d'Attribution

##### Sémantique de la relation

La relation d'Attribution est une relation qui relie l'énonciation du segment cible b à son émetteur, explicité dans le segment source a.

Le segment source est typiquement un segment de subjectivité SN, donc non porteur d'opinion. L'opinion est portée par le segment b, et éventuellement altérée par la nature des termes utilisés pour attribuer le discours à son émetteur.

### Règle associée

Pour le triplet résultant  $\langle Top, Sub, Val_{Set} \rangle$ , nous avons :

- $Top = topic(b)$
- $Sub = subj(b)$
- $Val_{Set}$ :
  - o Si le segment a est un segment SN de type reportage, conseil, ou recommandation,  $Val_{Set} = score(b)$
  - o Si le segment a est de subjectivité SN et de type *espérer*,  $Val_{Set} = score(a)$  (c'est-à-dire, puisque a ne porte pas d'opinion, l'ensemble résultant des propagations précédentes) : la modalité boulique induit une opinion de certitude nulle, donc non pertinente pour l'opinion globale

### Exemple

Dans le contexte d'un commentaire de série TV :

(15) [Je pense]<sub>a</sub> [qu'elle serait encore meilleure !]<sub>b</sub>

Le graphe de discours associé est le suivant :

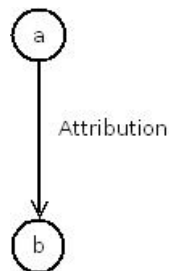


Figure 7.23 : Graphe discursif associé à l'exemple 15

Le tableau suivant présente les triplets de valeurs associés à chacun des deux segments :

	15.a	15.b
<b>topic</b>	Ind.	mt
<b>subj</b>	O (SN)	S (SE)
<b>score</b>	$\emptyset$	+3

Tableau 7.15 : Triplets de valeurs associés aux segments de l'exemple 15

Les deux segments sont ici de même polarité positive. Le *renouvellement continué* du scénario vient renforcer l'expression d'opinion présente dans le segment 15.a. Le triplet résultant  $\langle Top, Sub, Val_{Set} \rangle$  est donc :  $\langle mt, S, \{+3\} \rangle$  ( $Int_+(+2) = +3$ ).

### 7.4.3.8. Relation de Parallèle

#### Sémantique de la relation

Cette relation porte sur deux segments ayant une construction similaire (la plupart du temps syntaxique).

Par construction, la relation est symétrique. Seul le type de sujet de l'opinion permet d'établir une préférence relativement à la tâche de fusion sur l'un ou l'autre des segments.

#### Règle associée

Pour le triplet résultant  $\langle Top, Sub, Val_{Set} \rangle$ , nous avons :

- $Top$  :
  - o  $Top = topic(a) = topic(b)$  si  $topic(a) = topic(b)$
  - o  $Top = mt$  si  $topic(a) = mt$  ou  $topic(b) = mt$
  - o  $Top = pt$  si  $topic(a) = pt$  ou  $topic(b) = pt$
- $Sub = subj$ , valeur commune de subjectivité (la relation préserve, par définition, la subjectivité)
- $Val_{Set} = score(a) \cup score(b)$  : les opinions portées ont même importance

#### Exemple

Considérons la portion de discours suivante, extraite d'un commentaire de film :

(16) [Les acteurs paraissent ne pas y croire.]<sub>a</sub> [Les spectateurs non plus.]<sub>b</sub>

Le graphe de discours associé est le suivant :

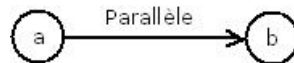


Figure 7.24 : Arbre discursif associé à l'exemple 16

Le tableau suivant présente les triplets de valeurs associés à chacun des deux segments :

	16.a	16.b
<b>topic</b>	pt (les acteurs)	mt
<b>subj</b>	S (SE)	S (SI)
<b>score</b>	-2	-2

Tableau 7.16 : Triplets de valeurs associés aux segments de l'exemple 16

Le triplet résultant  $\langle Top, Sub, Val_{Set} \rangle$  est donc :  $\langle pt, S, \{-2; -2\} \rangle$ .

### 7.4.3.9. Relation d'Encadrement (Frame)

#### Sémantique de la relation

La relation d'Encadrement est une relation subordonnante ayant pour argument un segment introducteur de cadre (segment source a), et le segment sur lequel porte ce cadre (segment cible b).

Le cadre indiqué par le segment a est en général objectif, ou comporte une opinion sur le cadre et non sur le sujet du document.

### Règle associée

Pour le triplet résultant  $\langle Top, Sub, Val_{Set} \rangle$ , nous avons :

- $Top = topic(b)$
- $Sub = subj(b)$
- $Val_{Set}$  :
  - o  $Val_{Set} = score(a) \setminus \{score(a_n)\} \cup score(b)$  si  $subj(b) = S$
  - o  $Val_{Set} = score(a) \setminus \{score(a_n)\}$  sinon

#### 7.4.3.10. Relation de But

### Sémantique de la relation

La relation de But relie deux segments, dont le second (b) présente de façon explicite le but, l'objectif, pour lequel l'action décrite dans le premier segment (a) est réalisée. Cette relation est coordonnante.

### Règle associée

Pour le triplet résultant  $\langle Top, Sub, Val_{Set} \rangle$ , nous avons :

- $Top$  :
  - o  $Top = mt$  si  $topic(a) = mt$  ou  $topic(b) = mt$
  - o  $Top = pt$  si  $topic(a) = pt$  ou  $topic(b) = pt$
- $Sub$ 
  - o  $Sub = subj(b)$ , si  $subj(b) = S$
  - o  $Sub = subj(a)$ , sinon
- $Val_{Set}$  :
  - o  $Val_{Set} = score(b)$  si  $topic(b) = mt$
  - o  $Val_{Set} = score(a)$  si  $topic(a) = mt$
  - o  $Val_{Set} = score(a) \cup score(b)$  sinon

### Exemple

Dans le contexte d'un commentaire de série TV :

(17) [Mais il faudra attendre la saison 5]<sub>a</sub> [pour atteindre une telle qualité.]<sub>b</sub>

Le graphe de discours associé est le suivant :

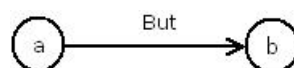


Figure 7.25 : Graphe discursif associé à l'exemple 17

Le tableau suivant présente les triplets de valeurs associés à chacun des deux segments :

	<b>17.a</b>	<b>17.b</b>
<b>topic</b>	Ind.	mt
<b>subj</b>	O (O)	S (SE)
<b>score</b>	∅	+3

Tableau 7.17 : Triplets de valeurs associés aux segments de l'exemple 17

Le triplet résultant  $\langle Top, Sub, Val\_Set \rangle$  est donc :  $\langle mt, S, \{+3\} \rangle (Int_+(+2) = +3)$ .

#### 7.4.3.11. Relation de Disjonction (Alternation)

##### Sémantique de la Relation

Cette relation coordonnante marque une disjonction entre deux segments a et b.

De par la structure de la relation, les segments partagent un sujet et un type de subjectivité commun. Nous avons donc :

- $topic(a) = topic(b) = topic$
- $subj(a) = subj(b) = subj$

##### Règle associée

La règle est strictement similaire à celle de la continuation, i.e., pour le triplet résultant  $\langle Top, Sub, Val\_Set \rangle$  :

- $Top = topic$  : le type de sujet commun
- $Sub = subj$  : la subjectivité commune
- $Val\_Set = score(a) \cup score(b)$  : les opinions portées ont même importance

##### Exemple

Dans le contexte d'un commentaire de série TV :

(18) [Les rebondissements sont soit trop prévisibles]<sub>a</sub> [soit tirés par les cheveux.]<sub>b</sub>

Le graphe de discours associé est le suivant :

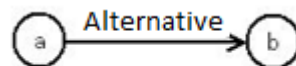


Figure 7.26 : Graphe discursif associé à l'exemple 18

Le tableau suivant présente les triplets de valeurs associés à chacun des deux segments :

	<b>18.a</b>	<b>18.b</b>
<b>topic</b>	pt (les rebondissements)	pt (les rebondissements)
<b>subj</b>	S (SE)	S (SE)
<b>score</b>	-2	-2

Tableau 7.18 : Triplets de valeurs associés aux segments de l'exemple 18



Le triplet résultant  $\langle Top, Sub, Val\_Set \rangle$  est donc :  $\langle pt, S, \{-2, -2\} \rangle$ . Comme pour la relation de Continuation, il est utile de garder les doublons de score d'opinion à des fins de pondération lors d'un recours éventuel à une heuristique de calcul numérique.

#### 7.4.3.12. Relation d'Arrière-Plan (Background)

##### Sémantique de la Relation

La relation d'Arrière-Plan est une relation où le segment cible  $b$  décrit la scène servant d'arrière-plan à l'événement décrit par le segment  $a$ .

##### Règle associée

Pour le triplet résultant  $\langle Top, Sub, Val_{Set} \rangle$ , nous avons :

- $Top = topic(b)$  :  $b$  sert à poser l'arrière-plan du discours
- $Sub = subj(b)$
- $Val_{Set}$  :
  - o  $Val_{Set} = score(b)$  si  $subj(b) = S$
  - o  $Val_{Set} = score(a) \setminus \{score(a_n)\}$

##### Exemple

Considérons l'exemple suivant, dans le contexte d'un commentaire sur un restaurant :

(19) [Je suis entré dans le restaurant.]<sub>a</sub> [Cela sentait la bonne cuisine.]<sub>b</sub>

Le graphe de discours associé est le suivant :

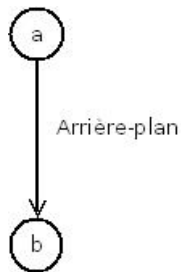


Figure 7.27 : Graphe discursif associé à l'exemple 19

Le tableau suivant présente les triplets de valeurs associés à chacun des deux segments :

	19.a	19.b
<b>topic</b>	Ind.	pt (odeur - cuisine)
<b>subj</b>	O (O)	S (SE)
<b>score</b>	$\emptyset$	{+2}

Tableau 7.19 : Triplets de valeurs associés aux segments de l'exemple 19

Le triplet résultant  $\langle Top, Sub, Val\_Set \rangle$  est donc :  $\langle pt, S, \{+2\} \rangle$ . Les valeurs de l'arrière-plan sont remontées, et contribueront à l'opinion globale de la même manière que si l'opinion du segment 19.b était exprimée dans le segment 19.a.

### 7.4.3.13. Relation de Conditionnelle

#### *Sémantique de la Relation*

Cette relation coordonnante relie deux segments où le premier (a) est une hypothèse et le deuxième (b) est la conséquence de l'hypothèse exprimée dans le segment a.

La relation de Conditionnelle est une relation complexe en ce qui concerne le calcul d'une opinion globale. Le segment a peut-être porteur d'opinion, mais dans ce cas, cette opinion n'est formulée qu'à titre d'hypothèse, comme dans l'exemple :

(20) [Si ce restaurant est bon,]<sub>a</sub> [je suis la reine d'Angleterre.]<sub>b</sub>

Par contre, une opinion exprimée dans le segment b peut être soit valide, soit invalide suivant la valeur de vérité de l'hypothèse exprimée :

(21) [Si je suis la reine d'Angleterre,]<sub>a</sub> [ce restaurant est le meilleur du monde.]<sub>b</sub>

(22) [S'il ne pleut pas,]<sub>a</sub> [la terrasse de ce restaurant est idéale.]<sub>b</sub> [Rassurez-vous,]<sub>c</sub> [il ne pleut quasiment jamais sur cette île.]<sub>d</sub><sup>56</sup>

Dans ces cas extrêmes, le segment (21.a) invalide l'opinion portée par le segment 21e.b). De même, sachant le contexte fournit par les segments (22.c) et (22.d), le segment (22.a) valide l'opinion portée par le segment (22.b).

Dans le cadre d'une automatisation du processus de remontée de l'opinion globale, nous prenons le parti d'ignorer les opinions des deux segments reliés par la conditionnelle, afin de ne pas remonter d'opinion contradictoire avec le message global du document.

#### *Règle associée*

Pour le triplet résultant  $\langle Top, Sub, Val_{set} \rangle$ , nous avons :

- $Top = topic(a)$  : conservation des informations issues de relations coordonnantes
- $Sub$  :
  - o  $Sub = subj(b)$  si  $subj(b) = S$
  - o  $Sub = subj(a)$  sinon
- $Val_{set} = score(a) \setminus \{score(a_n)\}$

#### *Exemple*

Considérons l'exemple suivant, dans le contexte d'un commentaire sur un restaurant :

(23) [Si ce restaurant est bon,]<sub>a</sub> [je suis la reine d'Angleterre.]<sub>b</sub>

---

<sup>56</sup> [Lanzarote](#) correspond bien à cette description.

Le graphe de discours associé est le suivant :

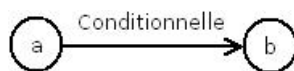


Figure 7.28 : Graphe discursif associé à l'exemple 23

Le tableau suivant présente les triplets de valeurs associés à chacun des deux segments :

	23.a	23.b
<b>topic</b>	mt (Ce restaurant)	Ind.
<b>subj</b>	S (SE)	O (O)
<b>score</b>	{+2}	$\emptyset$

Tableau 7.20 : Triplets de valeurs associés aux segments de l'exemple 23

Le triplet résultant  $\langle Top, Sub, Val\_Set \rangle$  est donc :  $\langle Ind., O, \emptyset \rangle$ . Ce triplet est clairement un choix par défaut, dans une optique de maintenir une cohérence dans le score remonté. Idéalement, un système capable d'identifier avec une précision raisonnable les assertions juste ou fausse remonterait le triplet :  $\langle mt, SI, \{-2\} \rangle$ , la conditionnelle pouvant être vue dans ce cas précis comme une négation implicite.

#### 7.4.3.14. Relation de Correction

##### Sémantique de la relation

La relation de Correction relie un segment source a présentant une éventualité qui est invalidée et corrigée par le segment cible b.

Les deux segments partagent un même sujet, mais diffèrent lorsqu'ils sont subjectifs de par l'opinion portée. L'opinion portée par le segment b est celle valide, qui vient remplacer celle portée par a.

##### Règle associée

Pour le triplet résultant  $\langle Top, Sub, Val_{set} \rangle$ , nous avons :

- $Top = topic$  : le topic commun
- $Sub$  :
  - o  $Sub = subj(b)$  si  $subj(b) = S$
  - o  $Sub = subj(a)$  sinon
- $Val_{set} = score(a) \setminus \{score(a_n)\} \cup score(b)$  : le segment b vient corriger le segment a : c'est cette valeur mise-à-jour qui est pertinente

##### Exemple

Considérons la portion de discours suivante, dans le contexte d'un commentaire de série TV :

(24) [C'est un peu stressant par moments,]<sub>a</sub> [même tout le temps,]<sub>b</sub>

Le graphe de discours associé est le suivant :

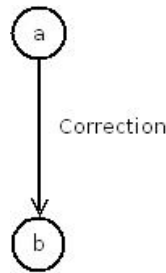


Figure 7.29 : Graphe discursif associé à l'exemple 24

Le tableau suivant présente les triplets de valeurs associés à chacun des deux segments :

	24.a	24.b
<b>topic</b>	mt	mt
<b>subj</b>	S (SE)	S (SE)
<b>score</b>	{-1}	{-2}

Tableau 7.21 : Triplets de valeurs associés aux segments de l'exemple 24

Le triplet résultant  $\langle Top, Sub, Val\_Set \rangle$  est donc :  $\langle mt, S, \{-2\} \rangle$ .

#### 7.4.3.15. Relations de Narration et de Flashback

##### Sémantique des relations

La relation de Narration indique que deux éventualités d'une même histoire sont chronologiquement ordonnés — le segment a se passe avant le segment b. À l'inverse, la relation de Flashback indique que deux éventualités d'une même histoire sont dans un ordre chronologique inverse — le segment b se passe avant le segment a.

Ces relations diffèrent de la relation de Continuation dans la mesure où l'on ne retrouve pas nécessairement un sujet commun entre les deux segments.

##### Remarque :

Dans le contexte de l'opinion, nous avons pu remarquer sur notre corpus que ces relations sont souvent accompagnées d'une relation de résultat. Par exemple, dans la portion de discours suivante :

(25) [Je viens d'aller le voir]<sub>a</sub> [et j'en suis encore toute émue.]<sub>b</sub>

Nous avons ici à la fois Narration(25.a, 25.b), et Résultat(25.a, 25.b).

##### Règle associée

Pour le triplet résultant  $\langle Top, Sub, Val_{set} \rangle$ , nous avons :

- *Top* :
  - o  $Top = mt$  si  $topic(a) = mt$  ou  $topic(b) = mt$
  - o  $Top = pt$  si  $topic(a) = pt$  ou  $topic(b) = pt$

- $Top = ot$  sinon
- $Sub$ 
  - $Sub = subj(b)$ , si  $subj(b) = S$
  - $Sub = subj(a)$ , sinon
- $Val_{Set}$  :
  - $Val_{Set} = score(a) \cup score(b)$  si  $Top = pt$  ou  $Top = ot$
  - $Val_{Set} = score(a)$  si  $topic(a) = mt$
  - $Val_{Set} = score(b)$  si  $topic(b) = mt$

### Exemples

- Considérons l'exemple suivant, illustrant le cas d'une narration, dans le contexte d'un commentaire sur un restaurant :

(26) [J'étais avec des amis,]<sub>a</sub> [et nous sommes tombés sur ce bon restaurant.]<sub>b</sub>

Le graphe de discours associé est le suivant :

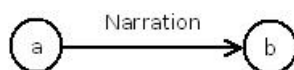


Figure 7.30 : Graphe discursif associé à l'exemple 26

Le tableau suivant présente les triplets de valeurs associés à chacun des deux segments :

	26.a	26.b
<b>topic</b>	Ind.	mt (Ce restaurant)
<b>subj</b>	O (O)	S (SE)
<b>score</b>	$\emptyset$	{+2}

Tableau 7.22 : Triplets de valeurs associés aux segments de l'exemple 26

Le triplet résultant  $\langle Top, Sub, Val\_Set \rangle$  est donc :  $\langle mt, S, \{+2\} \rangle$ .

- L'exemple suivant illustre le cas d'un flashback, dans le contexte d'un commentaire sur un restaurant :

(27) [J'ai mangé ici Lundi.]<sub>a</sub> [J'étais déjà venu manger dans ce lieu sympathique auparavant.]<sub>b</sub>

Le graphe de discours associé est le suivant :

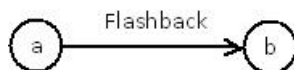


Figure 7.31 : Graphe discursif associé à l'exemple 27

Le tableau suivant présente les triplets de valeurs associés à chacun des deux segments :

	27.a	27.b
<b>topic</b>	Ind.	mt (ce lieu)
<b>subj</b>	O (O)	S (SE)
<b>score</b>	$\emptyset$	{+2}

Tableau 7.23 : Triplets de valeurs associés aux segments de l'exemple 27

Le triplet résultant  $\langle Top, Sub, Val\_Set \rangle$  est donc également :  $\langle mt, S, \{+2\} \rangle$ .

#### 7.4.4. Conclusion : quelles valeurs retenir pour le segment résultant ?

L'application de l'algorithme nous donne en sortie un segment équivalent au graphe – d'où le terme de graphe réduit. Ce segment possède trois attributs calculés :

- Une valeur unique *Top* pour le type de sujet de l'opinion
- Une valeur unique *Sub* pour le type de subjectivité
- Un ensemble de valeurs  $Val_{Set}$  représentant le ou les valeurs d'opinion du segment

L'objectif étant de remonter une unique valeur d'opinion pour le graphe en cours, plusieurs cas peuvent se présenter.

##### 7.4.4.1. Score d'opinion du segment résultant

Nous distinguons trois cas pour ce calcul :

**L'ensemble  $Val_{Set}$  est vide** : ce cas se présente si le segment résultant n'est pas porteur d'opinion. Dans ce cas, on ne peut pas remonter d'opinion pour ce graphe. S'il s'agit du graphe complet du document, cela s'interprète comme correspondant à un document ne portant pas d'opinion sur le sujet principal.

**L'ensemble  $Val_{Set}$  contient un unique élément val** : ce cas est le plus favorable, la valeur d'opinion du segment résultant est l'unique élément de la liste remontée.

**L'ensemble  $Val_{Set}$  contient N valeurs** : enfin, la liste peut contenir plusieurs éléments, remontés comme étant d'importance égale vis-à-vis de la structure discursive du document. Il est alors nécessaire d'appliquer une heuristique de calcul afin de simplifier cette liste. Dans le cas d'un discours cohérent, on peut s'attendre à ce que la liste contienne soit des valeurs similaires, soit des valeurs "proches" (i.e. dans ce contexte, de même polarité et d'intensités sensiblement similaires). Si ce n'était pas le cas – par exemple si la liste contient des valeurs d'opinion opposées – cela peut s'interpréter comme le signe d'une incohérence dans le discours, soit en raison d'une mauvaise rédaction du document, soit en raison d'une mauvaise reconnaissance automatique de la structure discursive.

##### 7.4.4.2. Type de sujet de l'opinion résultant

Le type de sujet de l'opinion résultant peut être soit "*mt*" (i.e. sujet principal du document), soit "*pt*" (i.e. sujet principal partiel), soit "*ot*" (autre sujet).

Dans le cas d'un sujet "*mt*", l'opinion résultante porte bien sur le sujet principal.

Dans le cas d'un sujet "*pt*", l'opinion résultante porte sur une partie du sujet principal. Il est donc possible que l'opinion calculée ne reflète pas la totalité du sujet principal. Dans le cadre d'une chaîne de traitement, il peut être intéressant de remonter cette incertitude à l'utilisateur final.

Enfin, dans le cas d'une opinion "ot", l'opinion résultant calculée ne porte pas sur le sujet principal. De même, cette information peut être remontée à l'utilisateur, par exemple en indiquant le graphe de discours comme non porteur d'opinion relativement au sujet principal.

## 7.4.5. Exemples complets

### 7.4.5.1. Cas d'étude

L'algorithme commence par la réduction des CDU les plus internes. Dans notre exemple (cf. section 7.2), et comme illustré par la figure 7.2, les deux CDU 13, et 14 ne possèdent pas de CDU imbriquées, et peuvent donc être réduites dans n'importe quel ordre. Par contre, la CDU 15 a parmi ses constituants la CDU imbriquée 16 : cette dernière doit donc être réduite en priorité.

#### Réduction de la CDU 13

La CDU 13 possède les constituants suivants, reliés par le graphe de discours de la figure 7.2 :

Segment	Subjectivité	Valeur d'opinion	Sujet associé <sup>57</sup>
2	SE	+3	mt (implicite)
3	SE	+2	mt (implicite)
4	SE	+3	mt (implicite)

Tableau 7.24 : Constituants de la CDU 13

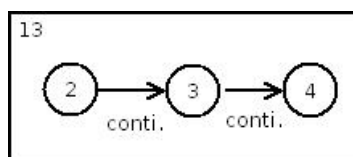


Figure 7.32 : Graphe interne à la CDU 13

Le segment racine, à partir duquel le calcul est effectué, est le segment 2.

#### Résolution des règles issues du segment 2 :

- Pas de relations subordonnantes
- Une relation coordonnante :
  - o Continuation(2,3)

Après application de la règle de Continuation, nous obtenons le segment calculé résultant {2,3}, de valeurs  $\langle mt, S, \{+3, +2\} \rangle$ .

#### Résolution des règles issues du segment 3 :

- Pas de relations subordonnantes
- Une relation coordonnante :
  - o Continuation(3,4)

<sup>57</sup> Le sujet associé est soit lexicalisé au sein d'un segment, soit implicite et inféré automatiquement. La section 6.5 décrit plus en détails l'obtention des données à partir de l'annotation.

Après application de la règle de Continuation, nous obtenons le segment calculé résultant 13 = {2,3,4}, de valeurs  $\langle mt, S, \{+3, +2, +3\} \rangle$ .

#### Réduction de la CDU 14

La CDU 14 possède les constituants et le graphe de discours suivants :

Segment	Subjectivité	Valeur d'opinion	Sujet associé
5	SE	+2	pt (mise-en-scène)
6	SE	+3	pt (mise-en-scène)
7	O	$\emptyset$	$\emptyset$

Tableau 7.25 : Constituants de la CDU 14

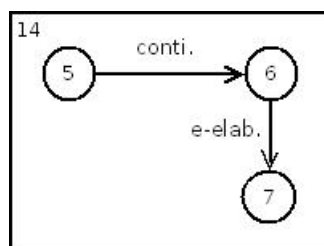


Figure 7.33 : Graphe interne à la CDU 14

Le segment racine est le segment 5.

#### Résolution des règles issues du segment 5 :

- Pas de relations subordonnantes
- Une relation coordonnante :
  - o Continuation(5,6)

Après application de la règle de Continuation, nous obtenons le segment calculé résultant {5,6}, de valeurs  $\langle pt, S, \{+2, +3\} \rangle$ .

#### Résolution des règles issues du segment 6 :

- Relation subordonnante :
- E-Elaboration(6,7)
  - o Pas de relations coordonnantes

Après application de la règle d'e-Elaboration, dans la mesure où le segment 7 n'est pas porteur d'opinion, nous obtenons le segment calculé résultant 14 = {5,6,7} suivant :  $\langle pt, S, \{+2, +3\} \rangle$ .

#### Réduction de la CDU 16

La CDU 16 possède les constituants et le graphe de discours suivants :

Segment	Subjectivité	Valeur d'opinion	Sujet associé
10	SEI	-2	pt (mise-en-scène)
11	SE	-1	pt (l'artiste)

Tableau 7.26 : Constituants de la CDU 16



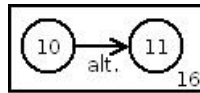


Figure 7.34 : Graphe interne à la CDU 16

Le segment racine est le segment 10.

*Résolution des règles issues du segment 10 :*

- Pas de relations subordonnantes
- Une relation coordonnante :
  - o Alternative(10,11)

Après application de la règle d'Alternative, nous obtenons le segment calculé résultant 16 = {10,11}, de valeurs  $\langle mt, S, \{+3, +2\} \rangle$ .

*Réduction de la CDU 15*

La CDU 15 possède, après réduction de la CDU 16, les constituants et le graphe de discours suivants :

Segment	Subjectivité	Valeur d'opinion	Sujet associé
9	SI	-2	mt (implicite)
16	S	{+3,+2}	pt

Tableau 7.27 : Constituants de la CDU 15

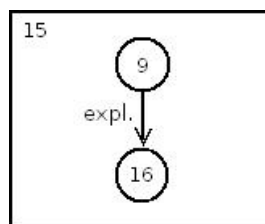


Figure 7.35 : Graphe interne à la CDU 15

Le segment racine est le segment 9.

*Résolution des règles issues du segment 9 :*

- Une relation subordonnante :
- Explication(9,16)
  - o Pas de relations coordonnantes

Après application de la règle d'Explication, nous obtenons le segment calculé résultant 15 = {9,16}, de valeurs  $\langle mt, S, \{-2\} \rangle$ . L'opinion du segment 9 prend le pas sur celle du segment complexe 16 expliquant cette opinion.

**Étape finale : résolution du graphe réduit**

Après réduction de ses constituants, le graphe résultant de discours se présente comme suit :

Segment	Subjectivité	Valeur d'opinion	Sujet associé
1	SN	$\emptyset$	$\emptyset$
8	SE	+3	pt (acteurs)
12	SE	+3	mt (implicite)
13	S	{+3;+2;+3}	mt
14	S	{+2;+3}	pt
15	S	{-2}	mt

Tableau 7.28 : Constituants du graphe

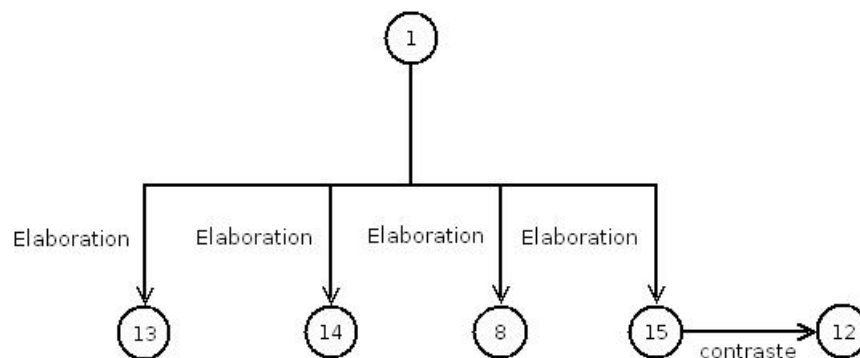


Figure 7.36 : Graphe résultant réduit

**Résolution des règles issues du segment 1 :**

- Quatre relations subordonnantes :
  - o Élaboration(1,13)
  - o Élaboration(1,14)
  - o Élaboration(1,8)
  - o Élaboration(1,15)
- Pas de relations coordonnantes

Le segment 1 étant non porteur d'opinion, l'application de la règle d'Élaboration(1,13) nous donne le segment résultant {1,13} :  $\langle mt, S, \{+3, +2, +2\} \rangle$ .

Ce segment résultant étant porteur d'opinion, le segment résultant final {1,8,12,13,14,15}, issu des relations Élaboration({1,13}, 14), Élaboration({1,13,14}, 8), et Élaboration({1,13,14,8}, {15,12}) sera également :  $\langle mt, S, \{+3, +2, +2\} \rangle$ .

À titre illustratif, la résolution de la relation de Contraste(15,12) donne le résultat {12,15}  $\langle mt, S, \{+3\} \rangle$  : le segment cible de la relation de contraste concerne le sujet principal, et est plus intense que le segment source, son opinion est donc conservée.

**Obtention de la note finale**

Le segment calculé contient une liste de valeurs d'opinion non limitée à un seul élément : la structure discursive ne permet pas de favoriser l'une ou l'autre de ces opinions. Par conséquent, il est nécessaire d'appliquer une heuristique numérique pour donner une valeur unique.

Le tableau suivant donne, pour plusieurs heuristiques parmi celles décrites en section 6.5, la valeur calculée.

Heuristique appliquée	Valeur calculée
Moyenne	+3 (+2.66, arrondi)
Médiane	+3
Intensité maximale	+3
Valeur la plus fréquente	+3
Première opinion	+3
Dernière opinion	+3
Vote (des heuristiques précédentes)	+3

Tableau 7.29 : Résultat final par heuristique

#### 7.4.5.2. Exemple complémentaire

##### Document

Le document suivant, sélectionné pour sa plus grande complexité discursive, est extrait de notre corpus de commentaires web de films et séries TV. Le sujet principal est le film *La Môme*<sup>58</sup>, qui raconte la vie d'Edith Piaf. L'annotateur a associé la note globale de -2 à ce document.

[J'avoue]<sub>1</sub> [que je suis très déçue de ce film.]<sub>2</sub> [D'abord, c'est complètement décousu,]<sub>3</sub> [et la personne [qui ne connaît pas la vie de PIAF]<sub>5</sub> aura bien du mal à se repérer.]<sub>4</sub> [On saute sans arrêt du coq à l'âne.]<sub>6</sub> [D'abord, on la voit, jeune,]<sub>7</sub> [puis vieille]<sub>8</sub> [et à nouveau jeune]<sub>9</sub> [si bien qu'on ne sait plus ou en en est.]<sub>10</sub> [Heureusement, que je connais sa vie,]<sub>11</sub> [mais je n'ai rien trouvé vraiment de cette vie.]<sub>12</sub> [On ne nous parle pas de son mari, de ses nombreux amants,]<sub>13</sub> [tel que MOUSTAKI,]<sub>14</sub> [et bien d'autres connus.]<sub>15</sub> [On nous la présente comme une véritable alcoolique...]<sub>16</sub> [Heureusement que l'actrice joue très bien son rôle.]<sub>17</sub> [Je m'attendais à mieux]<sub>18</sub>

Le tableau suivant reprend les valeurs d'opinions annotées pour chacun des segments.

Segment	Subjectivité	Valeur d'opinion	Sujet associé
1	SN	∅	∅
2	SE	-2	mt (implicite)
3	SE	-2	mt (implicite)
4	SE	-2	mt (implicite)
5	O	∅	∅
6	SE	-2	mt (implicite)
7	O	∅	∅
8	O	∅	∅
9	O	∅	∅
10	SI	-2	mt (implicite)
11	SEI	+1	pt (sa vie)
12	SI	-2	pt (sa vie)
13	SI	-2	pt (sa vie)

<sup>58</sup> [https://fr.wikipedia.org/wiki/La\\_M%C3%B4me](https://fr.wikipedia.org/wiki/La_M%C3%B4me)

14	O	∅	∅
15	O	∅	∅
16	SI	-2	pt (la [le personnage])
17	SE	+3	pt (l'actrice)
18	SE	-2	mt (implicite)

Tableau 7.30 : Valeurs de subjectivité et d'opinion pour chacun des segments

Enfin, la structure rhétorique du document a été annotée de la façon suivante :

Segments complexes :

- 19 : [1,2]
- 20 : [7,8,9]
- 21 : [14,15]
- 22 : [13,[14,15]]
- 23 : [3,4,5,6,10,11,12,16,[7,8,9],[13,[14,15]]]
- 24 : [[3,4,5,6,10,11,12,16,[7,8,9],[13,[14,15]]],17]

Relations rhétoriques :

- Attribution(1,2)
- Continuation(3,4)
- E-Elab(4,5)
- Explication(4,6)
- Narration(7,8)
- Flashback(8,9)
- Elaboration(6,[7,8,9])
- Result([7,8,9],10)
- Contrast(4,11)
- Contrast(11,12)
- Continuation(14,15)
- E-Elab(13,[14,15])
- Elaboration(12,[13,14,15])
- Elaboration(12,16)
- Contrast([3,4,6,10,11,12,16,7,8,9,13,14,15],17)
- Result([3,4,6,10,11,12,16,7,8,9,13,14,15,17],18)
- Elaboration([1,2],[3,4,6,10,11,12,16,7,8,9,13,14,15,17])

Cette structure peut être représentée graphiquement par le graphe de la figure 7.37.

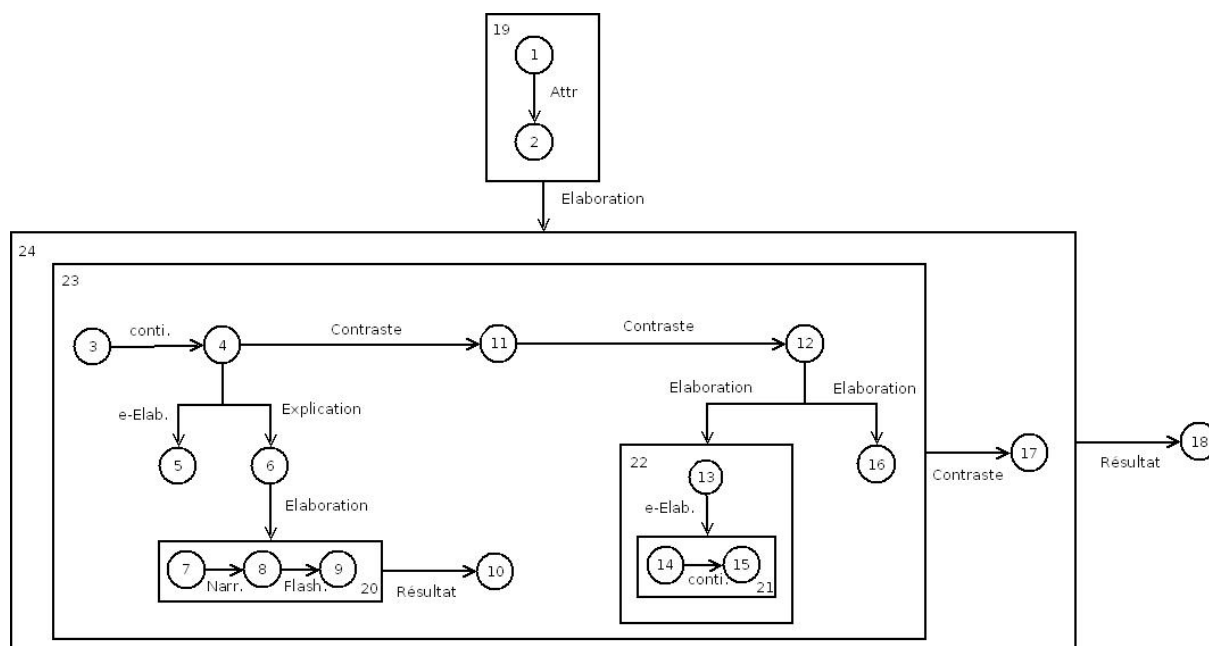


Figure 7.37 : Graphe discursif correspondant à l'exemple

### Réduction de la CDU 19

La CDU 19 possède les constituants suivants, reliés par le graphe de discours de la figure 7.37 :

Segment	Subjectivité	Valeur d'opinion	Sujet associé
1	SN	$\emptyset$	$\emptyset$
2	SE	-2	mt (implicite)

Tableau 7.31 : Constituants de la CDU 19

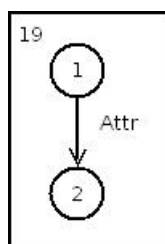


Figure 7.38 : Graphe interne à la CDU 19

Le segment racine, à partir duquel le calcul est effectué, est le segment 1.

### Résolution des règles issues du segment 1 :

- Une relation subordonnante :
- Attribution(1,2)
  - o Pas de relation coordonnante

Après application de la règle d'Attribution, nous obtenons le segment calculé résultant  $19 = \{1,2\}$ , de valeurs  $\langle mt, S, \{-2\} \rangle$ .

**Réduction de la CDU 20**

La CDU 20 possède les constituants suivants, reliés par le graphe de discours de la figure 7.37 :

Segment	Subjectivité	Valeur d'opinion	Sujet associé
7	O	∅	∅
8	O	∅	∅
9	O	∅	∅

Tableau 7.32 : Constituants de la CDU 20



Figure 7.39 : Graphe interne à la CDU 20

Le segment racine, à partir duquel le calcul est effectué, est le segment 7.

**Résolution des règles issues du segment 7 :**

- Pas de relation subordonnante
- Une relation coordonnante :
  - o Narration(7,8)

Après application de la règle d'Attribution, nous obtenons le segment calculé résultant {7,8}, de valeurs  $\langle \emptyset, O, \emptyset \rangle$ .

**Résolution des règles issues du segment 8 :**

- Pas de relation subordonnante
- Une relation coordonnante :
  - o Flashback(8,9)

Après application de la règle d'Attribution, nous obtenons le segment calculé résultant  $20 = \{7,8,9\}$ , de valeurs  $\langle \emptyset, O, \emptyset \rangle$ .

**Réduction de la CDU 21**

La CDU 21 possède les constituants suivants, reliés par le graphe de discours de la figure 7.37 :

Segment	Subjectivité	Valeur d'opinion	Sujet associé
14	O	∅	∅
15	O	∅	∅

Tableau 7.33 : Constituants de la CDU 21

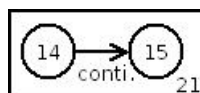


Figure 7.40 : Graphe interne à la CDU 21

Le segment racine, à partir duquel le calcul est effectué, est le segment 14.

**Résolution des règles issues du segment 14 :**

- Pas de relation subordonnante

- Une relation coordonnante :
  - o Continuation(14,15)

Après application de la règle de continuation, nous obtenons le segment calculé résultant 21 = {14,15}, de valeurs  $\langle \emptyset, 0, \emptyset \rangle$ .

### Réduction de la CDU 22

La CDU 22 possède les constituants suivants, reliés par le graphe de discours de la figure 7.37 :

Segment	Subjectivité	Valeur d'opinion	Sujet associé
13	SI	-2	pt (sa vie)
21	O	$\emptyset$	$\emptyset$

Tableau 7.34 : Constituants de la CDU 22

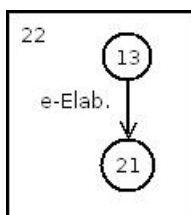


Figure 7.41 : Graphe interne à la CDU 22

Le segment racine, à partir duquel le calcul est effectué, est le segment 13.

### Résolution des règles issues du segment 13 :

- Une relation subordonnante :
  - o E-Elaboration(13,21)
- Pas de relation coordonnante

Après application de la règle d'Elaboration d'entités, nous obtenons le segment calculé résultant 22 = {13,14,15}, de valeurs  $\langle \emptyset, 0, \emptyset \rangle$ .

### Réduction de la CDU 23

La CDU 23 possède les constituants suivants, reliés par le graphe de discours de la figure 7.37 :

Segment	Subjectivité	Valeur d'opinion	Sujet associé
3	SE	-2	mt (implicite)
4	SE	-2	mt (implicite)
5	O	$\emptyset$	$\emptyset$
6	SE	-2	mt (implicite)
10	SI	-2	mt (implicite)
11	SEI	+1	pt (sa vie)
12	SI	-2	pt (sa vie)
16	SI	-2	pt (la [le personnage])
20	O	$\emptyset$	$\emptyset$
22	O	$\emptyset$	$\emptyset$

Tableau 7.35 : Constituants de la CDU 23

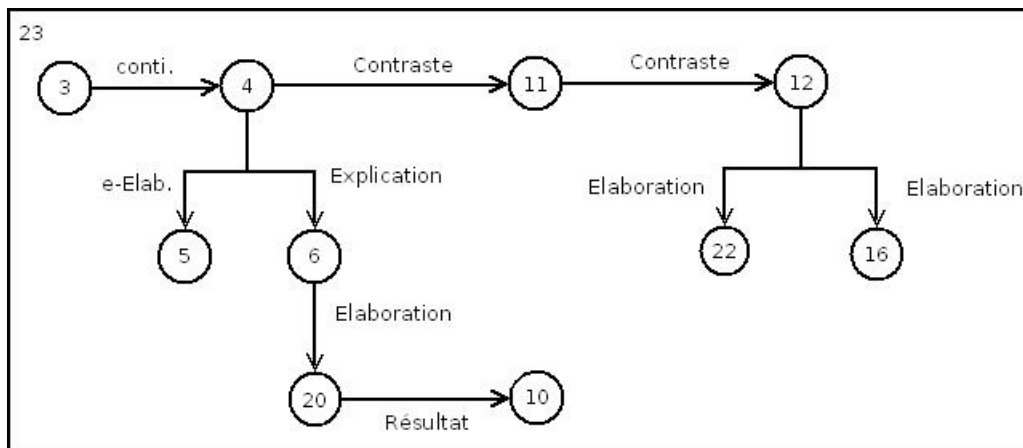


Figure 7.42 : Graphe interne à la CDU 23

Le segment racine, à partir duquel le calcul est effectué, est le segment 3. Le parcours en profondeur implique l'ordre suivant de résolution des nœuds : 3, 20, 6, 4, 11, 12.

*Résolution des règles issues du segment 3 :*

- Pas de relation subordonnante
- Une relation coordonnante
  - o Continuation(3,4)

Après application de la règle de Continuation, nous obtenons le segment calculé résultant {3,4}, de valeurs  $\langle mt, S, \{-2, -2\} \rangle$ .

*Résolution des règles issues du segment 20 :*

- Pas de relation subordonnante
- Une relation coordonnante
  - o Résultat(20,10)

Après application de la règle de Résultat et comme le segment complexe 20 n'est pas porteur d'opinion, nous obtenons le segment calculé résultant {20,10}, de valeurs  $\langle mt, S, \{-2\} \rangle$ .

*Résolution des règles issues du segment 6 :*

- Une relation subordonnante
  - o Elaboration(6,20)
- Pas de relation coordonnante

Le segment 6 étant porteur d'opinion, le segment résultant {6,20,10} est donc associé aux valeurs  $\langle mt, S, \{-2\} \rangle$ .

*Résolution des règles issues du segment 4 :*

- Deux relations subordonnantes
  - o E-Elaboration(4,5)
  - o Explication(4,6)
- Une relation coordonnante
  - o Contraste(4,11)



Nous commençons par appliquer la relation d'Elaboration d'entités entre {3,4} et 5. Le segment résultant {3,4,5} est le suivant :  $\langle mt, S, \{-2, -2\} \rangle$ . En effet, 5 n'est pas porteur d'opinion.

Nous appliquons ensuite la règle associée à l'Explication entre 4 et 6 pour obtenir {3,4,5,6,20,10} :  $\langle mt, S, \{-2, -2\} \rangle$ . Les valeurs sont là encore héritées de la source, puisque celle-ci est porteuse d'opinion.

Enfin, nous appliquons la règle associée au Contraste. Le segment calculé {3,4,5,6,20,10,11} résultant est le suivant :  $\langle mt, S, \{-1, -1\} \rangle$ , car l'intensité de l'opinion portée par le segment cible 11 est plus faible que celle portée par le segment source.

#### Résolution des règles issues du segment 11 :

- Pas de relation subordonnante
- Une relation coordonnante
  - o Contraste(11,12)

L'application de la règle de Contraste nous donne le segment calculé {3,4,5,6,20,10,11,12}, de valeurs :  $\langle mt, S, \{-2\} \rangle$  : le segment cible 12 étant plus intense que le segment source.

#### Résolution des règles issues du segment 12 :

- Deux relations subordonnantes
  - o Elaboration(12,22)
  - o Elaboration(12,16)
- Pas de relation coordonnante

Le segment source étant porteur d'opinion, l'application successive de la règle associée aux deux relations d'Elaboration donnera le segment calculé 23 = {3,4,5,6,20,10,11,12,22,16}, de valeurs :  $\langle mt, S, \{-2\} \rangle$ .

#### Réduction de la CDU 24

La CDU 24 possède les constituants suivants, reliés par le graphe de discours de la figure 7.37 :

Segment	Subjectivité	Valeur d'opinion	Sujet associé <sup>59</sup>
23	S	{-2}	mt
17	SE	+3	pt (l'actrice)

Tableau 7.36 : Constituants de la CDU 24

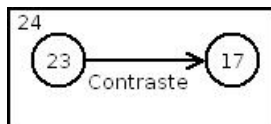


Figure 7.43 : Graphe interne à la CDU 24

Le segment racine, à partir duquel le calcul est effectué, est le segment 23.

<sup>59</sup> Le sujet associé est soit lexicalisé au sein d'un segment, soit implicite et inféré automatiquement. La section 6.5 décrit plus en détails l'obtention des données à partir de l'annotation.

*Résolution des règles issues du segment 12 :*

- Pas de relation subordonnante
- Une relation coordonnante
  - o Contraste(23,17)

Le sujet de l'opinion du segment source est le sujet principal, tandis que celui du segment cible est un sujet partiel. Par conséquent, l'application de la règle de Contraste conduit au segment calculé  $24 = \{23,17\}$  de valeurs  $\langle mt, S, \{-1\} \rangle$ .

*Étape finale : résolution du graphe réduit*

Après réduction de ses constituants, le graphe résultant de discours se présente comme suit :

Segment	Subjectivité	Valeur d'opinion	Sujet associé <sup>60</sup>
19	S	$\{-2\}$	mt
24	S	$\{-1\}$	mt
18	SE	-2	mt (implicite)

Tableau 7.37 : Constituants du graphe

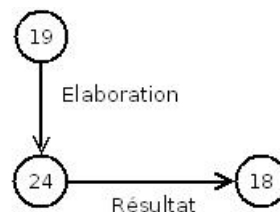


Figure 7.44 : Graphe réduit résultant

Le segment 19 constitue le nœud racine du graphe réduit. Le parcours en profondeur donne la résolution des nœuds dans l'ordre suivant : 24, 19 (les nœuds non sources de relation sont ignorés).

*Résolution des règles issues du segment 24 :*

- Pas de relation subordonnante
- Une relation coordonnante
  - o Résultat(24,18)

L'application de la règle de Résultat donne le segment calculé résultant  $\{24,18\}$ , de valeurs associées :  $\langle mt, S, \{-2\} \rangle$ .

*Résolution des règles issues du segment 19 :*

- Une relation subordonnante
  - o Elaboration(19,24)
- Pas de relation coordonnante

<sup>60</sup> Le sujet associé est soit lexicalisé au sein d'un segment, soit implicite et inféré automatiquement. La section 6.5 décrit plus en détails l'obtention des données à partir de l'annotation.

Le segment source 19 étant porteur d'opinion, le segment calculé  $\{19,24,18\}$  est donc associé aux valeurs :  $\langle mt, S, \{-2\} \rangle$ .

#### *Obtention de la note finale*

Le segment calculé contient une liste de valeurs d'opinion limitée à un seul élément : la note finale calculée est donc égale à cet unique élément, c'est-à-dire -2. Elle est en accord avec la note donnée par l'annotateur pour ce document.

## 7.5. Heuristiques de calcul de l'opinion globale

Les heuristiques de fusion que nous avons évaluées peuvent être classées en trois catégories :

- les heuristiques numériques qui se basent sur l'ensemble des segments porteurs d'opinion (développées en section 3.2.1).
- les heuristiques macro-discursives et discursives, qui se basent sur la structure discursive du texte (développées en section 3.2.2).
- les méta-heuristiques qui combinent plusieurs autres heuristiques pour déterminer l'opinion globale (développées en section 4.2.3).

### 7.5.1. Heuristiques numériques

Les heuristiques numériques se basent sur l'ensemble des segments porteurs d'opinion (i.e. de subjectivité SE, SI, ou SEI), en tant que sac d'opinions non ordonné.

#### 7.5.1.1. Moyenne des valeurs discrètes d'opinion

Ce modèle consiste à prendre directement la moyenne des valeurs renseignées par l'annotateur, sans normalisation préalable.

La formule est la suivante, pour un ensemble de segments considéré  $Seg$  de cardinal  $N$  :

$$Score_{Doc} = \frac{1}{N} \sum_{i=1}^N ValOp_{seg_i}$$

#### **Exemple - cas d'étude<sup>61</sup> :**

Sur l'exemple - cas d'étude, sans filtrage, l'application de l'heuristique conduit à la valeur résultante suivante (sur l'échelle de données  $[-3; +3] \cap \mathbb{Z}$ ):

$$Score_{Doc} = Round\left(\frac{(3 + 2 + 3 + 2 + 3 + 3 - 2 - 2 - 1 + 3)}{10}\right) = Round(+1.4) = +1$$

---

<sup>61</sup> Cf. section 7.2

### 7.5.1.2. Valeur d'opinion médiane

Ce modèle consiste à prendre la médiane de l'ensemble des opinions considérées. Contrairement à la moyenne, la médiane tend à ignorer les valeurs extrêmes aberrantes.

Pour une liste d'opinion  $Seg$  de cardinal  $N$  ordonnée suivant les valeurs croissantes, le score remonté est le suivant :

$$Score_{Doc} = \begin{cases} \text{si } N \text{ pair, } ValOp_{seg_i} \text{ avec } i = \frac{N}{2} \\ \text{sinon, } \frac{ValOp_{seg_i} + ValOp_{seg_{i+1}}}{2} \text{ avec } i = \frac{N}{2} \end{cases}$$

#### Exemple - cas d'étude :

Sur l'exemple - cas d'étude, sans filtrage, la liste ordonnée par valeur croissante est la suivante :  $Seg = [-1, -2, -2, +2, +2, +3, +3, +3, +3, +3]$ . La valeur calculée est donc la suivante :

$$Score_{Doc} = Round\left(\frac{+3 + 2}{2}\right) = +3$$

### 7.5.1.3. Sélection de l'intensité maximale

Ce modèle consiste à sélectionner l'intensité maximale parmi les opinions des segments considérés. Par convention, dans le cas – non rare – où l'ensemble de segments comporterait deux opinions de même intensité maximale et de polarité opposée, on sélectionne la polarité disposant du plus grand nombre d'instance parmi les segments. Dans le cas où l'on ait à nouveau égalité, on sélectionne (arbitrairement – nos données sont majoritairement positives) la polarité positive.

#### Exemple - cas d'étude :

Sur l'exemple - cas d'étude, l'intensité maximale est de 3, pour une polarité positive. Il n'y a pas d'opinion négative de même intensité (échelle  $[-3; +3] \cap \mathbb{Z}$ ).

$$Score_{Doc} = +3$$

### 7.5.1.4. Sélection de l'intensité maximale si polarité non mixte, moyenne sinon

Cette heuristique reprend l'idée de l'heuristique précédente de sélection de l'intensité maximale (section 2.1.4) dans le cas où les segments considérés sont tous de même polarité (polarité non-mixte). Dans le cas contraire, on applique une heuristique de moyenne.

#### Exemple - cas d'étude :

Sur l'exemple – cas d'étude, l'intensité maximale est de 3, pour une polarité positive. Il y a néanmoins deux opinions identifiées comme négatives : on applique donc les heuristiques de moyenne. On obtient donc le score suivant, calculé précédemment :

$$Score_{Doc} = +1.5 \text{ (sur } [-3; +3] \cap \mathbb{Z} \text{)}$$

### 7.5.1.5. Sélection de la valeur d'opinion la plus fréquente

Cette heuristique consiste à sélectionner la valeur d'opinion la plus fréquente parmi les valeurs d'opinion de l'ensemble de segments considéré.

Dans le cas où plusieurs valeurs d'opinion arriveraient à égalité, l'opinion choisie est celle comportant le plus d'opinions de valeurs "proches". Le tableau 7.38 suivant explicite quelles sont les valeurs d'opinions proches pour chaque valeur discrète de notre échelle.

Valeur d'opinion discrète (annotation)	Valeurs considérées comme proche
-3	{-2}
-2	{-3;-1}
-1	{-2;+1}
+1	{-1;+2}
+2	{+1;+3}
+3	{+2}

Tableau 7.38 : Valeurs d'opinion considérées comme proches d'une valeur annotée donnée

Enfin, dans le cas où l'on aurait encore une égalité, on sélectionne, arbitrairement, l'intensité la plus forte, et si besoin la polarité positive.

#### Exemple – cas d'étude :

Sur l'exemple – cas d'étude, on observe la répartition en fréquence suivante des valeurs d'opinion :

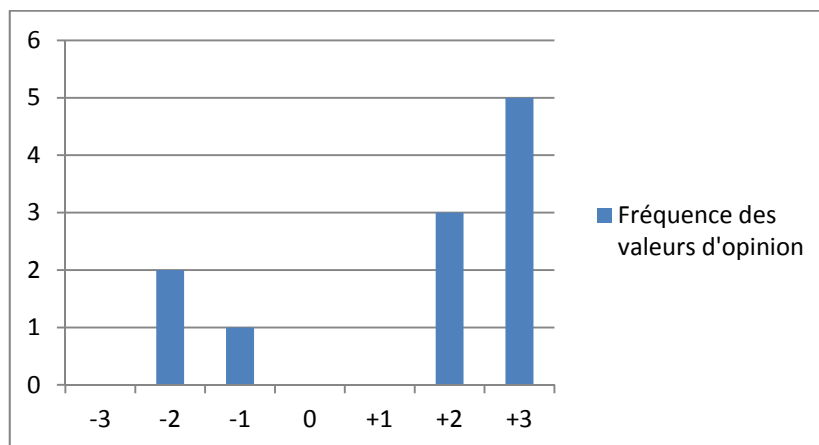


Figure 7.45 : Fréquence d'apparition de chaque valeur d'opinion dans l'ensemble des segments

La valeur d'opinion la plus fréquente est donc, sur une échelle  $[-3; +3] \cap \mathbb{Z}$  :

$$Score_{Doc} = +3$$

**Exemple complémentaire :** Considérons le texte suivant :

[Cette série est mauvaise.]\_1 [Les acteurs sont nullissimes,]\_2 [le scénario inexistant,]\_3 [les décors complètement ridicules...]\_4 [Bref, je ne recommande pas.]\_5

Le tableau suivant présente les valeurs d'opinion pour chacun des segments.

Segment	Valeur d'opinion
1	-2
2	-3
3	-2
4	-3
5	-1

Tableau 7.39 : Exemple complémentaire : valeurs d'opinion des segments

Pour cet exemple, la répartition en fréquences des valeurs d'opinion montre une ambiguïté entre les valeurs -2 et -3. Afin de résoudre cela, nous nous basons sur les valeurs proches telles que définies dans le Tableau 7.38.

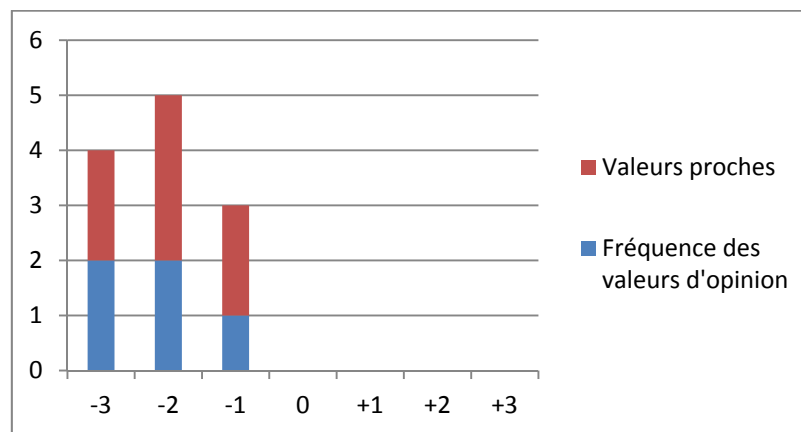


Figure 7.46 : Histogramme des fréquences des valeurs d'opinion du document, et valeurs proches cumulées

Au final, nous obtenons le score :  $Score_{Doc} = -2$ .

## 7.5.2. Heuristiques macro-discursives et discursives

Les heuristiques discursives se basent sur la structure discursive du texte pour remonter une valeur d'opinion globale, soit à partir de la structure macroscopique du discours, i.e. la position des segments dans le texte (heuristiques macro-discursives), soit sur le graphe du discours.

### 7.5.2.1. Sélection de l'opinion du premier segment subjectif

Cette heuristique consiste à attribuer au texte l'opinion portée par le premier segment porteur d'opinion de l'ensemble considéré.

#### Remarque :

Suivant les filtres utilisés en amont (cf. section 4 : stratégies avec élagage), ce segment peut ne pas être le premier segment du texte, et ce même si celui-ci est subjectif.

**Exemple – cas d'étude :**

Sur l'exemple – cas d'étude, le premier segment porteur d'opinion est le segment 2, de valeur d'opinion +3 sur une échelle  $[-3; +3] \cap \mathbb{Z}$ .

$$Score_{Doc} = +3$$

**7.5.2.2. Sélection de l'opinion du dernier segment subjectif**

Cette heuristique consiste à attribuer au texte l'opinion portée par le dernier segment porteur d'opinion de l'ensemble considéré.

**Remarque :**

Comme précédemment, suivant les filtres utilisés en amont, ce segment peut ne pas être le dernier segment du texte, même si celui-ci est subjectif.

**Exemple – cas d'étude :**

Sur l'exemple – cas d'étude, le dernier segment porteur d'opinion est le segment 12, de valeur d'opinion +3 sur une échelle  $[-3; +3] \cap \mathbb{Z}$ .

$$Score_{Doc} = +3$$

**7.6. Résultats et analyse****7.6.1. Données expérimentales**

Les données expérimentales utilisées sont celles annotées dans le cadre du projet CASOAR (cf. chapitre 2). Le premier corpus de données utilisé pour ces expérimentations est le corpus *gold* de commentaires web CASOAR. Le second est le corpus *gold* CASOAR des annotations de réactions d'internautes aux articles de presse du site [www.lemonde.fr](http://www.lemonde.fr).

Le corpus de commentaires web utilisé est constitué de 153 documents. Chaque document est en moyenne constitué de 12 segments élémentaires de discours (UDE ou EDU), groupés en, en moyenne, 13 segments complexes (CDU). 67 % des segments élémentaires subjectifs de ce corpus sont explicites ou mixtes. Concernant les sujets de l'opinion annotés, sur 1386 instances au total, 364 concernent le sujet principal, 868 un sujet partiel, et 154 hors-sujets.

Le corpus de réactions de presse est quant à lui constitué de 130 documents. Chaque document est constitué d'en moyenne 9 segments élémentaires, groupés en 10 segments complexes. Contrairement au corpus de commentaires web, seul 28 % des segments subjectifs de corpus sont explicites ou mixtes : les opinions y sont donc majoritairement implicites. Pour les sujets de l'opinion, sur 586 annotations, seules 33 concernent le sujet principal. 349 concernent un sujet partiel, et 204 sont hors-sujet.

### 7.6.1.1. Échelles d'annotations des données

Les données ont été annotées en utilisant l'échelle d'intensité et de polarité introduite au chapitre 1, tant au niveau du segment qu'au niveau du document. Ainsi, chaque segment porteur d'opinion et chaque document possèdent une annotation d'opinion sur une échelle discrète [-3;+3].

### 7.6.1.2. Généralisation des liens topic-opinion et holder-opinion

En raison de certaines difficultés rencontrées par les annotateurs avec l'outil d'annotation, le lien entre l'opinion et son sujet ou son émetteur n'est pas présent de manière fiable dans la totalité des données CASOAR brutes. Par conséquent, les liens "gold" entre les expressions d'opinion et leurs sujets associés ont été manuellement rajoutés pour ces expérimentations.

Cette reprise d'annotation s'est effectuée en ajoutant manuellement aux documents annotés une relation topic(X, Y) à deux arguments, qui peuvent être :

- Un identifiant d'expression d'opinion repéré par l'annotateur, et un identifiant de sujet de l'opinion repéré par l'annotateur
- Un identifiant d'expression d'opinion repéré par l'annotateur, et un type de sujet d'opinion associé (main topic, part-of topic, other topic), dans le cas où le sujet de l'opinion est absent de l'annotation, et n'a donc pas d'identifiant associé

Les expressions d'opinion, leurs valeurs associées, et les sujets d'opinion annotés n'ont pas été altérés lors de l'établissement de ce lien.

#### Lien segment-sujet de l'opinion

Ces liens entre expression d'opinion et sujet ont ensuite été généralisés en liens entre un segment subjectif et le type du sujet de l'opinion associé, en se basant sur l'heuristique de rattachement suivante :

Pour les segments subjectifs explicites (SE) :

- Si l'UDE ne comporte qu'une seule expression d'opinion, alors le type du topic de l'UDE est celui de cette expression d'opinion
- Si l'UDE comporte plusieurs opinions, alors le type du sujet est :
  - o « *main-topic* » si au moins une des expressions d'opinion est rattachée au sujet principal
  - o « *part-topic* » si aucune opinion n'est rattachée au sujet principal, mais au moins une des expressions d'opinion est rattachée à un sujet partiel
  - o « *other-topic* » si les expressions d'opinion sont toutes hors-sujet relativement au sujet principal du document

Pour les segments subjectifs implicites et mixtes (SI, SEI), le sujet de l'opinion est supposé être le sujet principal (type « *main-topic* »). L'idée derrière ce choix est que si une opinion est donnée implicitement, son sujet a de fortes chances d'être lui aussi implicite, et donc d'être le sujet principal. Cette hypothèse est assez forte, et mériterait éventuellement d'être étudiée plus en détail : néanmoins, elle nécessiterait une ré-annotation complète des sujet de l'opinion au niveau du segment.



### 7.6.2. Métriques d'évaluation

Afin d'évaluer si les sorties du modèle sont correctes vis-à-vis des données annotées, nous avons décidé d'employer deux métriques complémentaires : une mesure de l'accuracy globale des résultats (nombre de documents correctement classifiés / nombre total de documents), et une mesure de la corrélation de Pearson entre l'ensemble des valeurs d'opinion calculées et l'ensemble des notes attribuées par les annotateurs.

L'accuracy nous permet de capturer l'adéquation exacte d'une stratégie donnée avec les données annotées. Cette mesure est très sensible à un décalage au niveau de l'intensité renvoyé par un modèle, mais ne permet pas nécessairement de capter si certains documents se voient affecter des valeurs d'opinion aberrantes.

Le coefficient de corrélation de Pearson permet quant à lui de capturer l'adéquation globale d'un modèle avec les données : le coefficient sera élevé si le score calculé augmente lorsque la note attribuée manuellement augmente, et diminue lorsque la note diminue. Par contre, il ne permet pas de capter un décalage global constant d'intensité dans les sorties.

### 7.6.3. Résultats observés

Nous présentons les résultats pour deux tâches : tout d'abord, la reconnaissance de la polarité du document, puis dans une seconde partie, la reconnaissance de la polarité et de l'intensité du document.

Nous présentons tout d'abord les meilleurs résultats obtenus pour chaque heuristique de composition, avec le ou les filtres associés, ou la mention (*[rules]*) si applicable. Les résultats sont groupés par métrique, avec ou sans méthode de relaxation. Le coefficient de Pearson n'est présenté qu'avec relaxation : en effet, sans relaxation, le coefficient se retrouve calculé à partir de vecteurs partiels, et sa valeur est donc parfois artificiellement élevée, ce qui empêche l'interprétation correcte des résultats.

Ensuite, pour chaque métrique d'évaluation, nous comparons également les valeurs obtenues par :

- la baseline (i.e. par une approche sans discours ni filtrage, ne considérant par conséquent que les segments explicites)
- une approche sans discours, mais avec filtrage par type de sujet et par type de subjectivité (approche dite "*Sac de segments*") (cf. sections 7.3.1 et 7.3.2)
- les différentes approches de filtrage par discours (cf. section 7.3.4)
- l'approche par règles de remontée de l'opinion (cf. section 7.4)

Les premiers résultats nous permettent de comparer les différentes métriques, et de voir quels filtres/règles sont le mieux adaptées vis-à-vis de ces métriques. Les seconds nous permettent d'évaluer plus précisément l'apport de chaque ajout de complexité dans la méthode de fusion, et en particulier l'apport du discours.

Les résultats seront analysés dans la partie discussion, en section 7.6.4.

### 7.6.3.1. Repérage de la polarité du document

#### Corpus Reviews

Les tableaux suivants présentent les meilleurs résultats obtenus pour la reconnaissance de la polarité seule, sur le corpus de commentaires web.

Meilleur score: accuracy (sans relaxation)				
Modèle :	Filter: discourse	Filter: subjectivity	Filter: topic type	Accuracy: sans rel.
Average	keep_all	keep_all	keep_MT_PT	0.915
Mediane	keep_all	keep_all	keep_MT	0.902
Max Intensity	keep_all	keep_all	keep_all	0.908
Max Intensity or Average	[rules]			0.876
Max Frequency	keep_all	keep_all	keep_all	<b>0.922</b>
First Opinion	[rules]			0.902
Last Opinion	[rules]			0.895

Tableau 7.40 : Meilleurs résultats (accuracy), polarité seule, pour chaque heuristique de calcul, filtrages effectués sans relaxation

Meilleur score: accuracy (avec relaxation)				
Modèle :	Filter: discourse	Filter: subjectivity	Filter: topic type	Accuracy: avec relaxation
Average	top_simple	keep_all	keep_MT_PT	<b>0.954</b>
Mediane	top_simple	keep_all	keep_MT	0.948
Max Intensity	top_simple	keep_SI_SEI	keep_PT	0.922
Max Intensity or Average	[rules]			0.876
Max Frequency	top_simple	keep_all	keep_PT	0.961
First Opinion	[rules]			0.902
Last Opinion	top_simple	keep_all	keep_MT_PT	0.902

Tableau 7.41 : Meilleurs résultats (accuracy), polarité seule, pour chaque heuristique de calcul, filtrages effectués avec relaxation

Meilleur score: Pearson (avec relaxation)				
Modèle :	Filter: discourse	Filter: subjectivity	Filter: topic type	Pearson: with backup
Average	top_simple	keep_all	keep_MT_PT	0.913
Mediane	top_simple	keep_all	keep_MT	0.902
Max Intensity	top_simple	keep_SI_SEI	keep_PT	0.844
Max Intensity or Average	[rules]			0.862
Max Frequency	top_simple	keep_all	keep_PT	<b>0.927</b>
First Opinion	[rules]			0.824
Last Opinion	top_simple	keep_all	keep_MT_PT	0.827

Tableau 7.42 : Meilleurs résultats (Pearson), polarité seule, pour chaque heuristique de calcul, filtrages effectués avec relaxation

Le tableau suivant présente l'apport de chaque brique à la reconnaissance de la polarité, pour le corpus de commentaires web. Les métriques sont calculées avec application de la relaxation. Pour chaque score, nous précisons entre parenthèses les filtres appliqués ainsi que l'heuristique de composition employée.

	Accuracy	Pearson
<b>Baseline</b>	0.89 (Moyenne)	0.81 (Moyenne)
<b>Sac de segments</b>	0.92 (keep_all, keep_all, MaxFreq)	0.87 (keep_all, keep_all, Moyenne)
<b>Filtres discursifs</b>	0.96	0.94
- Filtre Sub_src	0.91 (keep_all, keep_all, Médiane)	0.84 (keep_all, keep_all, Médiane)
- Filtre Sub_src_not_Dest	0.91 (keep_all, keep_all, Médiane)	0.84 (keep_all, keep_all, Médiane)
- Filtre Top_simple	0.96 (keep_all, keep_all, MaxFreq)	0.94 (keep_all, keep_all, Médiane)
- Filtre Top_récurif	0.90 (keep_all, keep_all, Moyenne)	0.80 (keep_all, keep_all, Moyenne)
<b>Règles de remontée</b>	0.90 (MaxScore)	0.86 (MaxScore_Moyenne)

Tableau 7.43 : Apport des différents modules de fusion, corpus de commentaires web, polarité seule

**Corpus LeMonde**

Les tableaux suivants présentent les meilleurs résultats obtenus pour la reconnaissance de la polarité seule, sur le corpus de réactions aux articles de presse.

<b>Meilleur score: accuracy (sans relaxation)</b>				
<b>Modèle :</b>	<b>Filter:</b> <b>discourse</b>	<b>Filter:</b> <b>subjectivity</b>	<b>Filter:</b> <b>topic type</b>	<b>Accuracy:</b> <b>sans rel.</b>
<b>Average</b>	[rules]			<b>0.946</b>
<b>Mediane</b>	[rules]			<b>0.946</b>
<b>Max Intensity</b>	[rules]			<b>0.946</b>
<b>Max Intensity or Average</b>	[rules]			0.935
<b>Max Frequeny</b>	[rules]			<b>0.946</b>
<b>First Opinion</b>	[rules]			<b>0.946</b>
<b>Last Opinion</b>	[rules]			<b>0.946</b>

*Tableau 7.44 : Meilleurs résultats (accuracy), polarité seule, pour chaque heuristique de calcul, filtrages effectués sans relaxation*

<b>Meilleur score: accuracy (avec relaxation)</b>				
<b>Modèle :</b>	<b>Filter:</b> <b>discourse</b>	<b>Filter:</b> <b>subjectivity</b>	<b>Filter:</b> <b>topic type</b>	<b>Accuracy:</b> <b>avec relaxation</b>
<b>Average</b>	[rules]			0.946
<b>Mediane</b>	[rules]			0.946
<b>Max Intensity</b>	top_rec	keep_all	keep_all	<b>0.957</b>
<b>Max Intensity or Average</b>	[rules]			0.935
<b>Max Frequeny</b>	[rules]			0.946
<b>First Opinion</b>	[rules]			0.946
<b>Last Opinion</b>	[rules]			0.946

*Tableau 7.45 : Meilleurs résultats (accuracy), polarité seule, pour chaque heuristique de calcul, filtrages effectués avec relaxation*

Meilleur score: Pearson (avec relaxation)				
Modèle :	Filter: discourse	Filter: subjectivity	Filter: topic type	Pearson: avec relaxation
Average	[rules]			<b>0.824</b>
Mediane	[rules]			<b>0.824</b>
Max Intensity	top_rec	keep_all	keep_all	0.821
Max Intensity or Average	[rules]			<b>0.824</b>
Max Frequeny	[rules]			0.802
First Opinion	[rules]			0.802
Last Opinion	[rules]			0.784

Tableau 7.46 : Meilleurs résultats (Pearson), polarité seule, pour chaque heuristique de calcul, filtres effectués avec relaxation

Le tableau suivant présente l'apport de chaque brique à la reconnaissance de la polarité, pour le corpus de réactions aux articles de presse. Les métriques sont calculées avec application de la relaxation. Pour chaque score, nous précisons entre parenthèses les filtres appliqués et/ou l'heuristique de composition employée.

	Accuracy	Pearson
<b>Baseline</b>	0.88 (MaxScore)	0.52 (MaxScore)
<b>Sac de segments</b>	0.94 (keep_all, keep_all, MaxScore)	0.77 (keep_all, keep_all, MaxScore)
<b>Filtres discursifs</b>	0.96	0.82
- Filtre Sub_src	0.92 (keep_all, keep_all, Moyenne)	0.74 (keep_all, keep_all, Moyenne)
- Filtre Sub_src_not_Dest	0.92 (keep_all, keep_all, Moyenne)	0.74 (keep_all, keep_all, Moyenne)
- Filtre Top_simple	0.92 (keep_all, keep_all, Moyenne)	0.77 (keep_all, keep_all, MaxScore)
- Filtre Top_récurif	0.96 (keep_all, keep_all, MaxScore)	0.82 (keep_all, keep_all, MaxScore)
<b>Règles de remontée</b>	0.90 (MaxFreq)	<b>0.82</b> (MaxScore_Moyenne)

Tableau 7.47 : Apport des différents modules de fusion, corpus de réactions aux articles de presse, polarité seule

### 7.6.3.2. Repérage de la polarité et de l'intensité du document

#### Corpus Reviews

Les tableaux suivants présentent les meilleurs résultats obtenus pour la reconnaissance de l'intensité et de la polarité, sur le corpus de commentaires web.

Meilleur score: accuracy (sans relaxation)				
Modèle :	Filter: discourse	Filter: subjectivity	Filter: topic type	Accuracy: sans rel.
Average	[rules]			0.745
Mediane	[rules]			0.719
Max Intensity	[rules]			<b>0.752</b>
Max Intensity or Average	[rules]			0.732
Max Frequeny	[rules]			0.725
First Opinion	[rules]			0.739
Last Opinion	[rules]			0.725

Tableau 7.48 : Meilleurs résultats (accuracy) pour chaque heuristique de calcul, filtrages effectués sans relaxation

Meilleur score: accuracy (avec relaxation)				
Modèle :	Filter: discourse	Filter: subjectivity	Filter: topic type	Accuracy: avec relaxation
Average	[rules]			0.745
Mediane	[rules]			0.719
Max Intensity	[rules]			<b>0.751</b>
Max Intensity or Average	[rules]			0.732
Max Frequeny	[rules]			0.725
First Opinion	[rules]			0.739
Last Opinion	[rules]			0.725

Tableau 7.49 : Meilleurs résultats (accuracy) pour chaque heuristique de calcul, filtrages effectués avec relaxation

Meilleur score: Pearson (avec relaxation)				
Modèle :	Filter: discourse	Filter: subjectivity	Filter: topic type	Pearson: avec relaxation
Average	top_simple	keep_all	keep_MT_PT	0.927
Mediane	top_simple	keep_all	keep_MT	<b>0.940</b>
Max Intensity	top_simple	keep_SI_SEI	keep_PT	0.890
Max Intensity or Average	[rules]			0.913
Max Frequeny	top_simple	keep_all	keep_PT	0.934
First Opinion	[rules]			0.875
Last Opinion	[rules]			0.898

Tableau 7.50 : Meilleurs résultats (corrélation de Pearson) pour chaque heuristique de calcul, filtrages effectués avec relaxation

Le tableau suivant présente l'apport de chaque brique à la reconnaissance de la polarité et de l'intensité, pour le corpus de commentaires web. Les métriques sont calculées avec application de la relaxation. Pour chaque score, nous précisons entre parenthèses les filtres appliqués et/ou l'heuristique de composition employée.

	Accuracy	Pearson
<b>Baseline</b>	0.63 ( <i>MaxFreq</i> )	0.84 ( <i>Moyenne</i> )
<b>Sac de segments</b>	0.63 (keep_all, keep_all, <i>MaxFreq</i> )	0.91 (keep_all, keep_all, <i>Médiane</i> )
<b>Filtres discursifs</b>	0.65	0.94
- <b>Filtre Sub_src</b>	0.63 (keep_all, keep_all, <i>MaxScore</i> )	0.90 (keep_all, keep_all, <i>Médiane</i> )
- <b>Filtre Sub_src_not_Dest</b>	0.63 (keep_all, keep_all, <i>MaxFreq</i> )	0.90 (keep_all, keep_all, <i>Médiane</i> )
- <b>Filtre Top_simple</b>	0.63 (keep_all, keep_all, <i>Médiane</i> )	0.94 (keep_all, keep_all, <i>Médiane</i> )
- <b>Filtre Top_récurif</b>	0.65 (keep_all, keep_all, <i>MaxScore</i> )	0.84 (keep_all, keep_all, <i>Médiane</i> )
<b>Règles de remontée</b>	<b>0.75</b> ( <i>MaxScore</i> )	0.91 ( <i>Moyenne</i> )

Tableau 7.51 : Apport des différents modules de fusion, corpus de commentaires web, polarité et intensité

**Corpus LeMonde**

Les tableaux suivants présentent les meilleurs résultats obtenus pour la reconnaissance de l'intensité et de la polarité, sur le corpus de réactions aux articles de presse.

<b>Meilleur score: accuracy (sans relaxation)</b>				
<b>Modèle :</b>	<b>Filter: discourse</b>	<b>Filter: subjectivity</b>	<b>Filter: topic type</b>	<b>Accuracy: sans rel.</b>
<b>Average</b>	[rules]			<b>0.728</b>
<b>Mediane</b>	[rules]			0.707
<b>Max Intensity</b>	keep_all	keep_SI_SEI	keep_all	0.696
<b>Max Intensity or Average</b>	[rules]			0.685
<b>Max Frequeny</b>	[rules]			0.674
<b>First Opinion</b>	[rules]			0.696
<b>Last Opinion</b>	[rules]			0.674

Tableau 7.52 : Meilleurs résultats (accuracy) pour chaque heuristique de calcul, filtrages effectués sans relaxation

<b>Meilleur score: accuracy (avec relaxation)</b>				
<b>Modèle :</b>	<b>Filter: discourse</b>	<b>Filter: subjectivity</b>	<b>Filter: topic type</b>	<b>Accuracy: avec relaxation</b>
<b>Average</b>	[rules]			<b>0.728</b>
<b>Mediane</b>	[rules]			0.707
<b>Max Intensity</b>	top_simple	keep_SI_SEI	keep_all	0.707
<b>Max Intensity or Average</b>	[rules]			0.685
<b>Max Frequeny</b>	[rules]			0.674
<b>First Opinion</b>	[rules]			0.696
<b>Last Opinion</b>	[rules]			0.674

Tableau 7.53 : Meilleurs résultats (accuracy) pour chaque heuristique de calcul, filtrages effectués avec relaxation



Meilleur score: Pearson (avec relaxation)				
Modèle :	Filter: discourse	Filter: subjectivity	Filter: topic type	Pearson: avec relaxation
Average	[rules]			<b>0.839</b>
Mediane	[rules]			0.830
Max Intensity	keep_all	keep_all	keep_all	0.823
Max Intensity or Average	[rules]			0.832
Max Frequeny	[rules]			0.832
First Opinion	[rules]			0.831
Last Opinion	[rules]			0.794

Tableau 7.54 : Meilleurs résultats (corrélation de Pearson) pour chaque heuristique de calcul, filtrages effectués avec relaxation

Le tableau suivant présente l'apport de chaque brique à la reconnaissance de la polarité et de l'intensité, pour le corpus de réactions aux articles de presse. Les métriques sont calculées avec application de la relaxation. Pour chaque score, nous précisons entre parenthèses les filtres appliqués et/ou l'heuristique de composition employée.

	Accuracy	Pearson
<b>Baseline</b>	0.60 ( <i>MaxFreq</i> )	0.66 ( <i>Moyenne</i> )
<b>Sac de segments</b>	0.70 (keep_all, keep_all, <i>MaxFreq</i> )	0.82 (keep_all, keep_all, <i>Médiane</i> )
<b>Filtres discursifs</b>	0.70	0.80
- <b>Filtre Sub_src</b>	0.69 (keep_all, keep_all, <i>MaxScore</i> )	0.78 (keep_all, keep_all, <i>Médiane</i> )
- <b>Filtre Sub_src_not_Dest</b>	0.69 (keep_all, keep_all, <i>MaxFreq</i> )	0.77 (keep_all, keep_all, <i>Médiane</i> )
- <b>Filtre Top_simple</b>	0.70 (keep_all, keep_all, <i>Médiane</i> )	0.78 (keep_all, keep_all, <i>Médiane</i> )
- <b>Filtre Top_récurif</b>	0.68 (keep_all, keep_all, <i>MaxScore</i> )	0.80 (keep_all, keep_all, <i>Médiane</i> )
<b>Règles de remontée</b>	<b>0.73</b> ( <i>MaxScore</i> )	<b>0.84</b> ( <i>Moyenne</i> )

Tableau 7.55 : Apport des différents modules de fusion, corpus de réactions aux articles de presse, polarité et intensité

### 7.6.3.3. Discussion

Nous pouvons tout d'abord remarquer que la méthode de relaxation est utile et nécessaire : elle permet une amélioration substantielle des résultats. Les filtres à utiliser avec ou sans relaxation sont sensiblement similaires, sauf dans le cas de la reconnaissance de polarité seule sur le corpus de commentaires web.

Concernant les heuristiques de composition, il n'est pas possible de dégager une unique heuristique dominante. Pour la polarité, sur le corpus de commentaires web, les heuristiques de Moyenne et de Fréquence maximale donnent les meilleurs résultats. Sur le corpus de réactions aux articles de presse, les heuristiques de Moyenne, de Médiane, et d'intensité maximum donnent quant à elles les meilleurs résultats. Pour la reconnaissance de polarité et d'intensité, les heuristiques donnant les meilleurs résultats sont respectivement les heuristiques d'Intensité maximale et de Médiane pour le corpus de commentaires web, et l'heuristique de Moyenne pour le corpus de réactions aux articles de presse. Globalement, le choix de l'heuristique influe assez peu le score, pourvu que le filtre soit adapté. On remarque également, comme attendu, que le choix d'une heuristique n'influe que marginalement sur le score obtenu lorsque les règles de remontée de l'opinion sont appliquées. Cela montre en particulier que les règles de remontée de l'opinion renvoient un ensemble de valeurs d'opinion globalement cohérentes entre elles.

Pour la reconnaissance de la polarité, l'utilisation de filtres discursifs permet une amélioration substantielle des résultats par rapport à la baseline ou à un modèle de sac de segments. Par contre, les règles ne permettent dans ce cas pas une amélioration des résultats. Ceci peut s'expliquer en regard de la relative simplicité de la tâche : les résultats obtenus après filtrage discursif sont déjà très bons, avec plus de 95 % de reconnaissance correcte. Au niveau des filtres discursifs, les filtres conservant les composants du haut du graphe (stratégies *top\_simple* et *top\_recursive*) semblent être systématiquement plus efficace que les stratégies se basant sur la présence d'une subordonnée (stratégies *sub\_src* et *sub\_src\_not\_dest*).

En termes de reconnaissance de polarité et d'intensité, tâche plus difficile, le filtrage discursif se montre moins efficace : le gain par rapport aux filtres plus simples du modèle de sac de segments est limité, voire nul dans le cas des réactions aux articles de presse. Par contre, l'utilisation des règles de remontée permet un gain non négligeable en termes d'accuracy : entre 3 % et 10 % de gain par rapport au modèle "*sac de segments*". Le coefficient de corrélation ne reflète néanmoins pas ce gain par rapport aux filtres. Ceci montre une plus grande stabilité et une plus grande tolérance aux erreurs des filtres par rapport aux règles. En effet, l'introduction d'erreurs dans le graphe de discours repéré, notamment au niveau du type de chaque relation de discours, peut changer grandement l'opinion remontée, ce qui est capté par le coefficient de Pearson. À l'inverse, les filtres se limitent au caractère subordonnant/coordonnant de chaque relation, ce qui limite l'impact des erreurs de repérage.

En résumé, le choix entre les règles de remontée de l'opinion et l'application d'un filtre doit se faire en fonction de deux éléments :

- la tâche effectuée : pour une simple reconnaissance de la polarité, l'utilisation de filtres est suffisante pour obtenir de bons résultats.
- la confiance dans le graphe de discours repéré : si celui-ci comporte notoirement des erreurs, celle-ci risquent de remettre en cause totalement la validité de la valeur d'opinion remontée par les règles ; un jeu de filtre serait par conséquent plus adapté, et plus à même de donner des résultats proches de la réalité. Si au contraire le graphe repéré l'est avec une grande précision, l'application des règles semble la mieux adaptée pour remonter une opinion précise et fiable.

## 7.7. Conclusion et perspectives

### 7.7.1. Conclusion

Dans ce chapitre, nous avons proposé un principe de calcul de l'opinion globale au niveau d'un document, basé sur les valeurs d'opinion calculées au niveau du segment de discours (cf. chapitre précédents). Cette fusion se déroule en deux étapes :

- Une étape de remontée d'un ensemble de scores d'opinion pertinents
- Une étape de calcul numérique sur cet ensemble d'opinion

Pour la première étape, nous avons proposé et testé deux méthodes :

- Une méthode par filtres des segments : suivant la position dans le graphe de discours, suivant le type de subjectivité du segment, et suivant le type de sujet de l'opinion associé au segment
- Un algorithme et un jeu de règles de calcul et de remontée de l'opinion le long du graphe de discours, permettant de sélectionner finement quelles valeurs d'opinion garder, supprimer, ou altérer suivant le type de chaque relation.

Les résultats obtenus montrent que si les règles proposées pour la remontée de l'opinion donnent des résultats très satisfaisants en termes d'accuracy, celles-ci sont également sensibles aux erreurs de reconnaissance du graphe de discours.

Pour la seconde étape, nous avons testé plusieurs heuristiques de calcul ou de sélection d'une valeur d'opinion dans l'ensemble remonté. Parmi celles-ci, l'heuristique de sélection de l'opinion de plus forte intensité semble donner de meilleurs résultats que les autres.

### 7.7.2. Perspectives

Plusieurs pistes sont envisageables pour améliorer ces résultats, et élargir le domaine d'application.

Tout d'abord, il pourrait être intéressant d'introduire l'émetteur de l'opinion, aussi bien comme filtre complémentaire qu'au niveau de certaines règles de remontée de l'opinion – par exemple, la règle concernant la relation de Correction pourrait être améliorée en tenant compte de

l'émetteur de la correction. Cela pourrait permettre d'améliorer les performances sur un domaine où le type d'émetteur est reconnu comme important – par exemple, un corpus de presse.

Il pourrait également être intéressant de tester dans quelle mesure les résultats obtenus dans ce chapitre sur un document unique peuvent s'appliquer au niveau d'une collection de documents, et plus précisément d'un ensemble de commentaires web sur un même sujet, établissant un dialogue entre différents interlocuteurs. Là encore, la notion d'émetteur de l'opinion a intuitivement un rôle important à jouer.

Enfin, il serait intéressant de tester les méthodes appliquées dans ce chapitre sur des textes dont la structure discursive est imparfaite, telle qu'obtenue par reconnaissance automatique. Ceci permettrait de vérifier la stabilité des méthodes et leur tolérance aux erreurs engendrées par les algorithmes de construction automatique du graphe de discours.





## 8. Conclusion et perspectives

---

### 8.1. Travaux effectués et apports

Les travaux présentés dans cette thèse s'articulent autour de trois niveaux de granularité de l'extraction d'opinion.

Tout d'abord, au niveau de l'expression d'opinion, où nous proposons un modèle d'annotation fin de l'opinion et de son contexte immédiat. L'expression d'opinion est caractérisée par une polarité, une intensité, et une catégorie sémantique. Elle porte sur un sujet – pouvant être implicite, et est émise par un émetteur – pouvant également être implicite. Ses attributs peuvent être altérés par la présence d'opérateurs linguistiques portant sur l'opinion : opérateurs de négation, d'intensité et de modalité. Ces opérateurs peuvent altérer l'intensité et la polarité d'une expression d'opinion, mais également la certitude associée à l'opinion.

Toujours au niveau de l'expression de l'opinion, nous avons étudié deux pistes afin de générer un lexique d'expressions d'opinion pour le français. La première méthode, inspirée de la résolution d'un réseau de contraintes CSP, permet de générer un lexique d'expressions d'opinion en se basant sur la taxinomie générique du français de Synapse Développement. La seconde, inspirée des travaux de (Takamura, Inui, and Okumura 2005) et de l'algorithme de détermination du spin d'un ensemble de particules en interaction, est basée sur le graphe d'un dictionnaire de synonymes. Ces deux méthodes permettent d'obtenir un lexique polarisé de qualité correcte, désambiguïsé sémantiquement suivant le découpage en sens des dictionnaires de Synapse Développement : ce découpage possède l'avantage d'être repérable automatiquement via l'analyseur syntaxique de la société (Laurent, Nègre, and Séguéla 2009). Néanmoins, un simple lexique de polarité n'étant pas suffisant pour la chaîne de traitement que nous souhaitions mettre en place, nous avons choisi un lexique manuel d'expressions d'opinion, indiquant polarité, intensité et catégories sémantiques pour chaque entrée, validé linguistiquement, et désambiguïsé sémantiquement suivant le découpage précédemment mentionné.

Nous avons également étudié l'opinion à un niveau intermédiaire entre l'expression et le document : le niveau du segment de discours. Nous proposons une méthode de détermination de la subjectivité d'un segment de discours suivant quatre catégories : segment subjectif explicite (SE – le segment porte de manière explicite une opinion), segment subjectif implicite (SI – le segment porte de manière implicite une opinion, i.e. le contexte permet de déduire que le segment porte une opinion), segment subjectif non évaluatif (SN – segment introducteur de discours), segment objectif (O – segment non porteur d'opinion). Pour effectuer cette classification, nous utilisons deux classifieurs automatiques binaires SVM, divisant l'ensemble des segments respectivement en Subjectif Hors Contexte (SE et SN) et Objectifs Hors Contexte (O et SI), et en Évaluatif (SE et SI) et Non-Évaluatif (O et SN) ; les intersections deux à deux des quatre ensembles de segments ainsi générés permettent de retrouver les quatre classes proposées. Parmi les traits de classification, nous proposons de prendre en compte, de manière itérative, la classe des segments en relation

discursive avec le segment classé. Sur des relations marquées, nous pouvons observer une amélioration des résultats de classification, illustrant l'intérêt de la prise en compte du discours, même partiel, pour cette tâche.

Nous proposons également un modèle géométrique d'application des opérateurs sur une expression d'opinion. Ce modèle parabolique est basé sur une étude linguistique de l'effet des opérateurs de négation et de modalité, basée sur une série de questionnaires proposés à des volontaires de langue maternelle française. Les réponses apportées à ces questionnaires ont permis de valider un certain nombre d'hypothèses linguistiques, hypothèses également validées par le modèle. Ce modèle prend en compte la polarité et l'intensité. Ce modèle permet de déterminer, au niveau d'un segment de discours subjectif, la polarité et l'intensité de l'opinion portée par le segment, mais aussi la certitude associée à celle-ci.

Enfin, nous proposons au niveau d'un document dans son ensemble deux méthodes de détermination de l'opinion globale, en termes de polarité mais aussi d'intensité, basée sur la structure discursive du document. La première méthode consiste à appliquer une série de filtres – basés sur le sujet de l'opinion, sur le type de subjectivité du segment, et sur la position dans le graphe de discours du document – puis à appliquer une heuristique numérique de fusion. La seconde associe à chaque type de relation de discours une règle d'agrégation de l'opinion ; un parcours du graphe de discours en profondeur d'abord et à l'intérieur de segments complexes d'abord permet de remonter l'opinion au niveau global. La détermination de l'opinion globale par règles de remontée se montre plus efficace pour la détermination précise de l'intensité que celle par filtrage, mais également plus sensible aux erreurs : le choix d'une méthode devra entre autres prendre en compte le niveau de confiance dans les relations de discours déterminées par un *parser* automatique.

Ces différents niveaux d'extraction sont regroupés au sein d'une chaîne de traitement logicielle. Cette chaîne de traitement s'appuie d'une part sur le module d'analyse syntaxique de Synapse Développement, et d'autre part sur le langage de règles Myéline, proposé dans le cadre du projet CASOAR et décrit dans cette thèse, qui permet d'exprimer certains aspects de l'extraction sous forme de règles, directement compilées en code C++ compatible avec l'architecture logicielle interne de Synapse. Ceci permet une maintenance simplifiée de la chaîne de traitement, tout en conservant les avantages souhaités d'un point de vue industriel : distribution « légère » (sans plate-forme supplémentaire), et temps d'exécution optimisé.

## 8.2. Valorisation industrielle

La chaîne de traitement a été utilisée à plusieurs reprises au cours de la thèse, à divers stades d'avancée, dans le cadre de projets industriels et de recherche.

Dans le cadre du projet OpenSEM (consortium : Synapse Développement – Exalead – Priberam), une version prototype de la chaîne de traitement a été utilisée afin d'extraire les opinions portant sur les films et acteurs dans des critiques de films. La liste des acteurs de chaque film était extraite par les partenaires du projet (Exalead), et était par conséquent fournie en entrée de la chaîne de traitement. Le rattachement de l'opinion à son sujet était effectué via une heuristique baseline



d'association d'un sujet candidat avec l'opinion la plus proche au sein d'une phrase. À ce stade, la chaîne de traitement n'utilisait pas encore le modèle parabolique pour le traitement de la négation, mais une simple baseline de renversement de polarité. La figure suivante présente une capture d'écran du prototype de démonstration – les opinions étiquetées pour l'espagnol et le portugais l'ont été par Priberam<sup>62</sup>, partenaire sur ce projet :

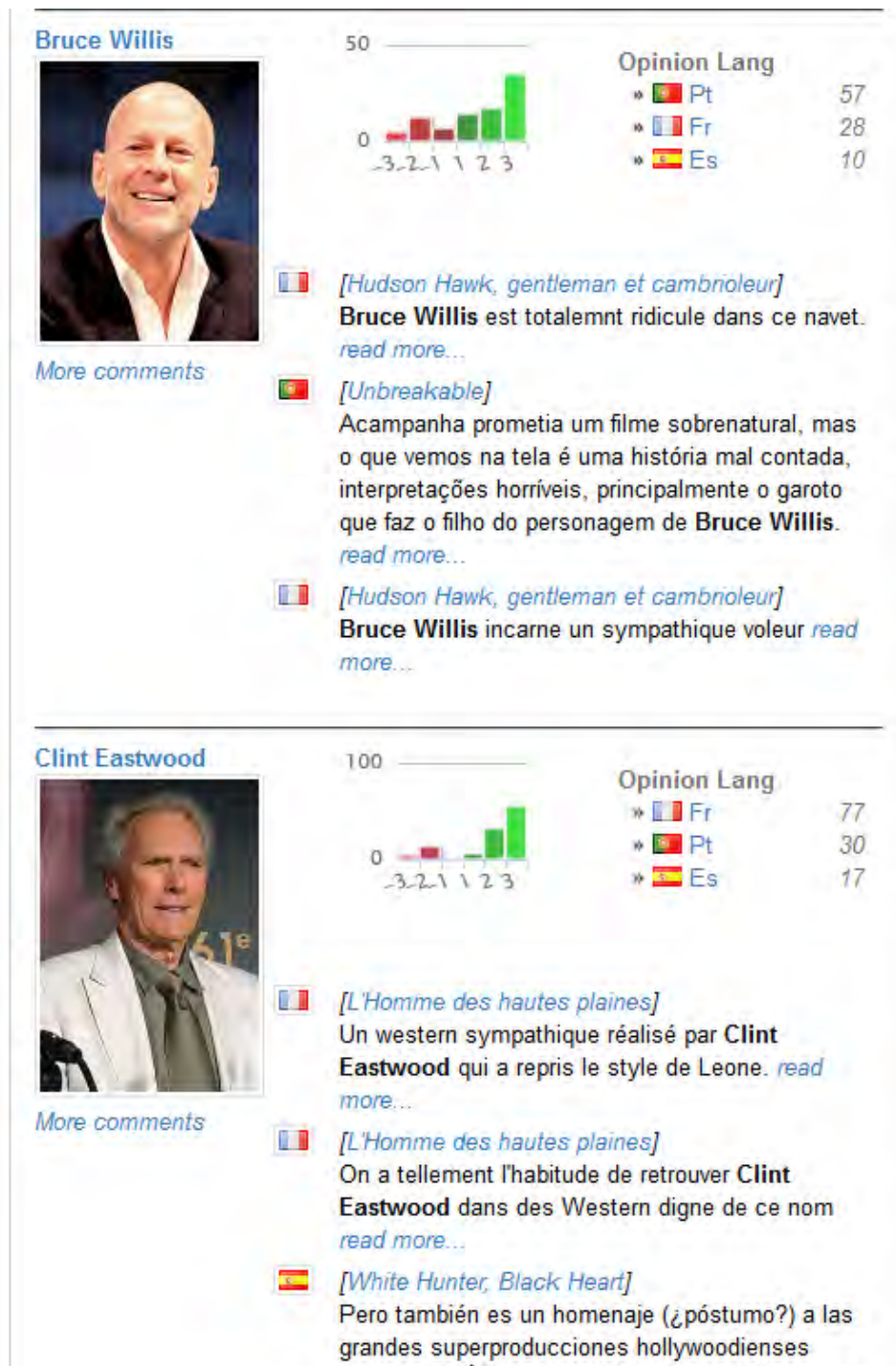


Figure 8.1 : L'interface d'affichage de résultats pour les acteurs du prototype OpenSem

<sup>62</sup> <http://www.priberam.pt/>

Nous avons également participé, dans le cadre du projet CASOAR (Synapse Développement – IRIT), à l'évaluation proposée par le projet DoXa sur les critiques de jeux-vidéo. La tâche comportait deux volets, la détermination de l'opinion au niveau d'un document d'une part (niveau macro), et au niveau d'un bloc de 1000 octets d'autre part (niveau meso). Dans le contexte de notre chaîne de traitement, ces deux niveaux sont traités en tant que document unitaire : en effet, le niveau meso, bien qu'interne au document complet, comporte typiquement plus d'une dizaine de segments de discours. La chaîne de traitement utilisée comportait, contrairement à celle utilisée dans le cadre du projet OpenSEM, une implémentation prototype du modèle parabolique pour le traitement de la négation. Pour chacun de ces deux niveaux, il fallait d'une part donner la polarité et l'intensité du bloc considéré, et d'autre part les catégories sémantiques DoXa prépondérantes du bloc. La sous-tâche de détermination de la polarité et de l'intensité du bloc était proche du traitement effectué par notre chaîne, et a donc nécessité peu d'adaptation. La seconde sous-tâche a par contre nécessité un développement spécifique. Afin de faire correspondre les catégories sémantiques DoXa aux catégories sémantiques CASOAR présentes dans notre lexique, nous avons établi une fonction de conversion, décrite dans le tableau ci-dessous. Cette fonction est basée à la fois sur les exemples fournis et sur la description donnée dans le guide d'annotation DoXa.

<b>Catégorie CASOAR</b>	<b>Catégorie DoXa équivalente</b>
Sentiment-Appréciation > Divertissement-Joie	Plaisir_Divertissement_Joie_Bonheur_Euphorie
Sentiment-Appréciation > Etonnement (polarité positive)	Etonnement_Positif
Sentiment-Appréciation > Espérer	Desir_Impatience_Envie
Sentiment-Appréciation > Apaisement	Soulagement_Apaisement
Sentiment-Appréciation > Intérêt-Passion (polarité positive)	Interet_Valorisation_Appreciation
Sentiment-Appréciation > tristesse-souffrance	Tristesse_Souffrance_Desespoir
Sentiment-Appréciation > Etonnement (polarité négative)	Etonnement_Negatif
Sentiment-Appréciation > Haine	Haine
Sentiment-Appréciation > Peur	Peur_Inquietude
Sentiment-Appréciation > Déception	Deception_Insatisfaction_Mecontentement
Sentiment-Appréciation > Ennui	Ennui
Sentiment-Appréciation > Intérêt-Passion (polarité négative)	Desinteret_Devalorisation_Depreciation
Jugement (polarité positive)	Interet_Valorisation_Appreciation
Jugement (polarité négative)	Desinteret_Devalorisation_Depreciation
Conseil > Recommander	Recommandation_Suggestion
Conseil > Suggérer	Recommandation_Suggestion

*Tableau 8.1 : Fonction de conversion des catégories sémantiques CASOAR vers les catégories DoXa*

Les résultats obtenus dans cette évaluation sont mitigés, mais tout à fait corrects compte tenu de la difficulté de la tâche. En effet, nous nous sommes confrontés à deux principaux problèmes. Tout d'abord, le domaine du jeu vidéo présente certaines spécificités au niveau lexical favorisant les

faux positifs : par exemple, la présence de monstres « horribles » dans un jeu n'est pas nécessairement négative. Cette particularité a nécessité de porter une attention toute particulière au lexique de repérage, de prendre en compte certains sens spécifiques au domaine, et de désactiver via une liste noire certains termes du lexique générique CASOAR menant à des contre-sens. D'autre part, le choix du bloc de 1000 caractères / octets comme unité intermédiaire de classification est fortement défavorable à une approche sémantique fine et / ou discursive. En effet, les blocs présentés étaient le plus souvent coupés au milieu d'une phrase ou d'un segment de discours. Ceci nuit grandement aux performances de l'analyse syntaxique, et donc du module de désambiguïsation sémantique ; ceci limite également le repérage de relations rhétoriques pertinentes, même marquées, dans la mesure où le bloc ne forme plus nécessairement une unité discursive. Les résultats complets de cette évaluation ne sont malheureusement pas diffusables, en raison d'un accord de confidentialité.

Enfin, le prototype final de la chaîne de traitement a été utilisé dans le cadre du projet CASOAR pour un partenariat avec le site [Marmiton.org](http://Marmiton.org)<sup>63</sup>. Le site *Marmiton* est un site collaboratif de recettes de cuisine : les internautes peuvent venir y déposer leur recette, commenter les recettes des autres internautes, et noter celles-ci. Récemment, le site a également lancé une section concernant les critiques de restaurants. L'application développée consiste à déterminer l'opinion au niveau d'une critique sur un restaurant, et, à l'intérieur de la critique, d'identifier sur quels plats portent les opinions unitaires. En entrée, les textes fournis sont accompagnés d'une liste de plats proposés par le restaurant ; ces plats ne sont par contre pas étiquetés dans le texte. Un module de repérage des plats en contexte a donc été développé par Synapse Développement, sur la base de la taxinomie générique des mots du français. Ces plats-candidats extraits ont ensuite été associés aux opinions suivant une amélioration de l'heuristique précédemment décrite : l'association est effectuée au sein d'une même proposition tant que l'expression et le candidat-sujet ne sont pas séparés par une virgule. Cette heuristique simple permet d'obtenir une bonne précision sans sacrifier trop de rappel. La figure suivante présente le prototype de démonstration :

---

<sup>63</sup> [www.marmiton.org](http://www.marmiton.org)



Figure 8.2 : Extrait de l'interface de l'application Marmiton, dans le cadre du projet Casoar

### 8.3. Perspectives

Nous pouvons identifier plusieurs pistes de perspectives pour faire suite aux travaux décrits dans cette thèse.

Tout d'abord, le modèle parabolique proposé prend en compte la polarité, l'intensité et le degré de certitude associés à une opinion. Néanmoins, il ne prend pas en compte l'engagement de l'émetteur. Cette notion, distincte de la notion de certitude (l'émetteur peut être fortement engagé sur le fait que l'opinion soit peu certaine, par exemple dans la phrase : « Je maintiens que la nourriture servie dans ce restaurant est potentiellement mauvaise. »), n'est actuellement pas prise en compte dans le modèle. La prise en compte de cette notion, nécessitant une dimension supplémentaire dans l'optique d'un modèle géométrique, pourrait par exemple mener à un modèle 3D basé sur des paraboloides au lieu de paraboles.

Dans le cadre de la remontée de l'opinion globale, il reste à tester en situation réelle, c'est-à-dire en se basant sur les sorties d'un *parser* de discours automatique plutôt que sur une annotation manuelle *gold*, les méthodes de détermination de cette opinion. Les résultats préliminaires laissent à penser que les règles dans leur formulation actuelle sont relativement sensibles aux erreurs. Si cela est confirmé pour le type d'erreurs issues d'une mauvaise reconnaissance automatique du discours, un travail d'affinage des règles et de détection des erreurs peut s'avérer nécessaire. Une piste pourrait être d'associer un coefficient de confiance à une règle de remontée.

Enfin, les modules présentés dans cette thèse utilisent parfois le sujet de l'opinion, mais aucun module générique n'est proposé pour résoudre cette tâche. Dans le contexte des applications

industrielles et des prototypes de démonstration pour les projets de recherche, un module *ad hoc* peut être développé. Néanmoins, il serait intéressant de déterminer dans quelle mesure ce travail d'adaptation peut être facilité via un module générique d'extraction de candidats-sujets de l'opinion.



# Bibliographie

---

- Adda, G., B. Sagot, K. Fort, and J. Mariani. 2011. "Crowdsourcing for Language Resource Development: Critical Analysis of Amazon Mechanical Turk Overpowering Use." In *LTC 2011: Proceedings of the 5th Language and Technology Conference*.
- Afantenos, Stergos, Nicholas Asher, Farah Benamara, Myriam Bras, C. Fabre, Mai Ho-dac, A. Le Draoulec, Philippe Muller, M. P. Péry-Woodley, and L. Prévot. 2012. "An Empirical Resource for Discovering Cognitive Principles of Discourse Organisation: The Annodis Corpus." In *Proceedings of LREC*.
- Afantenos, Stergos, Pascal Denis, Philippe Muller, and Laurence Danlos. 2010. "Learning Recursive Segments for Discourse Parsing." In *arXiv:1003.5372*. La Valette, Malte.
- Akkaya, Cem, Janyce Wiebe, and Rada Mihalcea. 2009. "Subjectivity Word Sense Disambiguation." In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*, 190–199.
- Andreevskaia, Alina, and Sabine Bergler. 2006. "Mining WordNet for Fuzzy Sentiment: Sentiment Tag Extraction from WordNet Glosses." In *Proceedings of EAACL'06*.
- Asher, Nicholas, Farah Benamara, and Yvette Yannick Mathieu. 2008. "Distilling Opinion in Discourse: A Preliminary Study." *Coling 2008: Companion Volume: Posters and Demonstrations*: 5–8.
- Asher, Nicholas, Farah Benamara, and Yvette Yannick Mathieu. 2009. "Appraisal of Opinion Expressions in Discourse." *Linguisticae Investigationes: Revue Internationale De Linguistique Française Et De Linguistique Générale* 32 (2): 279–292.
- Asher, Nicholas, and Alex Lascarides. 2003. *Logics of Conversation*. Cambridge Univ Pr.
- Asher, Nicholas, and L. Vieu. 2005. "Subordinating and Coordinating Discourse Relations." *Lingua* 115 (4): 591–610.
- Aue, A., and M. Gamon. 2005. "Customizing Sentiment Classifiers to New Domains: A Case Study." In *Proceedings of Recent Advances in Natural Language Processing (RANLP)*, 1:2–1.
- Béchet, F., A. Nasr, and F. Genet. 2000. "Tagging Unknown Proper Names Using Decision Trees." In *Proceedings of the 38th Annual Meeting on Association for Computational Linguistics*, 77–84.
- Benamara, Farah, Baptiste Chardon, Yvette Yannick Mathieu, and Vladimir Popescu. 2011. "Towards Context-Based Subjectivity Analysis." *IJCNLP 2011*.
- Benamara, Farah, Baptiste Chardon, Yvette Yannick Mathieu, Vladimir Popescu, and Nicholas Asher. 2012. "How Do Negation and Modality Impact on Opinions?" In *Proceedings of the ACL Workshop on Extra-Propositional Aspects of Meaning in Computational Linguistics (ExProM 2012)*, 10–18. Jeju, Republic of Korea.
- Benamara, Farah, Vladimir Popescu, Baptiste Chardon, Nicholas Asher, and Yvette Yannick Mathieu. à paraître. "Assessing Opinions in Texts: Does Discourse Really Matter?" In *Maite Taboada and Radoslava Trnavac Eds*. Brill publisher.
- Ben-David, S., J. Blitzer, K. Crammer, and F. Pereira. 2006. "Analysis of Representations for Domain Adaptation." *Advances in Neural Information Processing Systems* 19: 137.
- Bestgen, Yves, and G. Lories. 2009. "Un Niveau De Base Pour La Tâche 1 (corpus Français Et Anglais) De DEFT'09." *Actes Du Cinquième Défi Fouille De Textes*: 65.
- Bethard, Steven, Hong Yu, Ashley Thornton, Vasileios Hatzivassiloglou, and Daniel Jurafsky. 2004. "Automatic Extraction of Opinion Propositions and Their Holders." In *2004 AAAI Spring Symposium on Exploring Attitude and Affect in Text*, 22–24.

- Blitzer, J., M. Dredze, and F. Pereira. 2007. "Biographies, Bollywood, Boom-boxes and Blenders: Domain Adaptation for Sentiment Classification." In *Annual Meeting-Association For Computational Linguistics*, 45:440.
- Blitzer, J., R. McDonald, and F. Pereira. 2006. "Domain Adaptation with Structural Correspondence Learning." In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, 120–128.
- Boser, B. E., I. M. Guyon, and V. N. Vapnik. 1992. "A Training Algorithm for Optimal Margin Classifiers." In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, 144–152.
- Bradley, M. M, and P. J Lang. 1999. "Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings." *University of Florida: The Center for Research in Psychophysiology*.
- Burges, Christopher J. C. 1998. "A Tutorial on Support Vector Machines for Pattern Recognition." *Data Mining and Knowledge Discovery 2*: 121–167.
- Carlson, L., D. Marcu, and M. E. Okurowski. 2003. "Building a Discourse-tagged Corpus in the Framework of Rhetorical Structure Theory." *Current and New Directions in Discourse and Dialogue*: 85–112.
- Charaudeau, P. 1992. *Grammaire Du Sens Et De L'expression*. Hachette Paris.
- Chardon, Baptiste. 2010. "Catégorisation Automatique D'adjectifs D'opinion à Partir D'une Ressource Linguistique Générique." *Actes De RECITAL*.
- Chardon, Baptiste, Farah Benamara, Yvette Yannick Mathieu, Vladimir Popescu, and Nicholas Asher. 2013. "Sentiment Composition Using a Parabolic Model." In *Proceedings of IWCS 2013*. Potsdam, Germany.
- Chardon, Baptiste, Farah Benamara, Vladimir Popescu, Yvette Yannick Mathieu, and Nicholas Asher. 2013. "Measuring the Effect of Discourse Structure on Sentiment Analysis." In *Proceedings of CICLING 2013*. Samos, Greece.
- Chen, W. 2008. "Dimensions of Subjectivity in Natural Language." In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers*, 13–16.
- Chesley, P., B. Vincent, L. Xu, and R. K Srihari. 2006. "Using Verbs and Adjectives to Automatically Classify Blog Sentiment." *Training* 580 (263): 233.
- Choi, Yejin, and Claire Cardie. 2008. "Learning with Compositional Semantics as Structural Inference for Subsentential Sentiment Analysis." In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 793–801.
- Choi, Yejin, Claire Cardie, Ellen Riloff, and Siddharth Patwardhan. 2005. "Identifying Sources of Opinions with Conditional Random Fields and Extraction Patterns." In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP)*, 355–362. Vancouver, Canada.
- Collobert, R., S. Bengio, and J. Mariéthoz. 2002. *Torch: a Modular Machine Learning Software Library*. Technical Report IDIAP-RR 02-46, IDIAP.
- Cooper, M., S. De Givry, M. Sanchez, T. Schiex, and M. Zytnicki. 2008. "Cohérence D'arc Virtuelle Pour Les CSP Pondérés." In *JFPC 2008-Quatrièmes Journées Francophones De Programmation Par Contraintes*, 237–246.
- Cover, Thomas, and Peter Hart. 1967. "Nearest Neighbor Pattern Classification." *Information Theory, IEEE Transactions On* 13 (1): 21–27.
- Cunningham, Hamish, D Maynard, and V Tablan. 2000. *JAPE: a Java Annotation Patterns Engine (Second Edition)*.
- Cunningham, Hamish, Diana Maynard, and Kalina Bontcheva. 2011. *Text Processing with GATE*. University of Sheffield Department of Computer Science.



- Daelemans, W., J. Zavrel, K. Van der Sloot, and A. Van den Bosch. 2007. "Timbl: Tilburg Memory-based Learner." *Version 6*: 07–03.
- Daille, Béatrice, Estelle Dubreil, Laura Monceaux, and Matthieu Vernier. 2011. "Annotating Opinion—evaluation of Blogs: The Blogoscopy Corpus." *Language Resources and Evaluation* 45 (4) (June 29): 409–437.
- Danlos, Laurence. 2008. "Strong Generative Capacity of RST, SDRT and Discourse Dependency DAGs." *PRAGMATICS AND BEYOND NEW SERIES* 172: 69.
- Dave, Kushal, Steve Lawrence, and David M. Pennock. 2003. "Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews." In *Proceedings of the 12th International Conference on World Wide Web*, 519–528. Budapest, Hungary: ACM.
- Denecke, Kerstin. 2009. "Are SentiWordNet Scores Suited for Multi-domain Sentiment Classification?" In *Digital Information Management, 2009. ICDIM 2009. Fourth International Conference On*, 1–6.
- Ding, Xiaowen, B. Liu, and P. S. Yu. 2008. "A Holistic Lexicon-based Approach to Opinion Mining." In *Proceedings of the International Conference on Web Search and Web Data Mining*, 231–240.
- Dubreil, Estelle, Matthieu Vernier, Laura Monceaux, and Béatrice Daille. 2008. "Annotating Opinion – Evaluation of Blogs." In Marrakech, Morocco.
- Ebbinghaus, H. D., J. Flum, and W. Thomas. 1994. *Mathematical Logic*. Springer.
- Esuli, Andrea, and Fabrizio Sebastiani. 2006a. "SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining." In *Proceedings of LREC*. Vol. 6.
- Esuli, Andrea, and Fabrizio Sebastiani. 2006b. "Determining Term Subjectivity and Term Orientation for Opinion Mining." In *Proceedings the 11th Meeting of the European Chapter of the Association for Computational Linguistics (EACL-2006)*, 193–200.
- Fellbaum, Edited by Christiane. 1998. "WordNet: An Electronic Lexical Database" (May 15).
- Forest, Dominic, van Hoeydonck, D. Létourneau, and M. Bélanger. 2009. "Impacts De La Variation Du Nombre De Traits Discriminants Sur La Catégorisation Des Documents." *Actes Du Cinquième Défi Fouille De Textes*: 77.
- Freund, Y., and R. E. Schapire. 1996. "Experiments with a New Boosting Algorithm." In *MACHINE LEARNING-INTERNATIONAL WORKSHOP THEN CONFERENCE-*, 148–156.
- Gamon, M., and A. Aue. 2005. "Automatic Identification of Sentiment Vocabulary: Exploiting Low Association with Known Sentiment Terms." In *Proceedings of the ACL Workshop on Feature Engineering for Machine Learning in Natural Language Processing*, 57–64.
- Godard, Danièle. 2013. "Les Négateurs." In *La Grande Grammaire Du Français*. Abeillé, Anne; Godard, Danièle; Delaveau, Annie; (Ed.).
- Greene, Stephan, and Philip Resnik. 2009. "More Than Words: Syntactic Packaging and Implicit Sentiment." In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, 503–511. NAACL '09. Stroudsburg, PA, USA: Association for Computational Linguistics.
- Grouin, C., B. Arnulphy, J. B. Berthelin, S. El Ayari, A. García-Fernandez, A. Grappy, M. Hurault-Plantet, P. Paroubek, I. Robba, and P. Zweigenbaum. 2009. "Présentation De L'édition 2009 Du Défi Fouille De Textes (DEFT'09)." *Actes Du Cinquième Défi Fouille De Textes*: 35.
- Grouin, C., J. B. Berthelin, S. El Ayari, T. Heitz, M. Hurault-Plantet, M. Jardino, Z. Khalis, and M. Lastes. 2007. "Présentation De Deft'07 (défi Fouille De Textes)." *Actes Du Troisième Défi Fouille De Textes*: 3.
- Harb, A., G. Dray, M. Plantié, P. Poncelet, M. Roche, and F. Troussset. 2008. "Détection d'Opinion: Apprenons Les Bons Adjectifs!" In *Actes De l'Atelier FOuille Des Données d'OPinions (FODOP 08), Conjointement Conférence INFORSID*. Vol. 8.

- Hatzivassiloglou, Vasileios, and Kathleen R. McKeown. 1997. "Predicting the Semantic Orientation of Adjectives." In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics*, 181.
- He, H., B. Chen, L. Du, S. Li, H. Gao, W. Xu, and J. Guo. 2008. *PRIS in TREC 2008 Blog Track*. DTIC Document.
- Heerschop, B., F. Goossen, A. Hogenboom, F. Frasincar, U. Kaymak, and F. de Jong. 2011. "Polarity Analysis of Texts Using Discourse Structure." In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, 1061–1070.
- Hu, Mingqing, and Bing Liu. 2004. "Mining and Summarizing Customer Reviews." In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 168–177. Seattle, WA, USA: ACM.
- Johansson, R., and A. Moschitti. 2010. "Reranking Models in Fine-grained Opinion Analysis." In *Proceedings of the 23rd International Conference on Computational Linguistics*, 519–527.
- Kaji, N., and M. Kitsuregawa. 2007. "Building Lexicon for Sentiment Analysis from Massive Collection of HTML Documents." In *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, 1075–1083.
- Kamp, H., and U. Reyle. 1993. *From Discourse to Logic: Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. Vol. 42. Kluwer Academic.
- Kamps, Jaap, and Maarten Marx. 2002. "Words with Attitude." In *Proceedings of the 1st International WordNet Conference, Mysore, India*, 332–341.
- Kamps, Jaap, Maarten Marx, R. J. Mokken, and M. De Rijke. 2004. "Using WordNet to Measure Semantic Orientation of Adjectives." In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004)*, 4:1115–1118.
- Kanayama, H., and T. Nasukawa. 2006. "Fully Automatic Lexicon Expansion for Domain-oriented Sentiment Analysis." In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2006)*, 355–363.
- Kennedy, A., and D. Inkpen. 2006. "Sentiment Classification of Movie Reviews Using Contextual Valence Shifters." *Computational Intelligence* 22 (2): 110–125.
- Kim, Soo-Min, and Eduard Hovy. 2005. "Automatic Detection of Opinion Bearing Words and Sentences." In *Proceedings of IJCNLP 2005*. Vol. 5. Jeju Island, Korea.
- Kim, Soo-Min, and Eduard Hovy. 2006a. "Automatic Identification of Pro and Con Reasons in Online Reviews." In *Proceedings of the COLING/ACL on Main Conference Poster Sessions*, 483–490.
- Kim, Soo-Min, and Eduard Hovy. 2006b. "Extracting Opinions, Opinion Holders, and Topics Expressed in Online News Media Text." In *Proceedings of the Workshop on Sentiment and Subjectivity in Text*, 1–8. Sydney, Australia: Association for Computational Linguistics.
- Kobayakawa, T. S., T. Kumano, H. Tanaka, N. Okazaki, J. D. Kim, and J. Tsujii. 2009. "Opinion Classification with Tree Kernel Svm Using Linguistic Modality Analysis." In *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, 1791–1794.
- Kobayashi, N., T. Inui, and K. Inui. 2001. "Dictionary-based Acquisition of the Lexical Knowledge for P/n Analysis." In *Proceedings of Japanese Society for Artificial Intelligence, SLUD*, 33:45–50.
- Kuhn, R., and R. De Mori. 1995. "The Application of Semantic Classification Trees to Natural Language Understanding." *Pattern Analysis and Machine Intelligence, IEEE Transactions On* 17 (5): 449–460.

- Lafferty, J., A. McCallum, and F. Pereira. 2001. "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data." In *MACHINE LEARNING-INTERNATIONAL WORKSHOP THEN CONFERENCE-*, 282–289.
- Larrea, Paul. 2004. "L'expression De La Modalité En Français Et En Anglais (domaine Verbal)." *Revue Belge De Philologie Et D'histoire* 82 (3): 733–762.
- Laurent, Dominique, Sophie Nègre, and Patrick Séguéla. 2009. "L'analyseur Syntaxique Cordial Dans Passage." *Actes De TALN 9*.
- Laurent, Dominique, and Patrick Séguéla. 2003. "Proximité Synonymique Badauds: Flâneurs Ou Spectateurs?"
- Liu, Bing. 2010. "Sentiment Analysis and Subjectivity." *Handbook of Natural Language Processing*,: 627–666.
- Liu, J., and S. Seneff. 2009. "Review Sentiment Scoring via a Parse-and-paraphrase Paradigm." In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*, 161–169.
- Mann, W. C., and S. A. Thompson. 1988. "Rhetorical Structure Theory: Toward a Functional Theory of Text Organization." *Text* 8 (3): 243–281.
- Marcu, D. 2000. *The Theory and Practice of Discourse Parsing and Summarization*. The MIT press.
- Martin, J. R. 2005. *The Language of Evaluation: Appraisal in English*. Basingstoke: Palgrave Macmillan.
- Mathieu, Yvette Yannick. 2000. *Les Verbes De Sentiment. De L'analyse Linguistique Au Traitement Automatique*. CNRS éditions. Paris.
- Mathieu, Yvette Yannick. 2006. "A Computational Semantic Lexicon of French Verbs of Emotion." In *Computing Attitude and Affect in Text: Theory and Applications*, edited by James Shanahan, Yan Qu, Janyce Wiebe, and W. Bruce Croft, 20:109–124. The Information Retrieval Series. Springer Netherlands.
- Mathieu, Yvette Yannick. 2008. "Navigation Dans Un Texte à La Recherche Des Sentiments." *Linguisticae Investigationes* 31 (2): 313–322.
- Mathieu, Yvette Yannick, and C. Felbaum. 2010. "Verbs of Emotion in French and English." *Proceedings of GWC-2010, Mumbai, India*.
- Matsumoto, S., H. Takamura, and M. Okumura. 2005. "Sentiment Classification Using Word Sub-sequences and Dependency Sub-trees." *Advances in Knowledge Discovery and Data Mining*: 21–32.
- Miller, G. A., R. Beckwith, C. Fellbaum, D. Gross, and K. J. Miller. 1990. "Introduction to Wordnet: An On-line Lexical Database\*." *International Journal of Lexicography* 3 (4): 235–244.
- Miltsakaki, E., R. Prasad, A. Joshi, and B. Webber. 2004. "The Penn Discourse Treebank." In *Proceedings of the 4th International Conference on Language Resources and Evaluation*.
- Mirkin, S., I. Dagan, and S. Padó. 2010. "Assessing the Role of Discourse References in Entailment Inference." In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, 1209–1219.
- Mitchell, T. M. 1997. "Machine Learning. 1997." *Burr Ridge, IL: McGraw Hill*.
- Moilanen, K., and S. Pulman. 2007. "Sentiment Composition." In *Proceedings of the Recent Advances in Natural Language Processing International Conference*, 378–382.
- Mullen, Tony, and Nigel Collier. 2004. "Sentiment Analysis Using Support Vector Machines with Diverse Information Sources." In *Proceedings of EMNLP 2004*, 412–418. Barcelona, Spain.
- Musat, Claudiu, and Stefan Trausan-Matu. 2010. "The Impact of Valence Shifters on Mining Implicit Economic Opinions." In *Proceedings of the 14th International Conference on Artificial Intelligence: Methodology, Systems, and Applications*, 131–140. AIMSA'10. Berlin, Heidelberg: Springer-Verlag.

- Ng, Vincent, Sajib Dasgupta, and S. M. Niaz Arifin. 2006. "Examining the Role of Linguistic Knowledge Sources in the Automatic Identification and Classification of Reviews." In *Proceedings of the COLING/ACL on Main Conference Poster Sessions*, 611–618. COLING-ACL '06. Stroudsburg, PA, USA: Association for Computational Linguistics.
- Ortony, A., G. L. Clore, and A. Collins. 1988. *The Cognitive Structure of Emotions*. Cambridge Univ Pr.
- Osgood, C. E., G. J. Suci, and P. Tannenbaum. 1967. *The Measurement of Meaning*. Vol. 47. University of Illinois Press.
- Ounis, I., C. Macdonald, and I. Soboroff. 2008. *Overview of the Trec-2008 Blog Track*. DTIC Document.
- Pang, Bo, and Lillian Lee. 2004. "A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts." In *Proceedings of the 42nd Meeting of the Association for Computational Linguistics (ACL'04), Main Volume*, 271–278. Barcelona, Spain.
- Pang, Bo, and Lillian Lee. 2005. "Seeing Stars: Exploiting Class Relationships for Sentiment Categorization with Respect to Rating Scales." In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, 115–124. Ann Arbor, Michigan: Association for Computational Linguistics.
- Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. 2002. "Thumbs up?: Sentiment Classification Using Machine Learning Techniques." In *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing - Volume 10*, 79–86. Association for Computational Linguistics.
- Paroubek, P., J. B. Berthelin, S. El Ayari, C. Grouin, T. Heitz, M. Hurault-Plantet, M. Jardino, Z. Khalis, and M. Lastes. 2007. "Résultats De L'édition 2007 Du DEfi Fouille De Textes." *Actes De L'atelier De Clôture Du 3eme DEfi Fouille De Textes*: 9–17.
- Paroubek, P., A. Pak, and D. Mostefa. 2010. "Annotations for Opinion Mining Evaluation in the Industrial Context of the DOXA Project." In *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC'10), Valletta, Malta*.
- Péry-Woodley, Marie-Paule, Nicholas Asher, Patrice Enjalbert, Farah Benamara, Myriam Bras, Cécile Fabre, Stéphane Ferrari, et al. 2009. *ANNODIS: Une Approche Outillée De L'annotation De Structures Discursives*.
- Plantié, M., G. Dray, M. Roche, and others. 2007. "Défi DEFT07: Comparaison D'approches Pour La Classification De Textes D'opinion." *Actes De L'atelier De Clôture Du 3ème DEfi Fouille De Textes, Grenoble, France*: 57–69.
- Polanyi, Livia, and Martin van den Berg. 2011. "Discourse Structure and Sentiment." In *Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference On*, 97–102.
- Polanyi, Livia, and Annie Zaenen. 2006. "Contextual Valence Shifters." In *Computing Attitude and Affect in Text: Theory and Applications*, edited by James Shanahan, Yan Qu, Janyce Wiebe, and W. Bruce Croft, 20:1–10. The Information Retrieval Series. Springer Netherlands.
- Popescu, Ana-Maria, and Oren Etzioni. 2007. "Extracting Product Features and Opinions from Reviews." In *Natural Language Processing and Text Mining*, 9–28.
- Portner, P. 2009. *Modality*. Vol. 1. Oxford University Press, USA.
- Quirk, R, S Greenbaum, G Leech, and J Svartvik. 1985. *A Comprehensive Grammar of the English Language*. Longman.
- Redondo, Jaime, Isabel Fraga, Isabel Padrón, and Montserrat Comesaña. 2007. "The Spanish Adaptation of ANEW (Affective Norms for English Words)." *Behavior Research Methods* 39 (3): 600–605.

- Riloff, Ellen, Janyce Wiebe, and William Phillips. 2005. "Exploiting Subjectivity Classification to Improve Information Extraction." In *Proceedings of the National Conference On Artificial Intelligence*, 20:1106.
- Riloff, Ellen, Janyce Wiebe, and Theresa Wilson. 2003. "Learning Subjective Nouns Using Extraction Pattern Bootstrapping." In *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003-Volume 4*, 25–32.
- Schapire, R. E., and Y. Singer. 2000. "Booster: A Boosting-based System for Text Categorization." *Machine Learning* 39 (2): 135–168.
- Schmid, H. 1994. "Probabilistic Part-of-speech Tagging Using Decision Trees." In *Proceedings of International Conference on New Methods in Language Processing*, 12:44–49.
- Schmid, H. 1995. "Improvements in Part-of-speech Tagging with an Application to German." In *In Proceedings of the ACL SIGDAT-Workshop*.
- Somasundaran, Swapna. 2010. "Discourse-level Relations For Opinion Analysis". University of Pittsburgh.
- Somasundaran, Swapna, Janyce Wiebe, and Josef Ruppenhofer. 2008. "Discourse Level Opinion Interpretation." In *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1*, 801–808.
- Soricut, R., and D. Marcu. 2003. "Sentence Level Discourse Parsing Using Syntactic and Lexical Information." In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*, 149–156.
- Stone, P.-J. 1966. "General Inquirer - The MIT Press."
- Stoyanov, Veselin, and C. Cardie. 2008. "Annotating Topics of Opinions." In *Proceedings of LREC*.
- Strapparava, C., and A. Valitutti. 2004. "WordNet-Affect: An Affective Extension of WordNet." In *Proceedings of the 4th International Conference on Language Resources and Evaluation*, 1083–1086.
- Taboada, Maite, C. Anthony, and K. Voll. 2006. "Methods for Creating Semantic Orientation Dictionaries." In *Conference on Language Resources and Evaluation (LREC)*, 427–432.
- Taboada, Maite, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. 2011. "Lexicon-based Methods for Sentiment Analysis." *Computational Linguistics* 37 (2): 267–307.
- Taboada, Maite, K. Voll, and J. Brooke. 2008. "Extracting Sentiment as a Function of Discourse Structure and Topicality." *Simon Fraser University, Tech. Rep 20*.
- Takamura, H., T. Inui, and M. Okumura. 2005. "Extracting Semantic Orientations of Words Using Spin Model." In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, 140.
- Takamura, H., T. Inui, and M. Okumura. 2007. "Extracting Semantic Orientations of Phrases from Dictionary." In *Proceedings of NAACL/HLT*.
- Tanguy, L., F. Lalleman, C. François, P. Muller, P. Séguéla, and others. 2009. "RHECITAS: Citation Analysis of French Humanities Articles."
- Thomas, Matt, Bo Pang, and Lillian Lee. 2006. "Get Out the Vote: Determining Support or Opposition from Congressional Floor-debate Transcripts." In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, 327–335. Sydney, Australia: Association for Computational Linguistics.
- Toprak, Cigdem, and Iryna Gurevych. 2009. "Document Level Subjectivity Classification Experiments in Deft'09 Challenge." *Actes Du Cinquième Défi Fouille De Textes*: 91.
- Toprak, Cigdem, Niklas Jakob, and Iryna Gurevych. 2010. "Sentence and Expression Level Annotation of Opinions in User-Generated Discourse." In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, 575–584. Uppsala, Sweden: Association for Computational Linguistics.

- Torres-Moreno, J. M., M. El-Bèze, F. Béchet, and N. Camelin. 2007. "Comment Faire Pour Que L'opinion Forcée à La Sortie Des Urnes Soit La Bonne? Application Au Défi DEFT 2007." *Actes Du Troisième Défi Fouille De Textes*: 129.
- Torres-Moreno, J. M., M. El-Bèze, F. Béchet, and N. Camelin. 2009. "Fusion Probabiliste Appliquée à La Détection Et Classification D'opinions." *Actes Du Cinquième Défi Fouille De Textes*: 17.
- Trnavac, Radoslava, and Maite Taboada. 2011. "The Contribution of Nonveridical Rhetorical Relations to Evaluation in Discourse." *Language Sciences*.
- Turney, Peter. 2002. "Thumbs up or Thumbs down? Semantic Orientation Applied to Unsupervised Classification of Reviews." In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, 417–424. Philadelphia, Pennsylvania.
- Turney, Peter, and Michael Littman. 2002. "Unsupervised Learning of Semantic Orientation from a Hundred-billion-word Corpus." *Arxiv Preprint Cs/0212012*.
- Turney, Peter, and Michael Littman. 2003. "Measuring Praise and Criticism: Inference of Semantic Orientation from Association." *ACM Transactions on Information Systems (TOIS)* 21 (4): 315–346.
- Vernier, Matthieu. 2011. "Analyse à Granularité Fine De La Subjectivité". Nantes.
- Vernier, Matthieu, Yann Mathet, F. Rioult, Thierry Charnois, Stéphane Ferrari, and Dominique Legallois. 2007. "Classification De Textes D'opinions: Une Approche Mixte N-grammes Et Sémantique." In *Actes De L'atelier Défi Fouilles De Textes (DEFT'07)*, 93–108.
- Vernier, Matthieu, and L. Monceaux. 2010. "Enrichissement D'un Lexique De Termes Subjectifs à Partir De Tests Sémantiques."
- Vernier, Matthieu, Laura Monceaux, Béatrice Daille, and Estelle Dubreil. 2009. "Catégorisation Des Évaluations Dans Un Corpus De Blogs Multi-domaine."
- Vö, Melissa, Markus Conrad, Lars Kuchinke, Karolina Urton, Markus Hofmann, and Arthur Jacobs. 2009. "The Berlin Affective Word List Reloaded (BAWL-R)." *Behavior Research Methods* 41 (2): 534–538.
- Wang, S., and C. D. Manning. 2012. "Baselines and Bigrams: Simple, Good Sentiment and Topic Classification." In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, 90–94. Jeju, Republic of Korea.
- Whitelaw, Casey, Navendu Garg, and Shlomo Argamon. 2005a. "Using Appraisal Groups for Sentiment Analysis." In *Proceedings of the 14th ACM International Conference on Information and Knowledge Management*, 625–631. Bremen, Germany: ACM.
- Whitelaw, Casey, Navendu Garg, and Shlomo Argamon. 2005b. "Using Appraisal Taxonomies for Sentiment Analysis." In Bremen, Germany.
- Wiebe, Janyce. 1994. "Tracking Point of View in Narrative." *Computational Linguistics* 20 (2): 233–287.
- Wiebe, Janyce, Theresa Wilson, Rebecca Bruce, Matthew Bell, and Melanie Martin. 2004. "Learning Subjective Language." *Computational Linguistics* 30 (3): 277–308.
- Wiebe, Janyce, Theresa Wilson, and Claire Cardie. 2005. "Annotating Expressions of Opinions and Emotions in Language." *Language Resources and Evaluation* 39 (2): 165–210. doi:10.1007/s10579-005-7880-9.
- Wiegand, M., A. Balahur, B. Roth, D. Klakow, and A. Montoyo. 2010. "A Survey on the Role of Negation in Sentiment Analysis." In *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*, 60–68.
- Wilson, Theresa, and Janyce Wiebe. 2005. "Annotating Attributions and Private States." In *Proceedings of the Workshop on Frontiers in Corpus Annotations II: Pie in the Sky*, 53–60.
- Wilson, Theresa, Janyce Wiebe, and Paul Hoffmann. 2005. "Recognizing Contextual Polarity in Phrase-level Sentiment Analysis." In *Proceedings of the Conference on Human Language*

- Technology and Empirical Methods in Natural Language Processing*, 347–354. Vancouver, British Columbia, Canada: Association for Computational Linguistics.
- Wilson, Theresa, Janyce Wiebe, and Paul Hoffmann. 2009. “Recognizing Contextual Polarity: An Exploration of Features for Phrase-level Sentiment Analysis.” *Computational Linguistics* 35 (3): 399–433.
- Wilson, Theresa, Janyce Wiebe, and Rebecca Hwa. 2006. “Recognizing Strong and Weak Opinion Clauses.” *Computational Intelligence* 22 (2): 73–99.
- Wolf, F., and E. Gibson. 2006. *Coherence in Natural Language: Data Structures and Applications*. The MIT Press.
- Yu, Hong, and Vasileios Hatzivassiloglou. 2003. “Towards Answering Opinion Questions: Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences.” In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*, 129–136. Sapporo, Japan: Michael Collins and Mark Steedman.
- Zghal, H. B., M. A. Aoufa, and N. B. Mustapha. 2007. “A Model-driven Approach of Ontological Components for On-line Semantic Web Information Retrieval.” *Journal of Web Engineering* 6 (4): 309.
- Zhou, Lanjun, Binyang Li, Wei Gao, Zhongyu Wei, and Kam-Fai Wong. 2011. “Unsupervised Discovery of Discourse Relations for Eliminating Intra-sentence Polarity Ambiguities.” In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 162–171. EMNLP '11. Stroudsburg, PA, USA: Association for Computational Linguistics.
- Zirn, C., M. Niepert, H. Stuckenschmidt, and M. Strube. 2011. “Fine-Grained Sentiment Analysis with Structural Features.” In *Proceedings of the 5th International Joint Conference on Natural Language Processing*, 336–344. Chiang Mai, Thailand.





# Annexes

---

## A.1. Règles de découpage complémentaire pour la segmentation discursive

Comme présenté au chapitre 3, nous avons effectué un sur-découpage des segments discursifs par règles :

- un segment est dit découparable s'il contient au moins une expression d'opinion connue et repérée par lexique
- la découpe s'effectue :
  - o avant un marqueur de discours ou une conjonction de coordination
  - o après une virgule ou un point-virgule

Le repérage des marqueurs de discours s'effectue suivant la liste de marqueurs proposées dans le cadre du projet ANNODIS (Afantenos et al. 2012).

## A.2. Exemple d'utilisation de Myéline : Patrons de reconnaissance des comparatifs et de superlatifs relatifs

Ces patrons, utilisés pour la reconnaissance de subjectivité, ont été conçus pour couvrir les cas du corpus de développement. Il est possible que la couverture ne soit pas optimale sur d'autres types de textes. De même, même si un soin tout particulier a été apporté pour garantir la précision des patrons, il serait nécessaire de vérifier la non apparition de faux positifs sur un corpus différent avant de porter ce code.

Ce code est rédigé en Myéline (cf. chapitre 2 pour plus d'informations sur le langage).

### A.2.1. Reconnaissance de comparatifs

File: comparatifs.myel

```
#define CompPattern(ruleName)

Rule:comp1
{
    (Token:$begin && (Token.lemma=="plus" || Token.lemma=="moins"
                    || Token.lemma=="aussi"))
    (Token.cat=="ADJ" || Token.cat=="ADV")
    (Token)*
    (Token:$end && Token.lemma=="que")
}
=>
```

```
{
    CompPattern($begin.begin, $end.end, #RULENAME);
}

Rule:comp2
{
    (Token:$begin && Token.lemma=="meilleur")
    (Token.cat!="N")*
    (Token:$end && Token.lemma=="que")
}
=>
{
    CompPattern($begin.begin, $end.end, #RULENAME);
}

Rule:comp3
{
    (Token:$begin && Token.lemma=="mieux")
    (Token.cat!="N")*
    (Token:$end && Token.lemma=="que")
}
=>
{
    CompPattern($begin.begin, $end.end, #RULENAME);
}

Rule:comp4
{
    (Token:$begin && Token.cat=="ADV")
    (Token.lemma=="mieux" || Token.lemma=="plus")
    (Token.cat!="ADV")*
    (Token:$end && Token.lemma=="que")
}
=>
{
    CompPattern($begin.begin, $end.end, #RULENAME);
}

Rule:comp5
{
    (Token:$begin && Token.lemma=="plus")
    (Token.cat!="ADV")*
    (Token:$end && Token.lemma=="que")
}
=>
{
    CompPattern($begin.begin, $end.end, #RULENAME);
}

Rule:comp6
{
```

```

(Token:$begin && (Token.lemma=="plus" || Token.lemma=="moins"
                || Token.lemma=="autant"))
(Token.lemma=="de")
(Token.cat=="N")
(Token:$end && Token.lemma=="que")
}
=>
{
  CompPattern($begin.begin, $end.end, #RULENAME);
}

```

```

Rule:comp7
{
  (Token:$begin && Token.lemma=="moindre")
  (Token.lemma=="que")?
  (Token:$end && Token.cat=="N")
}
=>
{
  CompPattern($begin.begin, $end.end, #RULENAME);
}

```

```

Rule:comp8
{
  (Token:$begin && Token.lemma=="pire")
  (Token.cat=="N")
  (Token:$end && Token.lemma=="que")
}
=>
{
  CompPattern($begin.begin, $end.end, #RULENAME);
}

```

```

Rule:comp9
{
  (Token:$begin && Token.cat=="N")
  (Token:$end && Token.lemma=="moindre")
}
=>
{
  CompPattern($begin.begin, $end.end, #RULENAME);
}

```

```

Rule:comp10
{
  (Token:$begin && Token.lemma=="davantage")
  (Token.lemma=="que")
  (Token:$end && Token.cat=="N")
}
=>
{

```

```

    CompPattern($begin.begin, $end.end, #RULENAME);
}

Rule:comp11
{
    (Token:$begin && Token.lemma=="davantage")
    (Token.cat!="N")*
    (Token.lemma=="que" || Tokens.lemma==["être","égal"]
     || Token.lemma=="égaler")
    (Token.cat!="N")*
    (Token:$end && Token.cat=="N")
}
=>
{
    CompPattern($begin.begin, $end.end, #RULENAME);
}

Rule:comp12
{
    (Token:$begin && Token.cat=="V")
    (Token.cat!="ADV")*
    (Token.lemma=="plus" || Token.lemma=="moins"
     || Token.lemma=="autant")
    (Token:$end && Token.lemma=="que")
}
=>
{
    CompPattern($begin.begin, $end.end, #RULENAME);
}

Rule:comp13
{
    (Token:$begin && Token.cat=="V")
    (Token.cat!="ADV")*
    (Token:$end && Token.lemma=="davantage")
}
=>
{
    CompPattern($begin.begin, $end.end, #RULENAME);
}

Rule:comp14
{
    (Token:$begin && Token.cat=="V")
    (Token.lemma=="comme" || Tokens.string==["à","la","manière","de"]
     || Tokens.string==["à","l'","instar","de"])
    || Tokens.string==["à","l'","exemple","de"]
    || Tokens.string==["comparativement","à"])
    (Token.cat!="N")*
    (Token:$end && Token.cat=="N")
}

```

```

}
=>
{
    CompPattern($begin.begin, $end.end, #RULENAME);
}

```

## A.2.2. Reconnaissance de superlatifs relatifs

File: superlatifs.mye1

```

#define SupPattern(ruleName)

Rule:sup1
{
    (Token:$begin && Token.lemma=="le")
    (Token.lemma=="plus" || Token.cat=="moins")
    (Token:$end && Token.cat=="ADV")
}
=>
{
    CompPattern($begin.begin, $end.end, #RULENAME);
}

Rule:sup2
{
    (Token:$begin && Token.lemma=="le")
    (Token:$end && (Token.lemma=="meilleur" || Token.lemma=="pire"))
}
=>
{
    CompPattern($begin.begin, $end.end, #RULENAME);
}

Rule:sup3
{
    (Token:$begin && Token.cat=="P")
    (Token.lemma=="plus" || Token.cat=="moins")
    (Token:$end && Token.cat=="ADJ")
}
=>
{
    CompPattern($begin.begin, $end.end, #RULENAME);
}

Rule:sup4
{
    (Token:$begin && Token.cat=="V")
    (Token:$begin && Token.lemma=="le")
    (Token.lemma=="plus" || Token.cat=="moins")
    (Token:$end && Token.lemma=="de")
}

```

```
}  
=>  
{  
    CompPattern($begin.begin, $end.end, #RULENAME);  
}
```