UNIVERSITE BRETAGNE LOIRE

UNIVERSITÉ DE NANTES

Thèse de Doctorat

Joseph LARK

Mémoire présenté en vue de l'obtention du
grade de Docteur de l'Université de Nantes

sous le sceau de l'Université Bretagne Loire

École doctorale : Sciences et technologies de l'information, et mathématiques

Discipline : Informatique et applications, section CNU 27
Unité de recherche : Laboratoire des Sciences du Numérique de Nantes (LS2N)

Soutenue le 17 octobre 2017

Construction semi-automatique de ressources pour la fouille d'opinion

JURY

Présidente :	**M^me^ Pascale SÉBILLOT**, Professeur, INSA Rennes
Rapporteurs :	**M^me^ Chloé CLAVEL**, Maître de conférences - HDR, LTCI Telecom-ParisTech
	M. Xavier TANNIER, Professeur, Université Pierre et Marie Curie
Directeur de thèse :	**M. Emmanuel MORIN**, Professeur, Université de Nantes
Co-encadrant de thèse :	**M. Sebastián PEÑA SALDARRIAGA**, Docteur Ingénieur, Dictanova

Remerciements

Je souhaite remercier en premier lieu mes encadrants, M. Emmanuel MORIN et M. Sebastián PEÑA SALDARRIAGA, pour leur écoute et leurs conseils avisés dispensés tout au long de cette thèse avec une inébranlable bonne humeur.
Merci également à mes référents lors des comités de suivi de thèse, Mme Pascale SÉBILLOT et M. Pascal MOLLI, ainsi qu'à mes rapporteurs, Mme Chloé CLAVEL et M. Xavier TANNIER, pour avoir accepté d'évaluer cette thèse et pour leurs remarques enrichissantes sur mon travail.

Je remercie M. Fabien POULARD, fondateur et directeur de DICTANOVA, de m'avoir fait confiance pour apporter ma pierre à l'édifice technologique de l'entreprise, tout en me laissant la liberté que demandent des travaux de recherche.
Au cours de ces trois ans (un petit peu plus en réalité), j'ai pu voir DICTANOVA grandir, mûrir, déménager, se remettre en question, (se) déconstruire et (se) rebâtir, en un mot avancer. Je tiens à remercier toutes les personnes qui font ou ont un jour fait partie de cet équipage aux côtés duquel j'ai vécu une formidable épopée.

Merci aux membres de l'équipe TALN ainsi qu'à l'ensemble des équipes du LS2N, anciennement berceau de DICTANOVA, avec lesquels j'ai parfois pu mener des projets ou simplement échanger, et qui ont immanquablement permis de m'ouvrir à de nouvelles perspectives et ainsi donner un second souffle à mes pistes de travail.
Un grand merci à mes camarades de la première promotion du cursus ATAL, sans qui je n'aurais jamais eu l'idée de me lancer dans cette aventure.

Je souhaite remercier bien entendu mes proches et ma famille, qui m'ont soutenu à travers mes doutes, et qui ont aussi su me les faire oublier.
Enfin, tel la colline de Fourvière qui t'accueillais l'hiver pendant tes années lyonnaises, j'aimerais faire briller un grand Merci Marie, pour tout, y compris une relecture *in extremis* de ce tapuscrit.

1

Introduction

Identifier les leviers de satisfaction des consommateurs est aujourd'hui capital dans un monde où la relation que tisse une entreprise avec ses clients est sa plus grande richesse. Le domaine de la fouille d'opinion, dans lequel s'inscrit cette thèse, propose des méthodes permettant de répondre à ce besoin. Celles-ci nécessitent cependant une mise à jour constante de ressources spécialisées qui sont la pierre angulaire des outils d'analyse d'opinion.

Ce travail vise à développer des stratégies d'acquisition et de structuration de ces ressources, qui prennent la forme de lexiques, de patrons morpho-syntaxiques ou de textes annotés. Chacune de ces formes présente des difficultés d'acquisition propres, auxquelles s'ajoute la complexité de mettre à jour ces ressources en fonction de la langue à traiter ou du domaine des corpus analysés.

Dans ce chapitre d'introduction, nous abordons les enjeux de la fouille d'opinion et nous justifions nos motivations, avant de présenter plus en détail les objectifs de la thèse et enfin la structure du manuscrit.

1.1 La fouille d'opinion

Plusieurs ouvrages de référence définissent la fouille d'opinion comme l'équivalent de l'analyse de sentiment, soit le champ d'étude de l'expression des émotions humaines à travers le texte, le son ou l'image [Pang and Lee, 2008, Liu, 2012]. À notre sens, les opinions constituent en réalité une sous-partie de ces émotions, et la fouille n'est ici effectuée qu'au sein de corpus de textes, c'est pourquoi nous considérons la fouille d'opinion comme une discipline spécialisée de l'analyse de sentiment.

Cette discipline, telle que nous l'entendons, est aujourd'hui associée à l'informatique décisionnelle (ou *business intelligence*), c'est-à-dire une aide à la décision automatisée prenant généralement la forme d'un logiciel à destination des entreprises afin d'appuyer leurs choix stratégiques.

Pourtant, voilà bien longtemps que nous sondons la population en vue d'optimiser les échanges commerciaux. C'est notamment cette pratique qui a fortement contribué, au VI^{e} siècle avant notre ère, à l'avènement de la Grèce antique en tant que puissance thalassocratique internationale. Des informateurs situés tout au long du bassin méditerranéen permettaient d'éviter aux marchands grecs des voyages vers des territoires en guerre, peu propices au commerce, ou au contraire indiquaient une région subissant une pénurie ponctuelle, synonyme d'affaires fructueuses.

La notion de fouille d'opinion a évolué à travers les âges, concomitamment aux moyens de communication, et son informatisation correspond de nos jours à notre exploitation rapide et massive de l'information. De plus, il ne s'agit plus aujourd'hui de sonder une population de consommateurs passifs, mais de recueillir l'ensemble des avis et commentaires produits par ces utilisateurs lors d'enquêtes de satisfaction (comme illustré sur la figure 1.1), ou même des interactions spontanées avec une entreprise, par le biais des réseaux sociaux notamment.

FIGURE 1.1 – Exemple d'une demande de participation à une enquête de satisfaction

1.2 Mesurer la satisfaction

Afin de normaliser les avis recueillis et de faciliter l'analyse de ces sondages impliquant une masse de données conséquente, une première approche consiste à considérer l'opinion des consommateurs uniquement sous la forme d'une valeur numérique, telle que le NPS (*Net Promoter® Score*), communément admis en tant qu'indicateur quantitatif de la satisfaction. Le calcul du score NPS, indiqué par la formule 1.1, est effectué à partir de notes correspondant à la propension de l'utilisateur à recommander l'entreprise à un proche sur une échelle de 0 à 10, d'où la dénomination de *promoteurs* des répondants assignant une note supérieure à 8 et de *détracteurs* pour ceux dont la note est inférieure à 7.

$$NPS = \frac{\#Notes > 8}{\#Notes} \times 100 - \frac{\#Notes < 7}{\#Notes} \times 100 \tag{1.1}$$

La fiabilité d'une telle mesure, garantie par la constance de la question posée à l'utilisateur, est cependant à mettre en perspective suivant le domaine concerné. La différence de score visible entre plusieurs entreprises voire plusieurs secteurs d'activités, illustrée sur la figure 1.2 issue de la visualisation de NPS-Benchmark.com[1], n'est en effet pas nécessairement représentative de l'avis des utilisateurs. La manière dont est perçu le service fourni peut en réalité dépendre d'un a priori subjectif. Le bon fonctionnement des services de télécommunications, par exemple, peut être vu comme une normalité et leur dysfonctionnement vécu avec une forte insatisfaction, tandis que les services médicaux peuvent être jugés avec une plus grande clémence.

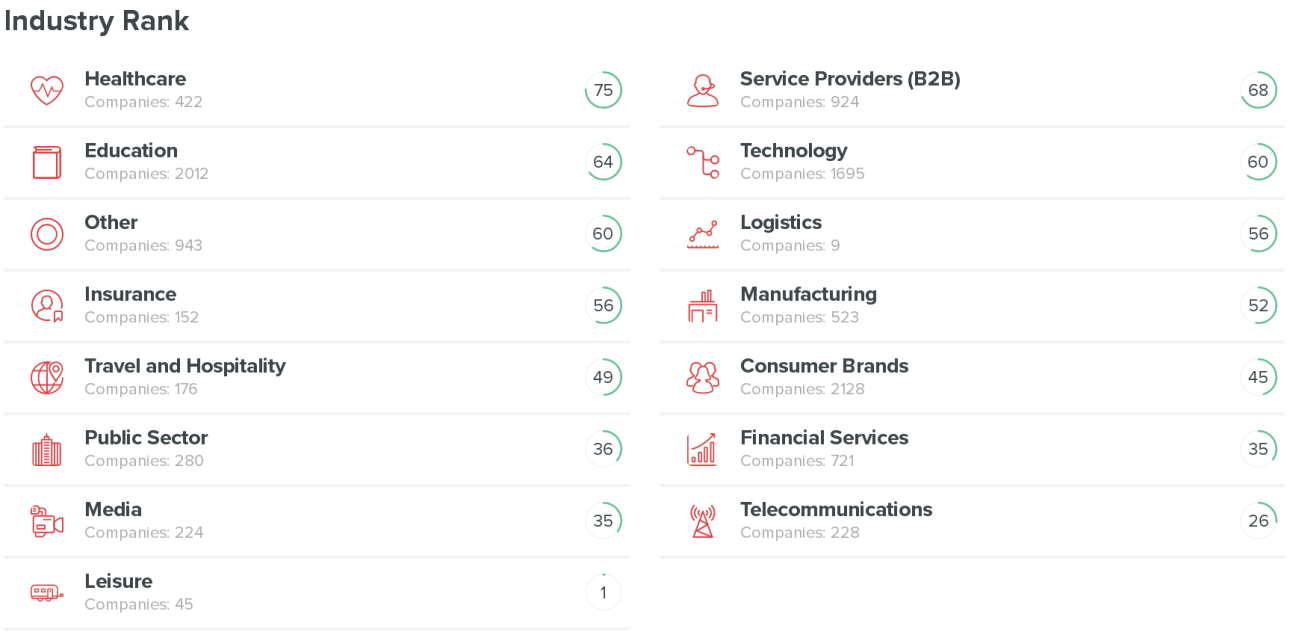


FIGURE 1.2 – Classement NPS des secteurs d'activité selon NPSBenchmarks.com en juin 2017

Par ailleurs, si cette approche permet d'évaluer de façon concise la satisfaction générale des utilisateurs, celle-ci n'apporte en revanche aucun enseignement quant à la source de cette satisfaction, limitant ainsi fortement le champ des possibles dans une optique d'aide à la décision. Il devient de plus en plus évident pour les entreprises soucieuses de l'avis de leurs clients qu'une valeur seule telle que le score NPS ne suffit pas à traduire réellement leurs attentes, c'est pourquoi elles se tournent vers l'analyse des contenus générés par les utilisateurs.

[1] npsbenchmarks.com/companies

1.3 L'opinion dans le texte

Les alternatives à cette note de satisfaction sont multiples et offrent chacune à l'utilisateur un certain degré de liberté. Nous notons parmi celles-ci les indicateurs binaires, les réponses à choix multiples, les questions appelant un commentaire libre mais orienté sur un sujet, et les commentaires totalement libres introduits par une demande neutre (« *Que pourrions-nous améliorer ?* », figure 1.5).

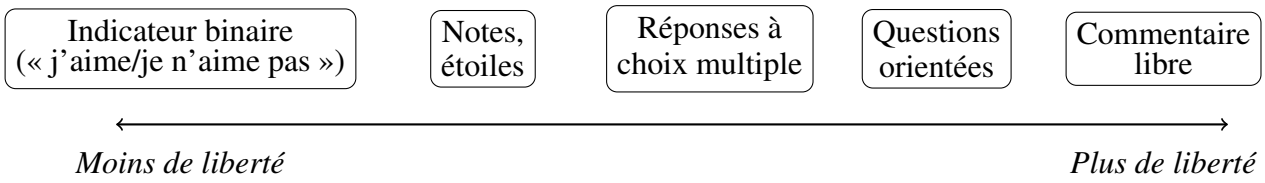
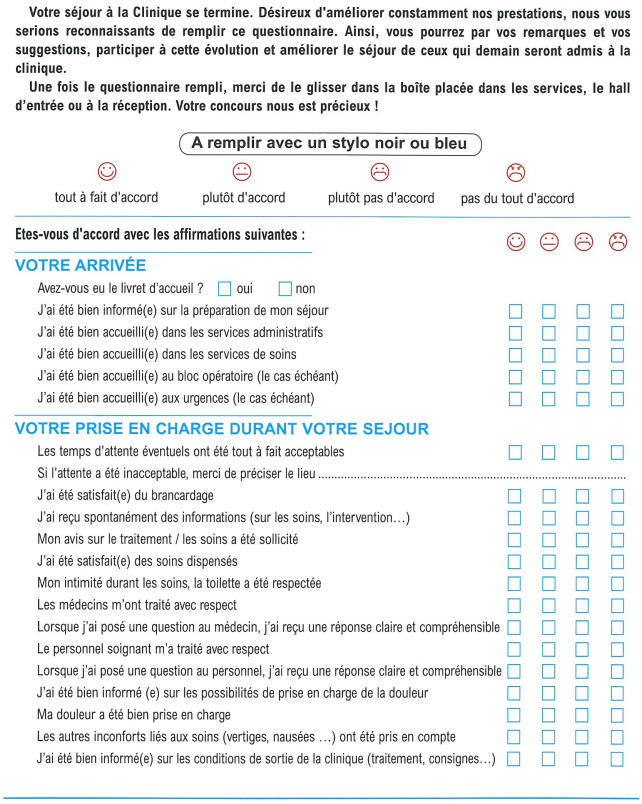


FIGURE 1.3 – Degrés de liberté des demandes d'avis client

Du point de vue des méthodes de fouille d'opinion, cette liberté d'expression est une difficulté majeure, car l'objectif est de retrouver les sujets de préoccupation des répondants au sein de textes non structurés. Cela justifie les travaux importants réalisés sur ce sujet dans le domaine du traitement automatique du langage naturel, propre à résoudre ce type d'extraction de données. En effet la nature bruitée de ces textes, caractérisée par la présence de fautes d'orthographe, intentionnelles ou non, d'erreurs de syntaxe ou encore de formatage du contenu, est un défi à plusieurs étapes de l'analyse du texte, du prétraitement (segmentation des phrases et des mots, catégorisation grammaticale et lemmatisation) à l'extraction de termes et d'opinions. Dans le cadre de notre travail, nous expérimentons uniquement nos méthodes sur des corpus d'avis issus d'enquêtes de satisfaction, éventuellement conditionnés par une question qui en oriente le contenu.



Votre séjour à la Clinique se termine. Désireux d'améliorer constamment nos prestations, nous vous serions reconnaissants de remplir ce questionnaire. Ainsi, vous pourrez par vos remarques et vos suggestions, participer à cette évolution et améliorer le séjour de ceux qui demain seront admis à la clinique.

Une fois le questionnaire rempli, merci de le glisser dans la boîte placée dans les services, le hall d'entrée ou à la réception. Votre concours nous est précieux !

A remplir avec un stylo noir ou bleu

tout à fait d'accord | plutôt d'accord | plutôt pas d'accord | pas du tout d'accord

Etes-vous d'accord avec les affirmations suivantes :

VOTRE ARRIVÉE

- Avez-vous eu le livret d'accueil ? ☐ oui ☐ non
- J'ai été bien informé(e) sur la préparation de mon séjour
- J'ai été bien accueilli(e) dans les services administratifs
- J'ai été bien accueilli(e) dans les services de soins
- J'ai été bien accueilli(e) au bloc opératoire (le cas échéant)
- J'ai été bien accueilli(e) aux urgences (le cas échéant)

VOTRE PRISE EN CHARGE DURANT VOTRE SEJOUR

- Les temps d'attente éventuels ont été tout à fait acceptables
- Si l'attente a été inacceptable, merci de préciser le lieu .....
- J'ai été satisfait(e) du brancardage
- J'ai reçu spontanément des informations (sur les soins, l'intervention…)
- Mon avis sur le traitement / les soins a été sollicité
- J'ai été satisfait(e) des soins dispensés
- Mon intimité durant les soins, la toilette a été respectée
- Les médecins m'ont traité avec respect
- Lorsque j'ai posé une question au médecin, j'ai reçu une réponse claire et compréhensible
- Le personnel soignant m'a traité avec respect
- Lorsque j'ai posé une question au personnel, j'ai reçu une réponse claire et compréhensible
- J'ai été bien informé (e) sur les possibilités de prise en charge de la douleur
- Ma douleur a été bien prise en charge
- Les autres inconforts liés aux soins (vertiges, nausées …) ont été pris en compte
- J'ai été bien informé(e) sur les conditions de sortie de la clinique (traitement, consignes…)

FIGURE 1.4 – Questionnaire à choix multiples

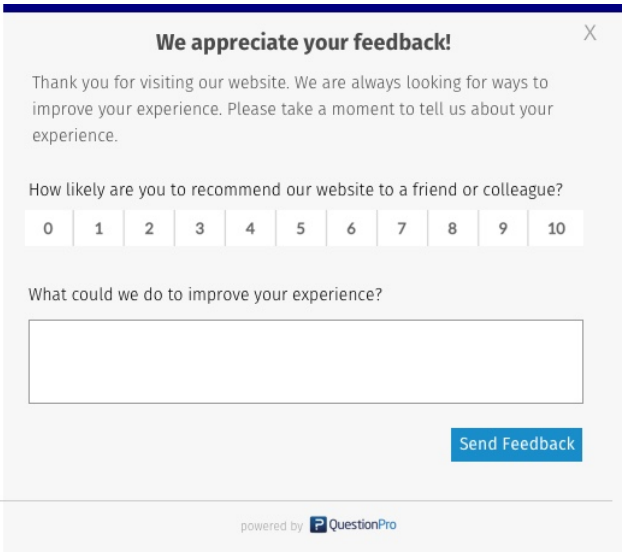


FIGURE 1.5 – Demande d'avis client au moyen d'une note et d'un commentaire libre

1.4 Dictanova

Cette thèse industrielle CIFRE est réalisée en partenariat avec la *startup* DICTANOVA[2], éditrice d'une suite logicielle sur le mode SaaS (*Software as a Service*) dont le cœur de métier est l'analyse des interactions clients.

Créée en 2011 par de jeunes chercheurs issus du Laboratoire d'Informatique de Nantes Atlantique (LINA), et révélée par le Concours national d'aide à la création d'entreprises technologiques innovantes en 2012 puis par l'Award 2013 du forum de l'Industrie Européenne des Technologies du Langage, l'entreprise est imprégnée d'une forte culture scientifique universitaire, particulièrement dans le domaine du traitement automatique des langues. DICTANOVA emploie aujourd'hui 20 personnes de différents métiers et propose aux entreprises une aide à la décision reposant sur les commentaires libres ainsi que sur les mails et les dialogues provenant de chats communautaires, en suivant le principe d'écoute du consommateur.

Bien qu'une valeur importante du produit réside dans l'accompagnement de l'utilisateur dans l'interprétation des résultats, la plate-forme développée par l'entreprise s'efforce de présenter de façon transparente les résultats de l'analyse de ces différents contenus. Cela se traduit dans notre travail par une forte contrainte de pertinence pour les éléments extraits par le système auquel nous contribuons.

1.5 Fouille d'opinion et ressources

L'ensemble des méthodes à travers lesquelles il est possible d'analyser les contenus générés par des utilisateurs nécessitent des appuis linguistiques, que nous qualifions de façon globale sous le terme « ressource ». Ces ressources, dont la création et la maintenance constitue l'objet principal de notre travail de recherche, ne peuvent toutefois pas être décorrélées de leur utilisation, c'est pourquoi une partie substantielle de ce manuscrit ne concerne pas directement ces appuis mais leur effet observé à travers le prisme des méthodes d'extraction d'opinion. Afin d'étudier l'apport des ressources, nous procédons à un raisonnement à contre-courant de cet effet en définissant tout d'abord les éléments que nous souhaitons extraire *in fine*, puis les méthodes permettant de les identifier, ce qui nous oriente enfin vers les questions pertinentes quant à la construction des ressources.

Comme dans le cas de nombreuses autres applications du traitement automatique des langues, la fouille d'opinion peut bénéficier de l'apport de multiples catégories de ressources. Des lexiques, tout d'abord, tels que des listes de mots caractéristiques de la présence d'une opinion associés à une valeur indiquant la tonalité du propos de l'utilisateur, appelées « lexiques affectifs », ou des listes de termes identifiés comme des sujets fréquents de l'opinion.

Les réseaux de connaissances sont également mis à contribution dans certains travaux du domaine. La couverture des expressions générées par des utilisateurs peut être ainsi améliorée en empruntant les liens de synonymie, ou de similarité sémantique au sens large. Parmi ces ressources nous comptons les dictionnaires de synonymes et d'antonymes, les bases de données construites sur le modèle d'ontologies, ou encore des modèles statistiques rapprochant des mots qui partagent des contextes lexicaux similaires, afin d'associer des éléments inconnus à un discours déjà traité.

Enfin, pour la mise en utilisation de modèles probabilistes supervisés, qui prévalent actuellement dans un large éventail d'applications en traitement automatique des langues, il est nécessaire de produire des ressources annotées, c'est-à-dire l'association d'un texte et d'une étiquette précisant l'information portée par celui-ci, ce qui peut être réalisé à l'échelle du document, de la phrase ou du mot. Nous constatons cependant une étrange disproportion entre l'utilisation massive de ces ressources et le faible nombre de travaux que nous avons retrouvé dans la littérature sur le sujet de leur construction.

[2] www.dictanova.com

La maintenance des ressources spécialisées est une tâche relativement coûteuse dans la mesure où il est nécessaire de les adapter à chaque nouveau domaine à traiter. D'une part, les sujets abordés par les utilisateurs, les mots marqueurs d'opinion, ainsi que la structure des phrases contenant une opinion peuvent varier selon la thématique de chaque commentaire. Il faut alors procéder à une désambiguïsation en contexte de ces éléments. D'autre part, de nouveaux indices de l'opinion peuvent émerger d'une nouvelle thématique. Il est alors impératif de pouvoir les détecter et de les associer aux ressources existantes. Dans ce cas, il convient de mettre en œuvre différentes stratégies pour les détecter dans des contextes où les données peuvent être dégradées comme éparses. La difficulté de cette tâche est démultipliée lorsqu'il s'agit de passer à une nouvelle langue.

Compte tenu de ces difficultés, il est possible d'envisager de définir un modèle global qui pourrait prendre en compte l'ensemble de ces caractéristiques. Ce n'est toutefois pas l'objectif poursuivi dans ce travail pour lequel la qualité des ressources prime sur la quantité. Ainsi les informations extraites passeront systématiquement au crible d'un humain qui jugera de leur validité *in fine*. Passer d'une ressource généraliste à une ressource spécialisée peut représenter une réduction majeure de l'information, qui doit ensuite être complétée manuellement en intégrant des connaissances issues de l'analyse semi-automatique du corpus de spécialité. Cette tâche peut rapidement se révéler fastidieuse et chronophage en fonction du domaine et de la taille du corpus.

1.6 Objectifs

Ces difficultés mettent en lumière trois problèmes définissant les objectifs de ce travail : le problème de la modélisation de la fouille d'opinion dans ce cadre industriel, celui du choix des méthodes à utiliser compte tenu des contraintes sur les données à traiter, et enfin le problème de la mise en place rapide de ces méthodes.

Notre objectif premier est de définir le problème de la fouille d'opinion de la manière la plus cohérente pour l'outil Dictanova. Il existe en effet plusieurs formulations de ce problème, et toutes ne permettent pas d'apporter les informations que recherchent les utilisateurs de la plate-forme. Plus qu'un simple choix d'une modélisation existante, il nous semble nécessaire de décomposer les divers éléments qui forment une opinion dans le texte afin de mieux comprendre l'information que ceux-ci peuvent transporter, et de quelle façon ces éléments peuvent être automatiquement identifiés.

Nous cherchons dans un second temps à évaluer différentes méthodes pouvant être appliquées dans ce cadre industriel, et suivant la modélisation définie. L'objectif est non seulement de comparer ces méthodes du point de vue des performances mais également du point de vue des difficultés relatives à leur mise en place, ce qui, comme nous l'avons précédemment mentionné, est un sujet peu présent dans la littérature.

Enfin notre travail consiste précisément à définir des processus pour la mise en place de ces méthodes, c'est-à-dire la génération des ressources nécessaires à leur utilisation. L'objectif de cette génération semi-automatique est d'accélérer la construction de ces ressources, encore largement réalisée de façon manuelle actuellement, tout en conservant un haut niveau de précision. Pour assurer une généralisation possible de cette construction, nous recherchons ici des processus qui ne dépendent ni du domaine du corpus analysé, ni de la langue employée.

1.7 Structure du manuscrit

La présente introduction constitue le **chapitre 1** de ce manuscrit, qui est ensuite composé de trois parties.

La **première partie** est axée autour des questions de modélisation de l'opinion dans le texte, de la définition d'une opinion à la façon d'en aborder l'extraction. Cette étude de la composition des expressions recherchées permet avant tout d'introduire les notions importantes utilisées dans la suite de ce travail et de fournir certaines clés de lecture lors de l'interprétation des résultats d'expérience.

- Dans le **chapitre 2**, nous cherchons à distinguer les éléments qui composent la notion d'opinion au sein de corpus d'avis clients.
- Dans le **chapitre 3**, nous étudions les échelles auxquelles la fouille d'opinion peut s'appliquer et nous proposons une modélisation de la fouille d'opinion adaptée à notre cadre de travail.

La **deuxième partie** concerne les méthodes de détection de l'opinion dans le texte, ainsi que leur combinaison. Après avoir présenté les deux paradigmes de fouille d'opinion que nous exploitons, nous étudions plusieurs approches pour leur hybridation, que nous comparons ensuite sur nos données d'étude. Cette partie permet notamment de définir plus précisément les types de ressource que nous cherchons à construire.

- Dans le **chapitre 4**, nous expliquons la création et le fonctionnement de règles symboliques pour la fouille d'opinion, et nous en listons les intérêts et les limites.
- Dans le **chapitre 5**, nous réalisons une présentation similaire du paradigme probabiliste, c'est-à-dire l'emploi d'un modèle de classification.
- Dans le **chapitre 6**, nous menons une étude des systèmes hybrides existants puis nous expérimentons plusieurs approches afin de déterminer la meilleure combinaison de ces deux paradigmes.

La **troisième partie** réunit nos travaux sur la construction de ressources pour la fouille d'opinion. Ayant établi que nos méthodes d'extraction requièrent un lexique affectif, des règles symboliques et des corpus annotés, nous définissons des techniques pour constituer efficacement chacune de ces ressources.

- Dans le **chapitre 7**, nous expérimentons plusieurs techniques pour la construction et la mise à jour de lexiques affectifs.
- Dans le **chapitre 8**, nous observons dans quelle mesure une liste de règles symboliques de fouille d'opinion peut être automatiquement enrichie ou créée à partir d'un corpus.
- Dans le **chapitre 9**, nous proposons une interface d'annotation facilitant la création de ressources pour les méthodes probabilistes supervisées.

Enfin, nous présentons nos conclusions sur l'ensemble de nos travaux dans le **chapitre 10**.

I

Éléments de la fouille d'opinion

Sachez écouter. Malheur à celui qui, sans la ramasser, laisse tomber une parole d'or de la bouche d'autrui.

Jules Renard

2

Éléments fondamentaux de l'opinion

Il est impossible d'identifier les expressions d'opinion dans un texte sans tout d'abord comprendre ce texte. Dans ce travail, nous commençons par définir la fouille d'opinion parmi des avis émis par des utilisateurs comme une application du domaine de la recherche d'information, puisqu'il s'agit avant tout d'identifier les sujets abordés par ces utilisateurs. Indépendamment de la notion même d'opinion ou de jugement, le fait que ces sujets soient spontanément évoqués est une information riche de sens. Nous présentons tout d'abord dans ce chapitre de quelle manière les éléments fondamentaux de l'opinion que sont les sujets apparaissent dans le texte, puis comment ces sujets peuvent être organisés afin d'initialiser l'application de fouille d'opinion, et enfin en quoi ces étapes font de la fouille d'opinion une application particulière du domaine de la recherche d'information. Dans un deuxième temps, nous développons la spécificité de la fouille d'opinion, c'est-à-dire l'identification des contextes qualifiant un sujet selon une orientation sémantique. Nous définissons les éléments fondamentaux de l'opinion que sont les marqueurs et nous montrons comment ces éléments s'associent aux sujets afin de former les expressions d'opinion.

2.1 Termes en fouille d'opinion

Les termes en fouille d'opinion s'inspirent et se différencient des termes tels qu'ils sont définis dans d'autres domaines du traitement automatique des langues. Partant d'une notion commune, nous spécialisons la définition de terme en fouille d'opinion afin de répondre aux besoins de l'analyse de contenus générés par des utilisateurs et nous listons les difficultés que pose l'extraction de ces termes dans le cadre de l'analyse d'avis clients.

2.1.1 La notion de terme

Clés de la compréhension du texte

Les unités linguistiques que nous appelons termes définissent les clés de la compréhension d'un document, au sens où ceux-ci sont les éléments utiles et nécessaires pour en résumer l'essence d'une part, et permettre de relier des documents partageant des points d'intérêt similaires d'autre part. En linguistique, un terme désigne un concept, soit une chose ou une idée [Sager, 1990], indépendamment de son contexte d'occurrence [Rastier, 1995]. En terminologie, ne sont termes que les occurrences de noms et syntagmes nominaux qui définissent un concept clé dans le domaine du corpus de l'occurrence [Grefenstette, 1993], et en ce sens, l'extraction terminologique vise à construire un lexique spécialisé d'un domaine. Dans le cadre d'une application pratique de l'analyse de documents au moyen de traitement automatique du langage, résumer un texte par ses termes permet d'indexer ces documents selon une organisation reposant sur des critères lexicaux voire sémantiques. Selon une telle organisation, il est non seulement possible de comprendre rapidement le contenu d'un document seul mais aussi le contenu de l'ensemble d'un corpus.

Termes	**Documents**
	Le service est très rapide !
service	*L'ambiance n'est pas géniale mais le service est correct*
	J'ai tout adoré dans ce restaurant : le service comme le menu !
ambiance	*L'ambiance n'est pas géniale mais le service est correct*
menu	*J'ai tout adoré dans ce restaurant : le service comme le menu !*
restaurant	*J'ai tout adoré dans ce restaurant : le service comme le menu !*

TABLEAU 2.1 – Exemples d'avis clients à propos de restaurants et termes extraits associés

L'extraction de termes est par conséquent fondamentale car celle-ci permet de connaître les sujets importants du corpus, ainsi que les documents partageant les mêmes sujets. Lier ces documents, c'est aussi lier leurs auteurs. Dans le cas de l'analyse de documents générés par des utilisateurs, ce sont bel et bien les communautés d'utilisateurs que nous cherchons à comprendre *in fine*, au travers de l'extraction des termes.

Définition dans notre cadre de travail

Si, comme nous le verrons par la suite, la fouille d'opinion partage de nombreux points communs avec les autres applications du domaine de la recherche d'information, il convient de dissocier la notion de *terme* (ou plus exactement de *terme clé*) en recherche d'information et la notion de *terme* en fouille d'opinion. Tandis que la première correspond à un nom ou syntagme nominal orthographiquement correct et dont la pertinence d'extraction est conditionnée par le domaine du corpus, le second correspond plus généralement aux sujets abordés par les auteurs des documents.

Les termes en fouille d'opinion ne sont donc pas limités au domaine du corpus, et plus généralement ne peuvent être contenus dans un ensemble fermé et défini. Il est en réalité primordial pour une application de fouille d'opinion que les sujets abordés, qu'ils portent un jugement ou non, émergent d'eux-mêmes afin de ne pas biaiser l'analyse globale en présupposant de la présence ou de l'absence de certains sujets. Autrement dit, l'objectif de ce type d'analyse est bien d'écouter et non de questionner.

2.1.2 Différentes formes de termes

À l'aspect ouvert de l'ensemble des termes que nous venons de décrire s'ajoute un aspect extrêmement éclectique, dans la mesure où les termes en fouille d'opinion se déclinent en de multiples catégories grammaticales et sous de multiples formes orthographiques.

Noms et syntagmes nominaux

La forme nominale est la forme la plus fréquemment considérée pour les termes en recherche d'information comme en fouille d'opinion, parmi lesquels est faite la distinction entre terme simple, soit un nom seul, et terme complexe, soit un syntagme nominal composé de plusieurs mots.

Un point de difficulté majeur lors de l'extraction de termes est la question de la délimitation des termes complexes. Définir pour un terme quels mots du voisinage inclure ou exclure, afin de déterminer un contexte utile et nécessaire, n'est pas aisé. Par extension, définir une méthode permettant de délimiter un tel contexte pour chaque terme du corpus est une tâche particulièrement ardue.

Il s'agit, dans l'idéal, de trouver pour chaque terme la fenêtre de mots présentant la plus grande pertinence pour l'analyse or cette pertinence est régulièrement décorrélée du simple cadre du traitement syntaxique du document mais est seulement perçue par l'analyste à la lecture des termes extraits. À titre d'exemple, dans la phrase « *Je préfère la livraison des colis le week-end* », nous ne comptons pas moins de six noms et syntagmes nominaux représentant des termes potentiels (« *livraison* », « *colis* », « *week-end* », « *livraison des colis* », « *colis le week-end* » et « *livraison des colis le week-end* »), parmi lesquels il n'est pas évident de sélectionner la combinaison la plus judicieuse.

Verbes et syntagmes verbaux

Du fait de la diversité des expressions employées dans des corpus générés par des utilisateurs, la notion de terme en fouille d'opinion doit être élargie afin de tenir compte de l'ensemble des occurrences possibles des sujets évoqués. Dans cette optique, nous étendons cette notion aux verbes et syntagmes verbaux, car les jugements que nous recensons portent non seulement sur des éléments substantivés mais également sur des actions qui sont exprimées par des verbes. Il est par ailleurs fréquent que la forme la plus commune d'un sujet dans un corpus soit la forme verbale et non son substantif. C'est le cas par exemple du syntagme verbal « *recevoir un colis* », plus naturellement rencontré que le syntagme nominal « *réception d'un colis* » parmi des avis rédigés à propos d'une livraison. Ignorer ces occurrences équivaut par conséquent à délaisser une partie non négligeable des documents à analyser, ce qui peut biaiser les conclusions de la fouille d'opinion sur le corpus.

Expressions idiomatiques

En sus des syntagmes nominaux et verbaux, qui peuvent être identifiés au moyen d'une sélection symbolique ou probabiliste des contextes étiquetés en rôles grammaticaux, nous identifions des sujets dont les occurrences correspondent à des expressions idiomatiques, ou tout du moins des collocations (« *avoir raccroché au nez* », « *être le dindon de la farce* », « *donner un coup de main* »). Ces termes « hors-normes » posent un réel problème dans la mesure où ce type d'expression requiert une reconnaissance lexicale, à la différence des syntagmes nominaux et verbaux, ce qui en complexifie l'extraction et la rend plus profondément dépendante de la langue du corpus. Or, de la même manière que dans le cas des syntagmes verbaux, ne pas prendre en considération ces expressions lors de la fouille d'opinion revient à délaisser une partie des opinions.

Questions incluant une collocation	**Réponses**
Le service était-il **à la hauteur de vos attentes** ?	*Le service a été à la hauteur de mes attentes* *Oui, le service a comblé mes attentes*
La livraison s'est-elle **passée sans encombres** ?	*Tout s'est passé sans encombres* *Ça se passe à chaque fois sans encombres*

TABLEAU 2.2 – Exemples de réponses d'utilisateurs reprenant l'expression d'une question.

En particulier, de telles expressions peuvent apparaître fréquemment lorsque le contenu d'un document provient d'une question, comme dans le cas des enquêtes de satisfaction, qui est un type de corpus que nous traitons ici. À l'image des exemples indiqués sur le tableau 2.2, cette question peut orienter fortement le vocabulaire employé dans la réponse, ainsi lorsque ce vocabulaire comporte des expressions idiomatiques, une partie non négligeable du corpus d'avis en contient.

Orthographe et uniformisation

Enfin, l'éclectisme des termes retrouvés en fouille d'opinion prend une toute autre dimension si nous considérons pour chacune des formes que nous venons de décrire la possibilité de variantes orthographiques. Les méthodes d'extraction de termes reposent généralement sur l'étiquetage grammatical des mots. Or les documents générés par des utilisateurs, de par leur nature bruitée, contiennent des fautes d'orthographe, des erreurs de saisie ou encore d'éventuels tronquages ou erreurs d'encodage lors de la phase de collecte. L'étiquetage syntaxique peut être lourdement affecté par ce type d'erreur, ce qui a une conséquence directe sur la capacité d'un système d'extraction de termes à faire émerger les éléments attendus [Jijkoun et al., 2008, Seddah et al., 2012, Petz et al., 2013].

D'autre part, la notion d'*erreur* orthographique induit la notion de *référence* orthographique à laquelle il est souhaitable de ramener une occurrence incorrecte, car considérer des variantes d'un même sujet comme deux sujets différents peut nuire là encore aux résultats de la fouille d'opinion et de surcroît créer une frustration de l'analyste pour qui le rapprochement est évident. Or cette référence orthographique nécessite par définition une ressource lexicale, qui enchaîne davantage un outil d'analyse réalisant cette correction à une langue particulière.

L'enjeu du rapprochement des différentes occurrences de termes similaires n'est par ailleurs pas limité au cas des erreurs orthographiques mais concerne également les formes que nous avons décrit précédemment. Afin de consolider des sujets identiques il est envisageable de regrouper des termes sous une forme commune [Jacquemin, 1994, Park et al., 2002], au moyen par exemple d'une substantivation dans le cas des syntagmes verbaux, ou plus simplement d'une uniformisation sous l'occurrence la plus fréquente dans le corpus, ce qui présente l'avantage de conserver la forme naturellement rencontrée des termes dans le texte et signifie une lisibilité accrue des résultats. Ce regroupement concerne uniquement les termes partageant une forte similarité orthographique. Les regroupements sémantiques de termes sans graphie commune (à l'instar de « *coût* » et « *prix* ») sortent du cadre du terme, et concernent davantage la notion d'entité comme nous le définissons en section 2.2.2

2.2 Domaines en fouille d'opinion

La notion de domaine est cruciale en fouille d'opinion et de nombreux travaux témoignent de la forte dépendance qui existe entre le discours utilisé spécifiquement dans un domaine et l'extraction des expressions d'opinion [Kanayama and Nasukawa, 2006, Pang and Lee, 2008, Jakob and Gurevych, 2010, Liu, 2012]. Pour autant, son importance n'a d'égale que son manque de définition. Dans cette section nous précisons le périmètre qui correspond au domaine dans notre cadre de travail, et en quoi ce périmètre peut influer sur la qualité de la fouille d'opinion.

2.2.1 La notion de domaine

Un domaine en recherche d'information est avant tout un périmètre conceptuel permettant de rassembler des documents partageant le même sujet principal, ou sujet de plus haut niveau selon une modélisation hiérarchique des concepts d'un corpus [L'Homme, 2004].

Définition dans notre cadre de travail

La difficulté qu'implique cette définition est de délimiter les frontières des sous-domaines spécialisant le sujet principal partagé. Dans le cadre de la fouille d'opinion, nous appelons généralement domaine un secteur d'activité ou un groupe de secteurs d'activités [Liu et al., 2005]. Cette définition est fortement liée à l'application pratique d'outils d'analyse des avis clients, puisque les analystes cherchent à comprendre les points forts et les points faibles d'entreprises de secteurs d'activités particuliers.

Importance du domaine en fouille d'opinion

La notion de domaine impacte de deux façons la fouille d'opinion. Premièrement, certaines expressions d'opinion sont particulières à un domaine et de ce fait, prévoir la possibilité de leur occurrence est une tâche non triviale. Cela se traduit par la nécessité d'étendre les ressources nécessaires à la fouille d'opinion à chaque nouveau domaine rencontré, afin de garantir une couverture optimale des expressions recherchées [Blitzer et al., 2007]. Deuxièmement, cette spécificité requiert, en plus de cette adaptation *ex nihilo*, une adaptation de l'existant [Marchand, 2013].

Dans le texte, la notion de domaine peut changer la valeur d'un mot en tant qu'indicateur de l'opinion car sa connotation ne reste pas identique d'un domaine à l'autre. Les éléments les plus touchés par ces deux phénomènes sont les sujets couverts par les expressions idiomatiques et les mots indiquant la présence d'une opinion, que nous détaillons par la suite.

2.2.2 La notion d'entité

Nous désignons par *entité* un intitulé regroupant un ensemble de termes sémantiquement liés. En ce sens, l'entité recouvre le même usage que le *concept*, tel qu'il est entendu en terminologie ou en ingénierie des connaissances. Le choix de vocabulaire ici est simplement motivé par un usage plus courant dans le domaine de la fouille d'opinion [Liu, 2012].

Cette notion est plus générale que la notion de terme et plus précise que la notion de domaine, dans la mesure où tous les termes d'un domaine ne sont pas sémantiquement proches. Une entité permet d'observer comme un tout l'ensemble des occurrences de termes employés par les utilisateurs pour décrire le même problème ou sujet d'intérêt. Les termes « *roman policier* » et « *livre de recettes* » sont par exemple contenus dans l'entité « *Livre* ». L'apport de cette notion en sus de la seule notion de terme est donc la possibilité de choisir une abstraction plus ou moins importante d'un sujet désigné. Si nous reprenons l'exemple précédent, nous pourrions choisir de spécifier l'entité « *Livre* » en plusieurs entités dont « *Roman* », où figurerait entre autres le terme « *roman policier* », et « *Guides pratiques* » où figurerait entre autres le terme « *livre de recettes* ».

Les entités structurent le domaine

L'entité est en quelque sorte une solution au problème de définition du domaine, car elle permet de moduler le périmètre de celui-ci selon les besoins de l'analyse, et une solution au problème de regroupement des termes afin que ceux-ci représentent des éléments cohérents. De la même façon qu'il paraît naturel de regrouper, par souci de cohérence pour l'analyste, des occurrences de termes ne différant que par une erreur orthographique, il est souhaitable d'associer des termes visant la même entité. Ce type d'association est cependant difficilement automatisable car la cohérence d'une entité peut varier d'une analyse (ou d'un analyste) à l'autre. Les termes « *prix* » et « *coût* », par exemple, représentent une entité cohérente à laquelle sont associés dans certains cas des termes tels que « *note* » ou « *facture* », tandis que dans d'autres cas ces derniers termes seront préférablement rapprochés de l'entité exprimant une idée distincte (celle d'un « *justificatif* » éventuellement).

Plus qu'une simple spécialisation de la notion de domaine, la notion d'entité permet de le structurer afin d'en faciliter l'analyse. Ce n'est véritablement qu'une fois structuré en entités qu'il est possible d'analyser finement un corpus d'un domaine particulier du point de vue de la fouille d'opinion [Liu et al., 2005]. Dans le cadre de l'analyse des avis clients, cela peut se traduire par un rapport de fouille d'opinion détaillé par entité afin de rendre compte des points forts et des points faibles des différents sujets qui constituent un secteur d'activité.

Hiérarchie d'entités

Enfin, la notion d'entité peut être elle aussi déclinée en une hiérarchie comprenant des sous-entités. C'est notamment nécessaire lorsque le projet d'analyse couvre un domaine large, qu'il convient de scinder en parties distinctes. Les entreprises permettant à leurs clients de s'exprimer à plusieurs étapes de leur « parcours d'achat » (visite en magasin, dialogue avec un conseiller ou vendeur, livraison, etc.) ont par exemple ce besoin. Dans chacune de ces parties, ou entités de premier ordre, les enseignements que nous cherchons parmi les documents concernés au moyen de la fouille d'opinion nécessitent de regrouper les termes par souci de cohérence, d'où le besoin de sous-entités, ou entités de second ordre. Dans l'exemple précédent, si les étapes du « parcours d'achat » sont les entités de premier ordre, les entités de second ordre peuvent instancier les différents produits, les vendeurs d'un magasin ou les multiples aspects d'une livraison (le délai, l'état du colis, etc.)

2.3 Fouille d'opinion en recherche d'information

Parce qu'elle repose en grande partie sur l'extraction et la structuration des termes au sein de corpus spécifiques, la fouille d'opinion peut être vue comme une spécialisation de la recherche d'information.

2.3.1 Richesse des termes extraits

Comme nous l'avons mentionné dans les sections précédentes, la fouille d'opinion telle que nous la concevons dans ce travail est limitée à des types de corpus précis : des avis clients doté d'une note, des mails de réclamation ou encore des conversations issues de chat communautaire. Au sein de ces corpus, les utilisateurs s'expriment majoritairement sur des sujets concernant directement l'entreprise ayant mis en place ces services permettant de collecter les retours de leurs clients, c'est pourquoi chaque document analysé et chaque terme extrait est important.

Indépendamment de la notion de sentiment porté par l'opinion, que nous développons dans la suite de ce chapitre, connaître les sujets abordés par ses clients représente d'ores et déjà une aide précieuse pour des entreprises soucieuses de la perception qu'ont les utilisateurs de leurs services. Ceci est d'autant plus vrai si ces sujets sont directement associés à une note, fournissant dès lors une première approche naïve de la fouille d'opinion.

2.3.2 Méthodes de recherche d'information pour la fouille d'opinion

En tant qu'application de la recherche d'information, la fouille d'opinion emprunte des méthodes permettant d'extraire et de rassembler des termes. Ces deux actions peuvent être réalisées par des approches symboliques ou probabilistes.

Extraction de termes

En ce qui concerne l'extraction de termes, ces méthodes s'appuient soit sur une représentation syntaxique de surface des termes, c'est-à-dire une séquence d'étiquettes grammaticales non structurée, tel que NOM (« *prix* », « *service* »), NOM PREP NOM (« *magasin de détail* », « *voiture avec chauffeur* ») ou NOM PREP DET NOM PREP NOM (« *fête de la science à Nantes* », « *demande de la preuve d'achat* »), soit une représentation syntaxique profonde indiquant la structure de la phrase, composée des relations de dépendances qui lient chacun des mots comme illustré sur la figure 2.1.

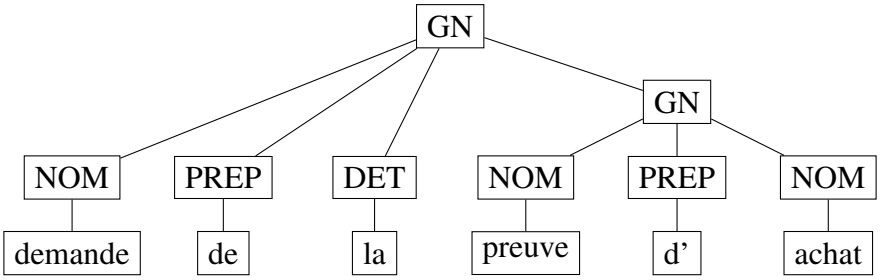


FIGURE 2.1 – Exemple d'un arbre de dépendances syntaxiques pour un syntagme nominal composé

Les méthodes symboliques réalisent une correspondance entre de telles représentations manuellement définies dans une ressource [Bourigault, 1994], tandis que dans le cas des méthodes probabilistes, il s'agit d'inférer ces représentations à partir de corpus dont nous connaissons d'ores et déjà les termes à extraire, autrement dit une ressource annotée [Awasthi et al., 2006, Socher et al., 2010] (nous considérons ici uniquement les méthodes d'apprentissage supervisé).

Association de termes

Nous distinguons trois types de méthodes de regroupement de termes, correspondant aux associations que nous avons décrit précédemment.

Premièrement, les méthodes recherchant une similarité lexicale permettent de regrouper les erreurs orthographiques sous une forme commune de référence. Parmi ces méthodes nous comptons les distances d'édition reposant sur les lettres (distance de Levenshtein [Levenshtein, 1966] et variantes telles que Needleman - Wunch [Needleman and Wunsch, 1970]) ou sur les syllabes (distances phonétiques SOUNDEX, PHONETEX), éventuellement couplées avec l'utilisation d'un lexique afin de disposer de la référence orthographique. Deuxièmement, les méthodes de mise en correspondance de variantes terminologiques [Jacquemin, 1994] permettent de relier des termes désignant le même concept et partageant en partie la même graphie. La différence qu'il peut exister entre ces termes est une variation d'une préposition liant deux noms (« *magasin de Paris* », « *magasin à Paris* »), l'ajout d'un mot (« *magasin de Paris* », « *magasin de Paris Ouest* »), ou encore un mot dont la catégorie grammaticale est différente (« *magasin de Paris* », « *magasin Parisien* »). Enfin, le troisième type de regroupement de termes concerne la recherche d'une similarité sémantique entre des termes ne partageant pas la même graphie, soit au moyen d'une ressource listant des liens de synonymie ou d'hyperonymie [Morin and Jacquemin, 2004], comme une ontologie ou un réseau de connaissance tel que Wordnet [Fellbaum, 1998], soit au moyen de méthodes distributionnelles reposant sur des cooccurrences, tel que l'Information Mutuelle [Church and Hanks, 1990], ou reposant sur une similarité contextuelle, tel que l'algorithme BASILISK [Riloff et al., 2003] ou WORD2VEC [Mikolov et al., 2013].

Méthode dans notre cadre de travail

Dans le cadre de ce travail nous exploitons en priorité une méthode symbolique s'appuyant sur une représentation syntaxique de surface. Nous justifions le choix d'une méthode symbolique par le manque de données annotées en termes d'une part, et par le fait que ce type de ressource doit être sans cesse renouvelée afin d'opérer l'adaptation au domaine nécessaire que nous évoquions précédemment d'autre part. Une approche symbolique permet par ailleurs de réaliser l'extraction terminologique de manière agnostique, c'est-à-dire sans influence d'une annotation pouvant présenter un certain biais. Enfin, le choix de l'analyse en représentation de surface et non en arbres de dépendances est majoritairement lié à la qualité moyenne de ce traitement sur les contenus générés par des utilisateurs.

2.3.3 Termes cibles de l'opinion

Comme nous l'avons vu jusqu'ici, extraire les termes est la première étape de la fouille d'opinion. Tous les termes n'ont cependant pas le même rôle pour ce type d'analyse ; s'il est important de noter chaque occurrence d'un sujet au sein des corpus d'avis, toutes les occurrences de ce sujet ne sont pas des cibles d'opinion. Nous dissocions les occurrences d'un terme dans un contexte neutre (« *Nous nous sommes rendus à ce* ***restaurant*** *hier soir.* ») des occurrences d'un même terme lorsqu'il est qualifié par un avis (« *C'est un* ***restaurant*** *très chaleureux.* »), car ce sont principalement ces cibles d'opinion qui sont en mesure de fournir à l'analyste les clés de lecture quant aux attentes et déceptions exprimées dans les avis.

Identifier les contextes permettant de distinguer ces deux catégories de termes est l'étape centrale de la fouille d'opinion, et en fait sa spécificité en tant qu'application de recherche d'information. La prochaine section est consacrée à la définition de ces contextes.

2.4 Subjectivèmes

Si les termes cibles d'opinion se situent au voisinage d'un ensemble de mots caractéristiques de l'expression d'un jugement, une unique unité linguistique est véritablement le marqueur de l'opinion : le subjectivème.

2.4.1 La notion de subjectivème

La subjectivité dans le texte

Afin de comprendre la notion d'opinion, qui est l'objectif de cette partie, nous revenons dans un premier temps à la notion de subjectivité. Le linguiste français Benveniste en donne la définition suivante [Benveniste, 1958] :

> " *La « subjectivité » dont nous traitons ici est la capacité du locuteur à se poser comme « sujet ». Elle se définit, non par le sentiment que chacun éprouve d'être lui-même (ce sentiment, dans la mesure où l'on peut en faire état, n'est qu'un reflet), mais comme l'unité psychique qui transcende la totalité des expériences vécues qu'elle assemble, et qui assure la permanence de la conscience. Or nous tenons que cette « subjectivité », qu'on la pose en phénoménologie ou en psychologie, comme on voudra, n'est que l'émergence dans l'être d'une propriété fondamentale du langage. Est « ego » qui* dit *« ego ». Nous trouvons là le fondement de la « subjectivité », qui se détermine par le statut linguistique de la « personne ».* "
>
> De la subjectivité dans le langage, in *Problèmes de linguistique générale*, p.258–266

Cette subjectivité, que nous comprenons comme la transparence de l'auteur dans son message, peut-être détectée au moyen de deux types de marqueurs lexicaux appelés déictiques et subjectivèmes. Tandis que les déictiques, qui comprennent un ensemble défini de marqueurs (les pronoms tels que « *je* », « *nous* », ou les

adverbes de lieu et de temps tels que « *ici* », « *là* » ou « *maintenant* »), sont en essence des témoins de cette transparence, les subjectivèmes forment une classe complexe de mots et d'expressions portant l'argument que l'auteur exprime alors même qu'il transparaît dans son récit. La notion de subjectivème proposée en français [Kerbrat-Orecchioni, 1980] et également formalisée en anglais par la théorie de l'*appraisal* [Martin and White, 2003], est un phénomène linguistique qui peut être aussi facilement identifiable que difficilement modélisable et nous n'avons eu cesse de le redéfinir au cours de notre travail. Il est en effet facile, pour un locuteur natif du moins, de reconnaître dans une phrase si un jugement est émis à l'encontre d'un sujet, cependant les éléments linguistiques qui construisent ce jugement dans le texte ne peuvent être aisément désignés, de par la multiplicité de leurs formes ou leur caractère implicite.

Cas des avis clients

Dans le cas des corpus d'avis clients, la transparence de l'auteur dans le message est en réalité un prérequis ; ces messages sont d'ailleurs souvent appelés « *prises de parole* » pour cette raison. L'intérêt des déictiques est donc réduit pour ce type de corpus, c'est pourquoi nous nous concentrons sur la notion de subjectivème qui apporte, elle, des informations supplémentaires sur le discours subjectif de l'auteur et en particulier les éléments de jugement et l'orientation sémantique de son message.

Par ailleurs, une spécification des éléments recherchés peut également être opérée au sein même des subjectivèmes car ceux que nous voyons apparaître parmi les avis clients sont en majorité des marqueurs explicites, ou tout du moins des jugements explicites pouvant contenir des marqueurs explicites ou implicites. En effet, par son message, l'auteur cherche avant tout à être entendu et compris, et rares sont les cas où les avis envoyés présentent une prose sibylline. Toutefois, cette simplification de la classe des subjectivèmes que nous considérons – au demeurant complexe – s'accompagne de la difficulté de les détecter dans des données bruitées, de la même manière que dans le cas des termes que nous évoquions précédemment.

2.4.2 Hétérogénéité des subjectivèmes

Dans la mesure où les subjectivèmes sont des éléments qualifiant les sujets abordés, il est naturel de constater que la catégorie grammaticale majoritairement représentée des subjectivèmes est celle des adjectifs. En effet, les exemples canoniques d'expressions d'opinion suivent le schéma de syntagmes nominaux contenant un adjectif épithète (« *un mauvais accueil* »), ou de syntagmes verbaux contenant un adjectif attribut (« *le vendeur était agréable* », « *joindre le SAV est impossible* »). Ces expressions sont toutefois loin de se limiter à ces schémas, et, à l'instar des termes, les subjectivèmes sont représentés sous de multiples formes.

Catégories grammaticales

En dehors des adjectifs que nous venons d'évoquer, les subjectivèmes peuvent tout aussi bien se présenter sous la forme de noms et syntagmes nominaux (« *une catastrophe* », « *la bonne humeur* »), verbes et syntagmes verbaux (« *aimer* », « *détester* »), d'adverbes (« *bien* », « *malheureusement* ») voire d'interjections (« *Ce film, pff...* ») ou encore de ponctuations (« *Ça, c'est un restaurant ?!* »), surtout si nous incluons dans celles-ci les émoticônes (« *:)* », « *;D* »...).

Cette diversité ne facilite ni l'acquisition de ces marqueurs lors de la construction d'un lexique de subjectivèmes, ressource cruciale pour la fouille d'opinion, ni l'application de ces marqueurs notamment en ce qui concerne la désambiguïsation syntaxique. Ainsi, des homographes peuvent être considérés ou non comme subjectivèmes selon leur catégorie grammaticale :

- l'adjectif « *bien* » est subjectivème, mais ce n'est pas le cas du nom « *bien* »
- le nom « *manie* » est subjectivème, mais ce n'est pas le cas du verbe conjugué « *manie* » (*manier*)
- le verbe conjugué « *recommandé* » (*recommander*) est subjectivème, mais ce n'est pas le cas du nom « *recommandé* »

Expressions

Enfin, il est parfois plus cohérent lors de l'analyse en opinion d'une phrase de considérer l'entièreté d'une expression comme un subjectivème, comme par exemple dans le cas des expressions idiomatiques. Il est fréquent que ce type d'expression soit composé d'éléments qui, pris indépendamment, ne transportent pas l'argument subjectif propre au subjectivème (« *pris la main dans le sac* »), ou pire encore du point de vue de l'ambiguïté de l'analyse, transportent un type d'argument différent (« ***bon** vent !* », « *je me suis **bien** fait avoir...* »).

Importance du contexte

Le problème de cohérence que nous venons de rencontrer pour les expressions peut être en réalité étendu aux contextes des subjectivèmes de façon générale. La propriété subjective des mots et expressions marqueurs d'opinion est non pas inhérente à leur forme lexicale, mais est extrêmement dépendante du contexte dans lequel se trouvent ces marqueurs. Le cas des propositions conditionnelles en est un exemple canonique [Liu, 2012], car celui-ci illustre particulièrement l'aspect volatile de la subjectivité. Ainsi, le jugement positif porté par une proposition telle que « *la qualité est meilleure* », devient négatif dans une proposition conditionnelle (« *si la qualité avait été meilleure...* »).

2.5 Le couple terme–subjectivème

Nous avons jusqu'ici décrit les éléments fondamentaux de la fouille d'opinion que sont les termes et les subjectivèmes de façon parallèle, or il est important de comprendre que ces rôles de sujets et marqueurs de l'opinion fonctionnent de manière tout à fait complémentaire.

2.5.1 La notion d'opinion

Définition dans notre cadre de travail

Nous définissons une opinion comme l'union d'un terme et d'un subjectivème présentant une relation de qualifiant et de qualifié. Dans le cadre de l'analyse d'avis clients, une opinion est une occurrence d'un sujet abordé par un utilisateur pour laquelle il est possible d'identifier un argument précisant le jugement porté sur ce sujet.

En sus de la détection de termes et de subjectivèmes il faut donc procéder à l'identification de couples en relation, ce qui, là encore, est une étape présentant ses difficultés propres lors de l'analyse. En particulier, la distance qui sépare chacun des éléments est très variable : si nous avons donné l'exemple canonique d'un adjectif qualificatif juxtaposant un nom, cela ne couvre en réalité qu'une certaine partie des cas rencontrés. Le subjectivème se trouve ainsi régulièrement éloigné de plusieurs mots du terme qu'il qualifie, comme dans le cas d'une subordonnée (« *ce restaurant, qui est de loin le meilleur, se trouve rue Joffre* ») voire dans une phrase différente du sujet (« *Nous sommes allés jeudi chercher notre commande. Impossible de la récupérer.* »).

Pour autant, cette seule distance ne peut être considérée comme un critère déterminant. Un texte court tel que « *prix correct, emballage beaucoup moins* » en montre l'exemple : le terme « *emballage* » et le subjectivème « *correct* », ici séparés d'une virgule, ne sont pas en relation. Au sein de corpus d'avis clients souvent bruités, cette virgule est même parfois omise, faisant de la désambiguïsation de la phrase une tâche particulièrement complexe.

Dans ce travail nous posons deux hypothèses guidant nos recherches :

- Nous ne considérons que les opinions contenues dans une seule phrase, c'est-à-dire que l'occurrence du sujet abordé ainsi que l'occurrence du marqueur d'opinion doivent se trouver toutes deux dans la même proposition. Plus spécifiquement, ces occurrences doivent être explicites au sens où il ne

peut s'agir, dans le cas du terme par exemple, d'un pronom reprenant par anaphore un terme cité plus explicitement dans une phrase précédente. Dans le cas des subjectivèmes, cet écho peut se manifester par un rappel implicite du jugement (« *La qualité de la photo est extraordinaire. Il en est de même pour la vidéo* »)

- Nous prenons le parti de ne pas nous appuyer sur l'analyse en dépendances de la phrase. Bien qu'il s'agisse d'une méthode spécialisée pour la recherche d'éléments en relation au sein d'une phrase, les expériences préliminaires que nous avons réalisées en exploitant ces outils n'ont pas permis de valider la capacité des modèles actuellement disponibles, en français tout du moins, à s'adapter à la nature bruitée des corpus de documents générés par des utilisateurs.

Ces hypothèses, bien que limitant d'une certaine façon notre champ de recherche, offrent surtout un cadre permettant de mieux définir la notion d'opinion dans ce travail et de consolider nos connaissances dans un périmètre restreint, afin de mieux développer par la suite une définition plus large de la fouille d'opinion.

Opinion sans terme

Comme une règle ne vient pas sans exceptions, notre définition de l'opinion comme l'association d'un terme et d'un subjectivème est mise à mal par les expressions dont l'un ou l'autre est manquant. Les opinions sans terme sont les interjections (« *Génial !* ») ou les propositions dont le sujet est rhétorique, par l'emploi par exemple du pronom « *tout* » (« *Tout était bien* »). Ces expressions d'opinion ne sont pour autant pas dénuées de sujet, seulement celui-ci n'est pas matérialisé par un terme selon la définition que nous en avons donné. L'omission du terme ciblé provient parfois du fait que le sujet global du corpus est implicite. Si les exemples précédents apparaissaient dans des avis sur des restaurants, ce pourrait être l'établissement dans son entièreté qui est jugé.

Nous distinguons deux alternatives pour traiter ces cas dans le cadre de l'analyse d'avis clients. Soit il est effectivement reconnu que le sujet visé est le concept de plus haut niveau dans le corpus, et peut alors être assimilé à ses occurrences plus explicites, soit l'expression est considérée trop ambigüe pour être comptabilisée sur un sujet précis. Si la première solution, que nous détaillerons au chapitre suivant, est celle communément admise dans la modélisation de la fouille d'opinion [Liu, 2012, Pontiki et al., 2015], nous choisissons la seconde dans ce travail. Ce choix est motivé par le souhait de fournir une analyse fidèle des prises de parole recueillies, or supposer du sujet visé par l'auteur d'un avis lorsque celui-ci n'est pas explicite nous paraît être une interprétation qui ne respecte pas les propos de cet auteur d'une part, et qui peut par ce biais fausser l'analyse d'autre part.

Opinion sans subjectivème

Nous avons vu dans la section précédente que les subjectivèmes possèdent une formidable diversité, or un cas particulier de cette diversité est le recouvrement entre terme et subjectivème. Ainsi, certains des exemples de termes donnés en section 2.1.2 (« *avoir raccroché au nez* », « *un coup de main* ») peuvent tout à fait être également considérés comme des subjectivèmes, créant alors une double attribution qui rend floue la notion de rôle dans la structuration de la phrase du point de vue de la fouille d'opinion. Ces expressions sont généralement désambiguïsées au cas par cas, puisque leur utilisation en tant que terme ou subjectivème dépend de plusieurs paramètres subjectifs (leur connotation, leur pertinence pour le corpus analysé). Le nom subjectivème « *erreur* » peut paraître par exemple trop général pour être considéré comme un terme informatif, et conserve donc uniquement son rôle de marqueur d'opinion, tandis qu'un syntagme plus spécifique tel que « *erreur de saisie* » en précise le sens et peut être utile à l'analyse.

2.5.2 Orientation sémantique

Nous avons jusqu'ici parlé de *jugement* porté par les expressions d'opinion, or cette propriété première de l'opinion n'est cependant pas une information suffisante lorsqu'il s'agit de déterminer quelle position prennent les auteurs sur un sujet visé. Cette position, généralement définie selon une polarité ou un type d'émotion, prend la forme d'une valeur appelée orientation sémantique qui est associée au subjectivème. Enfin, quelle que soit la valeur attribuée à la position de l'auteur, celle-ci peut se placer sur un axe d'évaluation représentant un type de jugement.

Polarité

Le premier type de valeur rencontré, aussi bien chronologiquement qu'en ordre d'importance d'utilisation, est la polarité. C'est en général en deux classes, correspondant aux expressions *positives* et *négatives*, que sont rangés les subjectivèmes, et par extension les opinions. Cette simple dichotomie offre l'avantage d'une analyse claire, par exemple en répondant aux questions telles que

- Sur quels sujets les auteurs expriment leur contentement ?
- À propos desquels font-ils part de leur insatisfaction ?
- Y a-t-il plus de sujets de contentement que d'insatisfaction ?

L'approche binaire seule peut cependant montrer quelques limites quant à la finesse d'analyse. Lorsque les expressions employées sont plus ambigües notamment, une unique valeur peut ne pas suffire. L'occurrence du sujet « *service* » dans la phrase « *le service est correct mais un petit peu lent* », par exemple, ne peut être attribuée à une polarité de façon triviale. Pour des cas tels que celui-ci, il est nécessaire d'envisager une troisième valeur représentant une opinion « *mixte* », ou bien d'attribuer à la fois une polarité positive et négative au sujet.

Par ailleurs, alors même que l'orientation sémantique d'un sujet semble évidente entre positif et négatif, il peut être souhaitable de savoir différencier plusieurs nuances au sein d'une même polarité. La phrase « *Le repas était plus que convenable, mais si je reviens dans ce restaurant c'est vraiment pour l'ambiance* » en est une illustration. Ici le sujet de l'« *ambiance* » pourrait être mis en avant dans la mesure où cela semble un critère déterminant pour l'auteur de l'avis. D'un autre côté, la notion de nuance de jugement est extrêmement subjective et il ne s'agit pas ici de commettre l'erreur de sur-interprétation du discours d'un utilisateur telle que nous l'avons exposé précédemment.

Émotion

Une manière alternative d'affiner l'analyse du point de vue de l'orientation sémantique est d'attribuer une valeur à celle-ci parmi une étendue plus large et plus proche des réels sentiments humains, c'est-à-dire les émotions.

Les contributions en modélisation et en classification des émotions définissent de nombreuses classes et sous-classes d'émotions, suivant les travaux de R. Plutchik et E. Ekman, les plus cités en ce qui concerne l'extraction d'émotions [Liu et al., 2003, Généreux and Evans, 2006, de Albornoz et al., 2012, Yang et al., 2014, Cambria et al., 2015]. Ekman, qui a davantage travaillé sur les émotions non-verbales, définit six émotions primaires (la colère, la joie, la surprise, le dégoût, la tristesse et la peur) tandis que Plutchik en propose huit primaires, se déclinant en plusieurs niveaux de combinaisons d'émotions, tel qu'illustré sur la figure 2.2 [1]. La qualité d'analyse obtenue par cette approche est toutefois discutable, d'autant plus si cela est mis en comparaison avec le travail nécessaire en modélisation. Comme en attestent les travaux existants en extraction d'émotions [Yang et al., 2014], cette classification est loin d'être évidente. Nos propres expériences sur le sujet nous ont montré cette difficulté, qui est, de manière naturelle, croissante en fonction de la finesse des classes d'émotions demandées.

[1] Source : Wikipédia (https://fr.wikipedia.org/ wiki/Robert_Plutchik)

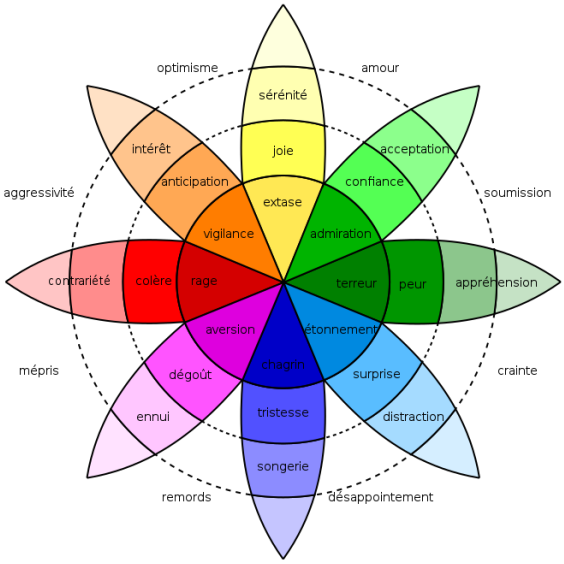


FIGURE 2.2 – Roue des émotions de Plutchik.

Pour ce qui est de l'application de cette extraction aux avis clients, les classes possibles parmi les émotions primaires (de Ekman ou Plutchik) sont relativement restreintes, dans la mesure ou plusieurs émotions sont rarement représentées (la tristesse, la peur...). Dès lors, il parait plus intéressant de rechercher des éléments parmi les sous-classes des émotions les plus présentes (la joie, la colère), ce qui peut complexifier le modèle de classification en rendant les classes plus sensibles, et diminuer la qualité de l'analyse en conséquence. Pour ces raisons, nous n'exploitons pas de modèle d'émotions dans le cadre de ce travail, et nous préférons scinder l'opinion en polarité. Toutefois, afin de mieux décrire le contenu généré par les utilisateurs, nous définissons également des profils types d'avis, que nous détaillons dans le chapitre suivant, permettant à la fois une dichotomie plus claire et une restitution plus riche de la fouille d'opinion.

Axiologie

Enfin, la polarité comme l'émotion portée par un subjectivème peut être précisée par une catégorie axiologique, c'est-à-dire une valeur morale qu'il représente. L'axiologie du discours dans les documents générés par des utilisateurs comprend cinq catégories [Vernier et al., 2011] : l'éthique, l'esthétique, l'affectif, le pragmatique et le cognitif. L'axiologie d'un subjectivème est une information au croisement entre le sujet et le marqueur de l'opinion, dans la mesure où les axes d'évaluation propres à un subjectivème ne peuvent s'adresser qu'à des types identifiables de sujets. Si nous prenons l'exemple d'un avis sur un produit électronique, les axes « *esthétique* » et « *pragmatique* » s'appliquent respectivement aux entités du visuel et de la praticité du produit. En ce sens, identifier l'axiologie des subjectivèmes peut être vu comme un moyen de pallier l'absence de terme telle que nous l'avons détaillée en section 2.5.1.

2.6 Stabilité des subjectivèmes

L'orientation sémantique des subjectivèmes, en particulier la polarité à laquelle ceux-ci peuvent être associés, est, nous l'avons vu, la pierre angulaire de la fouille d'opinion puisque cela représente la spécificité de cette application parmi les autres du domaine de la recherche d'information. Or cette pierre angulaire est très fragile, car l'orientation sémantique d'un subjectivème peut changer d'un corpus à l'autre. Plus précisément, nous distinguons les subjectivèmes *stables*, dont l'orientation sémantique ne dépend pas du domaine, des subjectivèmes *instables* dont non seulement l'orientation sémantique mais également la nature même de subjectivème dépend du domaine du corpus d'apparition.

Il est important ici de comprendre que cette notion de stabilité s'applique de façon absolue à un subjectivème sur l'ensemble d'un domaine et n'entre pas en contradiction avec la notion de dépendance au contexte du subjectivème que nous avons développé en section 2.4.2.

2.6.1 Subjectivèmes stables et instables

Instabilité de la subjectivité

Nous avons déjà évoqué l'importance du domaine en fouille d'opinion. Parmi les liens fragiles que la fouille d'opinion tisse avec la notion de domaine, nous retrouvons l'instabilité de la subjectivité. Cette instabilité se traduit dans le texte par le fait que l'occurrence d'un mot ou d'une expression représentant la même entité linguistique (*i.e.* même étiquetage grammatical et même sens) peut être un subjectivème pour un domaine, mais pas pour un autre. Certains domaines sont plus touchés par ce phénomène, à l'image des critiques de littérature ou de cinéma :

- les subjectivèmes désignant aussi des genres littéraires : « *merveilleux* », « *fantastique* », « *horreur* »
- les subjectivèmes employés aussi dans le domaine du cinéma : « *un film catastrophe* », « *un scénario rocambolesque* »

C'est le cas pour un grand nombre d'expressions imagées, souvent usitées uniquement dans le cadre d'un domaine spécifique, mais dont la forme de surface peut aussi apparaître dans des corpus différents. Les avis sur la grande distribution, par exemple, nécessitent une désambiguïsation des expressions imagées sur la nourriture, souvent associés à un aspect négatif : « *un navet* », « *un panier de crabes* », « *des salades* » ; les termes du domaine de la joaillerie, pour en donner l'exemple inverse, ont généralement une connotation positive : « *un bijou* », « *un diamant* », « *une perle* ».

Dans bien des cas la désambiguïsation de ces expressions problématiques est résolue en imposant un contexte élargi lors de la détection de celles-ci, afin de dissocier une occurrence subjective d'une occurrence neutre. Cette spécification du subjectivème n'est pas une étape simple puisque cela nécessite de définir un contexte suffisamment précis pour réussir cette désambiguïsation, et suffisamment large pour couvrir un grand nombre des occurrences du subjectivème. D'autre part, il peut exister un certain nombre de ces contextes pour un seul subjectivème, notamment lorsque celui-ci n'appartient pas à une expression idiomatique mais peut être en cooccurrence avec de multiples mots dans des expressions subjectives, ce qui décuple le travail requis lors de la construction du lexique affectif.

Instabilité de la polarité

Le deuxième phénomène d'instabilité que nous rencontrons est le changement de polarité. Certains subjectivèmes sont en effet stables du point de vue de la subjectivité, au sens où ils conservent leur rôle de marqueur d'opinion quelque soit le domaine d'apparition, cependant l'orientation sémantique de ces marqueurs n'est pas identique dans tous les domaines. Ce problème d'adaptation au domaine a été largement étudié [Blitzer et al., 2007, Read and Carroll, 2009, Choi and Cardie, 2009, Jakob and Gurevych, 2010], y compris en français [Garcia Fernandez and Ferret, 2012, Marchand et al., 2014].

Les subjectivèmes sur lesquels ce phénomène est le plus observé sont les adjectifs de mesure (« *grand* », « *petit* », « *énorme* », « *rapide* », « *long* »...). Au contraire des subjectivèmes instables en subjectivité, les marqueurs instables en polarité peuvent difficilement être spécifiés par un contexte plus large, car c'est bien le sens intrinsèque du mot ou de l'expression qui change suivant le domaine du corpus d'apparition. Par conséquent, ces éléments requièrent une *activation* par domaine. Cela en fait un problème d'un ordre de grandeur supérieur au précédent car cette activation conditionnelle implique de maintenir une ressource spécifique par domaine.

2.6.2 À la recherche d'une stabilité

Dans la section précédente, nous avons exposé l'instabilité inhérente aux marqueurs de l'opinion, et présenté quelques pistes pour y remédier. La recherche de stabilité est une problématique majeure en fouille d'opinion, tant du point de vue de la recherche en traitement automatique des langues (Comment donner une cohérence aux différentes expressions d'opinions ? Comment lier le discours subjectif dans différents domaines ?) que de celui des motivations industrielles. La multiplicité des ressources nécessaires est effectivement un frein au développement d'une application de fouille d'opinion, d'autant plus si cela doit être répété dans plusieurs langues (qui ne partagent pas nécessairement les mêmes marqueurs instables). L'adaptation au domaine est également un point de difficulté du point de vue de la qualité perçue par un utilisateur d'une application de fouille d'opinion car ce qui peut être un subjectivème simplement non adapté est vu comme une grave erreur d'analyse.

Stabilité des couples terme–subjectivème

En dehors de ces palliatifs s'appliquant directement sur le lexique affectif, les solutions au problème de l'instabilité des subjectivèmes, et subséquemment des opinions, sont à construire autour des couples terme–subjectivème. Dans les faits, si l'orientation sémantique des marqueurs d'opinion dépend du domaine du corpus dans lequel ils apparaissent, cette information est surtout guidée par le terme qui est cible du jugement porté par le subjectivème.

Autrement dit, si nous observons dans le détail le problème d'adaptation au domaine, nous nous apercevons qu'il s'agit d'une adaptation à un ensemble de termes, ces termes étant certes fortement conditionnés par le domaine. À notre sens, il est par conséquent plus judicieux de parler d'une polarité du couple terme–subjectivème lors de l'analyse des opinions d'un corpus, car cela décrit non seulement l'instabilité des subjectivèmes d'un domaine à l'autre, mais également l'instabilité des marqueurs d'opinion au sein d'un même corpus. [Liu, 2012] cite l'exemple de l'adjectif « *long* » au sein d'avis sur des produits électroniques, positif si le sujet visé est la durée de la batterie ou la durée de vie du produit, mais négatif si ce sujet est le temps d'allumage par exemple.

Il est extrêmement rare que des couples terme–subjectivème soient instables au sein d'un même corpus. Les occurrences dont nous pouvons donner l'exemple concernent des expressions particulièrement ambigües dont l'orientation sémantique dépend du *ton* de l'avis, ou même de la perception propre à l'auteur de cet avis, ce qui ne peut être déduit du texte seul. Citons pour illustrer ceci le cas de l'adjectif « *rapide* » dans le domaine de la restauration :

- « *Parfait ! service très rapide !* » (avis associé à une note haute)
- « *Le service est très rapide, on a à peine eu le temps de s'installer...* » (avis associé à une note basse)

Dans le cadre de ce travail, nous veillons par conséquent à construire et à maintenir des ressources telles que les lexiques affectifs en nous appuyant sur les couples terme–subjectivème, éventuellement après un amorçage de la ressource à l'aide de subjectivèmes stables seuls.

Modélisation de la fouille d'opinion

Nous l'avons vu au cours de cette section, la notion d'opinion comprend des éléments fortement instables et pourtant primordiaux pour mener à bien l'analyse. C'est entre autres pour canaliser ces éléments qu'il est nécessaire de proposer une modélisation de la fouille d'opinion, permettant à la fois de s'adapter à l'ensemble des corpus traités et de proposer une expressivité représentant tous les cas d'expressions d'opinion rencontrés. Les différentes modélisations possibles, que nous détaillons dans le chapitre suivant, ont pour objectif de proposer un cadre cohérent au sein duquel les résultats de la fouille d'opinion sont stables.

2.7 Synthèse

Dans ce chapitre, nous avons détaillé les définitions basiques de la fouille d'opinion, et nous avons précisé dans quelle mesure celles-ci s'appliquent au contexte d'une application analysant les contenus générés par des utilisateurs. Nous avons expliqué que la notion d'opinion repose sur deux éléments fondamentaux du discours :

- le **sujet** abordé par l'utilisateur, et à propos duquel celui-ci exprime une satisfaction ou une insatisfaction, correspondant dans le texte aux termes et expressions-clés généralement sous la forme de noms, de syntagmes nominaux et verbaux ;
- le **marqueur d'opinion**, outil linguistique polymorphe dont l'utilisateur se sert pour qualifier le sujet abordé, et qui fournit à l'expression d'opinion son orientation sémantique, soit la valeur du jugement porté, généralement dépeinte de façon binaire sous la forme d'une polarité, au sens où un jugement est *positif* ou *négatif*.

Ces deux éléments sont indissociables de la notion de **domaine**, soit le thème global des documents analysés, dans la mesure où ce thème définit le champ lexical que nous retrouvons parmi les sujets abordés, et influe fortement sur l'orientation sémantique des marqueurs d'opinion. Ces considérations nous mènent à définir une opinion comme l'association d'un sujet et d'un marqueur pour un domaine donné.

Dans le contexte d'une application traitant des avis recueillis et restituant une analyse à un utilisateur, la recherche de ces associations est confrontée à certains obstacles :

- le contenu des avis recueillis est **bruité**, c'est-à-dire que le langage qui y est employé comporte des erreurs, ce qui entrave l'identification des sujets, des marqueurs d'opinion et de leur lien ;
- les utilisateurs de cet outil de fouille d'opinion n'ont pas tous la même **perception de la restitution** des sujets abordés, celle-ci doit donc pouvoir s'adapter à leur besoin ;
- la spécificité au domaine des sujets et marqueurs d'opinion induit une **adaptation des ressources linguistiques** à chaque nouveau domaine traité.

3

Granularité de la fouille d'opinion

Une manière naturelle d'aborder les différentes modélisations de l'opinion dans le texte est de les concevoir sur différentes échelles. Dans ce chapitre nous étudions les intérêts et les difficultés de la fouille d'opinion à l'échelle d'un document, d'une phrase, d'un mot ou d'une expression. Nous décrivons en quoi les éléments fondamentaux de l'opinion, soit les sujets et marqueurs vus précédemment, sont indissociables de la fouille d'opinion quel que soit le niveau d'analyse et comment ceux-ci peuvent être exploités au mieux pour rendre compte des opinion énoncées par les utilisateurs. Enfin, nous proposons une modélisation adaptée à la production d'un résumé d'opinion en abordant d'une nouvelle manière le problème de la dépendance au domaine de la fouille d'opinion.

3.1 Fouille d'opinion à forte granularité

La fouille d'opinion à forte granularité désigne l'ensemble des méthodes cherchant à inférer l'orientation sémantique de prises de parole dans leur entièreté. Ces méthodes considèrent principalement un document ou une phrase comme unité textuelle porteuse de jugement. Nous indiquons ici certains des travaux fondateurs de cette approche qui ont permis d'orienter notre travail.

3.1.1 Travaux existants

Fouille d'opinion à l'échelle du document

Afin de trier les commentaires hostiles (ou *flames*) parmi des messages échangés entre utilisateurs, [Spertus, 1997] expérimente une méthode reposant sur l'identification de types d'expression caractéristiques de l'hostilité, lesquels sont modélisés par des règles pour être utilisés par un arbre de décision. Si ce système parvient effectivement à réduire le taux de *flames* parmi l'ensemble des messages, les ambiguïtés sémantiques, par ailleurs soulevées par l'auteur lors de la définition des expressions caractéristiques, induisent des erreurs de classification. [Turney, 2002] propose d'inférer l'orientation sémantique des avis provenant de plusieurs domaines (avis sur des automobiles, des banques, des films et des lieux touristiques) en s'appuyant sur l'information mutuelle spécifique (*Pointwise Mutual Information*), une mesure de pertinence de cooccurrence. Cette mesure est établie entre les bigrammes d'un document dont un des deux mots est un adjectif ou un adverbe et un ensemble de mots positifs, puis un ensemble de mots négatifs. La plus grande similarité atteinte avec l'un ou l'autre de ces ensembles définit la polarité du document. [Pang et al., 2002] présente également une classification de documents en polarité, cependant les auteurs comparent des méthodes d'apprentissage automatique (*Naive Bayes*, *SVM*) avec une approche sac-de-mots, en s'appuyant sur l'ensemble des mots ou bigrammes de mots du document indépendamment de leur ordre d'apparition, à la manière d'une catégorisation de textes. Plus récemment, les travaux en classification de documents selon une notion de sentiment ont connu un regain d'intérêt avec l'avènement des réseaux sociaux. Les ateliers SemEval ABSA 2014, 2015 et 2016 (*Aspect-Based Sentiment Analysis*) [Pontiki et al., 2014, Pontiki et al., 2015, Pontiki et al., 2016] ont ainsi proposé une tâche évaluée et des ressources annotées pour des systèmes de classification en polarité de tweets [1] en anglais. L'atelier DEFT 2015 (Défi Fouille de Textes) [Hamon et al., 2015] a proposé le même type d'évaluation en français.

Fouille d'opinion à l'échelle de la phrase

Afin de réduire le risque d'ambiguïté sémantique, une seconde approche consiste à réduire le périmètre de l'analyse à chaque phrase du document. [Wiebe et al., 1999] s'intéresse à la détection de phrases présentant un caractère subjectif. Le système proposé est un modèle classification supervisé utilisant des traits de classification correspondant aux déictiques, tels que les pronoms ou certains adverbes et adjectifs. [Yu and Hatzivassiloglou, 2003] séparent des phrases comportant une opinion de phrases présentant un fait dans un corpus d'articles journalistiques. Une différenciation est faite entre phrases négatives, positives, mixtes

[1] Messages échangés sur la plateforme sociale Twitter.

(comportant une opinion positive et une opinion négative) et ambigües (dont la polarité est non triviale). Les traits de classification utilisés sont très similaires à ceux d'une méthode de catégorisation de texte, soit les n-grammes de mots ainsi que leur catégorie grammaticale, auxquels est ajoutée la polarité d'adjectifs présents dans la phrase, préalablement définie de façon manuelle. En reprenant ces bases, [Kim and Hovy, 2004] introduisent les notions d'expression d'opinion (au sens d'un segment de phrase où l'opinion est exprimée) et d'émetteur de l'opinion, pouvant être catégorisé dans l'une des deux catégories d'entités nommées PERSONNE ou ORGANISATION. La polarité des marqueurs d'opinion utilisés pour classifier ces phrases est inférée par synonymie au moyen du réseau de connaissance WORDNET [Fellbaum, 1998] à partir d'un petit nombre de marqueurs constitués à la main, en suivant le raisonnement selon lequel les synonymes d'un mot portent la même orientation sémantique, et les antonymes en portent l'orientation opposée. Comme l'expliquent les auteurs, ce raisonnement présente une limite due à la dérive sémantique de la recherche de synonymes, c'est-à-dire le fait que le sens du mot initial, et par extension son orientation sémantique, est peu à peu perdu au fil des connections dans le réseau de connaissance. Dans ce travail, une solution proposée à cette limite consiste à mesurer le degré d'association *positif* et *négatif* de chaque synonyme – seuls les mots présentant une forte association avec une seule polarité sont considérés comme des marqueurs d'opinion pertinents. [Täckström and McDonald, 2011] proposent un moyen de réduire le travail manuel nécessaire à l'élaboration de ressources annotées pour la fouille d'opinion à l'échelle de la phrase en employant un étiquetage de séquence sur les phrases d'avis clients dont la polarité est considérée connue car inférée d'une note indiquée par l'auteur de l'avis. Enfin, [Wang and Cardie, 2014] détectent les messages échangés entre utilisateurs faisant l'objet d'une dispute en utilisant une méthode de détection similaire. Il est intéressant de noter que les auteurs sont toutefois en désaccord sur le fait que les traits de classification utilisant un lexique affectif sont bénéfiques, ou du moins nécessaires.

3.1.2 Polarité d'une phrase ou d'un document

Une application limitée

Les principaux objectifs de la fouille d'opinion à granularité forte sont soit d'évaluer l'orientation sémantique d'un corpus, en retrouvant automatiquement une évaluation globale du sujet principal, soit d'identifier les documents d'une orientation sémantique particulière au sein de ce corpus.

Or cette orientation sémantique à l'échelle du document n'est triviale que lorsque le document est pleinement positif ou pleinement négatif, cependant nous avons vu que ce n'est régulièrement pas le cas. Comment, dès lors, juger de la polarité d'un document s'il comporte des jugements de polarités différentes ? Attribuer la polarité la plus présente, à l'image de [Turney, 2002] par exemple, suppose que tous les sujets et subjectivèmes ont le même « poids » pour l'orientation sémantique. En dehors du cas limite d'une égalité entre positifs et négatifs, cela pose une nouvelle fois la question de l'interprétation du discours de l'auteur de l'avis, car ces sujets n'ont pas nécessairement une importance égale et cette répartition n'est certainement pas identique pour chacun des auteurs.

[Pang et al., 2002] illustre cette répartition inégale par l'exemple des critiques de films, dans lesquelles les utilisateurs sont régulièrement amenés à concéder les qualités (ou respectivement, les défauts) d'une œuvre puis donner un jugement en opposition avec cette liste de points positifs (respectivement, négatifs), tel que dans l'exemple suivant :

> *This film should be brilliant. It sounds like a great plot, the actors are first grade, and the supporting cast is good as well, and Stallone is attempting to deliver a good performance. However, it can't hold up.* (noté 2/5)

Dans cette modélisation à granularité forte, la définition de l'opinion comme l'association d'unités linguistiques s'applique difficilement car, bien que l'analyse s'appuie également sur des indices précis de l'opinion dans le texte, tels que les marqueurs d'opinion ou les déictiques, un seul sujet global peut être considéré par document (dans l'exemple précédent, il s'agit du film, et non des acteurs ou du scénario,

pourtant cités) et il en est de même de l'orientation sémantique du jugement qui lui est porté. Un premier pas vers une granularité plus fine de l'analyse est de considérer l'opinion à l'échelle de la phrase. Réaliser la fouille d'opinion à ce niveau permet de conserver une certaine cohérence du discours, et de qualifier la phrase entière comme pertinente ou non pour l'analyse, en écartant par exemple les propositions conditionnelles ou interrogatives qui ont généralement pour effet de rendre ambigüe l'orientation sémantique des opinions.

Cependant, si chaque proposition est effectivement cohérente du point de vue du discours, nous y retrouvons de multiples combinaisons d'orientation sémantique associées à des sujets. Un avis peut ainsi comporter plusieurs sujets de même orientation sémantique (1), plusieurs sujets portant des jugements différents (2,3) ou encore plusieurs jugements différents portés sur un même sujet (4).

1. *La **commande** était incomplète et l'**article** reçu pas celui commandé*
2. *Le **livreur** n'a pas sonné et a simplement déposé le **colis** devant ma porte !*
3. ***Colis** parfait, par contre **amabilité du livreur** à revoir*
4. ***Commande** envoyée rapidement, mais pas complète...*

Exploiter les éléments d'opinion à granularité fine pour réaliser une fouille d'opinion à granularité forte, c'est-à-dire retrouver une opinion unique comme dans le cas des travaux cités en section 3.1.1, semble contradictoire avec la fouille d'opinion parmi des avis clients. Il est important de garder à l'esprit l'objectif final de l'application de fouille d'opinion que nous développons, qui est de comprendre les besoins, les sources de satisfaction et d'insatisfaction des utilisateurs. Il s'agit d'une tâche complexe pour laquelle, à l'image d'un puzzle, chaque pièce contribue à la compréhension de l'image globale, ce qui se traduit dans notre cadre de travail par l'ensemble des enseignements à extraire du corpus.

Or, par définition, la fouille d'opinion à granularité forte ne permet pas cette fragmentation de l'information. Les enseignements fournis par ce type d'analyse sont par conséquent inexploitables pour rendre compte des arguments dispersés dans le corpus, dans la mesure où le contenu même du corpus n'est pas pris en considération, au sens où il n'est pas question ici de l'expliquer, mais simplement de se servir du discours comme un élément différenciant parmi les différents documents. Du fait de cette simplification trop importante de l'analyse, il est impossible de comprendre les opinions des utilisateurs, et par extension les pistes à suivre pour améliorer la relation d'une entreprise avec ceux-ci.

Une annotation accessible

Nous venons d'évoquer la difficulté d'attribuer une opinion unique à une phrase ou à un document. Pour autant, il est généralement demandé à l'utilisateur, en sus de son avis rédigé, de soumettre une note résumant son expérience globale. C'est par ailleurs cette notation qui rend d'une certaine façon la fouille d'opinion à granularité forte caduque, car l'information est alors non seulement donnée, mais – hors erreur de saisie – elle est aussi de qualité car c'est l'auteur lui-même qui, en attribuant la note, donne son véritable jugement prenant en compte un certain nombre de paramètres qui peuvent être invisibles dans le texte, telle que l'importance que cet auteur accorde à chacun des sujets abordés et non abordés. Pour un système reposant sur un apprentissage supervisé, la présence d'une note est un moyen rapide de construire une ressource annotée.

La possibilité d'une telle acquisition est rare pour une application de classification de textes. Parmi les corpus de contenus générés par des utilisateurs, peu d'indices sémantiques sont ainsi requis de façon systématique. À notre connaissance, les autres types d'annotation « spontanée » prennent lieu sur les réseaux sociaux, où les utilisateurs indiquent par exemple un sujet commun avec une balise spécifique (*hashtag* (#) sur Twitter, sélection d'une catégorie dans les forums) ou encore emploient des émoticônes (« :) », « :(», « ;-D »...) en tant qu'annotation spontanée de l'émotion. [Pak and Paroubek, 2010] extraient ainsi un corpus

de tweets contenant des émoticônes afin d'étudier des phénomènes linguistiques récurrents parmi ce type de contenu et de construire automatiquement un modèle de classification de tweets en polarité.

Dans un corpus d'avis clients, un même procédé peut être opéré à partir des documents portant une note extrême. Puisque les notes et les émoticônes ne dépendent pas de la langue, ceci est un moyen d'obtenir un corpus trié selon l'orientation sémantique dans une langue inconnue.

3.1.3 Autres applications

Nonobstant cet intérêt limité pour la fouille d'opinion à proprement parler, d'autres applications de l'analyse à l'échelle du document peuvent être utiles à la compréhension d'un corpus d'avis clients. Ainsi, il est utile à une entreprise de comprendre les types d'avis exprimés par leurs clients.

Suggestions

Nous désignons par « suggestion » une expression de souhait décrite par un utilisateur, qui rend compte d'une décorrélation entre l'attente et l'expérience vécue d'un service. Des indicateurs de ce type d'expression sont l'emploi du conditionnel, ou de la notion de manque :

- *J'aurais aimé des conseils plus précis...*
- *Pas assez de choix au rayon livre !*
- *Si le magasin avait été plus propre, je me serais arrêté*

Du point de vue des entreprises souhaitant analyser les retours écrits de leurs clients, ces suggestions ont une valeur inestimable car celles-ci représentent un moyen très direct d'orienter des choix difficiles lorsqu'il s'agit de redéfinir leurs produits, leurs services ou leurs processus. L'extraction d'avis contenant une suggestion peut tirer parti des méthodes de classification de phrases car d'une part ces expressions sont parfois exprimées de façon implicite, et d'autre part la prise en compte de la phrase entière apporte un supplément de contexte non négligeable pour cette application.

Attrition

Le spécialiste en stratégie d'entreprise F. Reichheld évalue qu'un client perdu est bien plus impactant qu'un client acquis [Reichheld, 2001]. Selon l'auteur, réduire de 5% le taux d'attrition, c'est-à-dire la proportion d'utilisateurs se séparant des services de l'entreprise, permet d'augmenter les revenus de l'entreprise de 25 % et plus. Cela s'explique notamment par une réduction du besoin d'acquisition de nouveaux clients (dont le coût est important) et par la tendance des clients plus loyaux à dépenser davantage. L'identification des clients présentant un risque d'attrition est par conséquent une pratique vers laquelle se tournent bon nombre d'entreprises. Une méthode manuelle utilisée consiste à parcourir les avis négatifs portant une note extrême (0/5, 0 à 1/10, etc.) et de relever parmi ces avis ceux exprimant le souhait de résilier une offre, ou encore d'acheter un produit chez un concurrent. Cette méthode est fastidieuse et présuppose que les clients concernés sont cloisonnés dans le périmètre des avis portant cette note basse.

Une classification à l'échelle de la phrase peut ici grandement aider ces entreprises en reconnaissant automatiquement ces clients en risque d'attrition au travers de leur discours, dans la mesure où certaines expressions caractéristiques de l'attrition sont récurrentes :

- « aller ailleurs » : *Si le service ne s'améliore pas, j'irai ailleurs*
- « la dernière fois » : *C'est la dernière fois que je commande sur votre site*
- « ne plus acheter » : *Je n'achèterai plus jamais rien chez vous !*

Ces expressions ont de commun qu'elles visent rarement un sujet directement, au contraire des opinions que nous avons définies, c'est pourquoi la classification de phrases semble une meilleure option. Par ailleurs, la seule présence d'indices lexicaux tels que les exemples présentés ne suffit pas à déterminer un risque d'attrition (à titre d'exemple, les phrases « *La dernière fois que je suis venu, les vendeurs ont été aimables* », « *Je n'irai jamais acheter un appareil ailleurs* » ne seraient pas pertinentes), mais nécessite une prise en compte d'un contexte plus large, ce qui est possible si l'analyse est réalisée à l'échelle de la phrase.

Évaluation du personnel

Enfin, notons également la classification à l'échelle de la phrase comme un moyen d'évaluer les échanges entre les salariés d'une entreprise et les clients. L'application industrielle ici concerne l'animation du personnel qui est en contact direct avec l'utilisateur (vendeur, conseiller, téléconseiller...) et constitue le sujet des avis concernés. Il est donc intéressant pour l'ensemble de l'équipe du personnel de découvrir les avis de ce type, indépendamment du fait que ceux-ci leur soit favorables.

- *Le conseiller a parfaitement su me répondre*
- *Elle a été très sympa*
- *On m'a bien conseillé*
- *On m'a raccroché au nez*
- *Je ne veux plus traiter avec ce conseiller*

Le jugement au sein de ces avis est par définition adressé à un sujet, cependant celui-ci est régulièrement présent sous une forme implicite (par l'emploi d'une anaphore ou par métonymie notamment), c'est pourquoi bon nombre de ces expressions ne peuvent être détectées de la même façon que les opinions formées d'un terme et d'un subjectivème explicites.

Parcours

La dernière application de l'analyse à forte granularité que nous identifions en fouille d'opinion pour la relation client est l'identification de récit de « parcours ». Il peut être difficile pour une entreprise de se rendre compte précisément des étapes par lesquelles passe un client dans le cas d'un processus relativement long comme un retour de produit, une contestation ou simplement une demande d'information. Afin d'améliorer ce type de processus, une analyse à l'échelle du document peut mettre en avant les utilisateurs fournissant un récit détaillé de leur parcours.

> *Après accord verbal de la conseillère [...] j'ai fini par recevoir une lettre de refus du service financier. Sauf, qu'entre temps, j'avais signé une demande d'ouverture de compte courant à la banque en me promettant la gratuité de la gestion de compte pendant un an. Une fois le compte ouvert, on m'a débité des frais bien que pas utilisé. Ne souhaitant évidemment pas payer ces frais, j'ai reçu un courrier d'huissier. Que d'incompétence !*

Suite à l'identification de ces avis, une analyse supplémentaire de séparation des différentes étapes, à l'échelle de la phrase, permet d'identifier les points les plus bloquants du processus. À l'image des émotions illustrées à l'aide de la roue de Plutchik (figure 2.2), ces types d'expressions ne sont pas exclusifs et peuvent être combinés (par exemple, une expression relative à l'attrition peut suivre un récit d'un parcours difficile...). L'analyse croisée de ces expressions offre dès lors des pistes d'explication riches en enseignements vis-à-vis des comportements d'utilisateurs.

3.2 Fouille d'opinion à granularité fine

Dans le cadre de ce travail, nous nous intéressons davantage à la fouille d'opinion à granularité fine. Il s'agit d'un cadre d'application d'extraction des opinions ciblées sur des sujets, telles que nous les avons décrites au chapitre 2, et par conséquent plus adapté à l'analyse des avis clients. Nous détaillons la définition de ce problème d'extraction à travers une modélisation existante, et nous en présentons l'utilisation pour la production d'un résumé d'opinion à partir d'avis clients. Dans une dernière partie, nous questionnons la pertinence de cette modélisation et nous en proposons une nouvelle, intervenant à une granularité intermédiaire permettant d'approcher le problème de la dépendance au domaine de manière plus efficace.

3.2.1 Définition du problème

Travaux existants

Les travaux que nous considérons comme précurseurs pour ce type d'analyse sont ceux de [Hu and Liu, 2004] et de [Wilson et al., 2005]. Dans le premier, les auteurs recherchent les adjectifs cooccurrents des sujets abordés parmi des avis clients. La polarité des adjectifs inconnus est inférée par synonymie à la manière de [Kim and Hovy, 2004], ce qui peut entraîner une dérive sémantique comme nous l'avons vu en section 3.1.1. Le second travail présente une méthode de classification en polarité de segments de phrases à partir de traits syntaxiques et d'un lexique affectif. Les auteurs soulignent en particulier les difficultés de cette tâche dues aux ambiguïtés syntaxiques et sémantiques comme la détection de la négation ou l'instabilité de la polarité de certains marqueurs d'opinion. [Wiebe and Mihalcea, 2006] explorent davantage la question de ces ambiguïtés et proposent d'adapter des méthodes distributionnelles provenant du domaine de la désambiguïsation lexicale afin de qualifier avec une plus grande précision la subjectivité de segments de phrases. [Ding et al., 2008] définissent des patrons syntaxiques permettant de désambiguïser la polarité d'opinions afin de retrouver *in fine* la polarité globale associée à un sujet dans un corpus. [Lavalley et al., 2010] expérimentent un modèle SVM (*Support Vector Machine*) pour l'extraction de chaînes de mots relatives à des opinions parmi des avis clients en français et suggèrent un croisement entre opinions et thématiques tel que nous le présentons dans la suite de ce chapitre.

Certains travaux en fouille d'opinion à granularité fine font usage de l'analyse en dépendances syntaxiques dans le but d'identifier les relations entre un sujet et un marqueur d'opinion. Dans cette voie, [Qiu et al., 2009] montrent qu'une sélection manuelle de règles d'extraction reposant sur l'analyse en dépendances permet d'atteindre des performances intéressantes à la fois en extraction d'opinion et en inférence de polarité. [Jakob and Gurevych, 2010] exploitent les dépendances syntaxiques comme traits d'un modèle markovien CRF (*Conditional Random Field*) dans le même but. Enfin, [Liu et al., 2015] proposent une méthode syntaxique permettant de s'affranchir de la sélection manuelle de règles et conservant des résultats intéressants.

Retrouver les couples terme – subjectivème

Formellement, l'objectif de la fouille d'opinion à granularité fine est de retrouver dans chaque document les occurrences des sujets abordés sur lesquels un jugement est porté et l'orientation sémantique (que nous représentons ici par une valeur de polarité) de chacun de ces sujets. En utilisant les définitions exposées précédemment, ce problème revient à retrouver les couples terme–subjectivème, puisque ce sont les subjectivèmes qui indiquent le jugement porté sur le terme.

Nous notons principalement deux paradigmes envisageables pour approcher ce problème. Soit l'expression d'opinion est détectée comme un tout portant une unique information de la forme d'un couple {*sujet*, *polarité*}, soit les éléments contribuant à retrouver cette information sont détectés distinctement, comme illustré sur la figure 3.1. Dans ce travail, nous adhérons au second paradigme en suivant une logique aristotélicienne selon laquelle « qui peut le plus peut le moins ». Considérer chacun de ces éléments permet effectivement, entre autres, de retrouver les expressions d'opinion recherchées selon le premier paradigme.

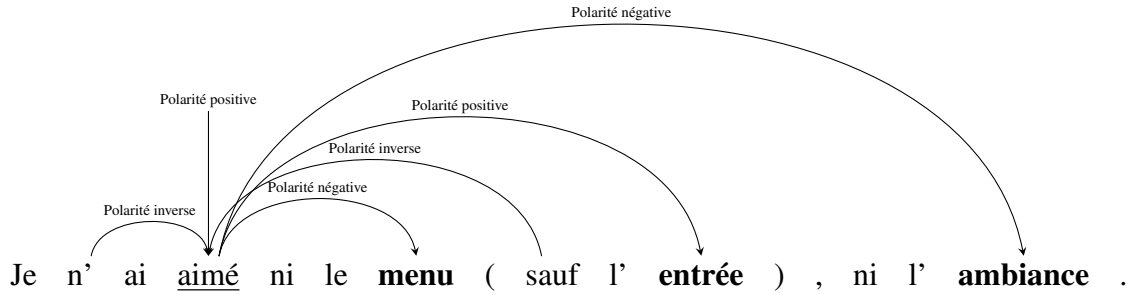


FIGURE 3.1 – Exemple d'une phrase analysée en opinion à granularité fine. Les sujets sont « menu », « entrée » et « ambiance », sur lesquels est projetée la polarité du marqueur d'opinion « aimer ». La particule de négation « ne » et la préposition « sauf » influent sur cette polarité en l'inversant.

En résumé, si à la lecture d'un avis un sujet est explicitement évoqué dans un jugement positif ou négatif, alors nous considérons que des indices dans le texte de ce jugement existent. Retrouver automatiquement ces indices constitue tout le défi de la fouille d'opinion à granularité fine.

3.2.2 Une approche adaptée au résumé d'opinion

L'application finale de la fouille d'opinion à granularité fine que nous définissons ici est la production d'un résumé d'opinion. Ce résumé, pouvant prendre diverses formes et notamment celle d'un rapport statique ou d'un tableau de bord (ou *dashboard*) dynamique, est avant tout le résultat d'une analyse qui se veut à la fois le plus complet et le plus lisible possible. En suivant ces objectifs, le travail de [Liu et al., 2005] propose un résumé visuel, où sont évalués les différents aspects d'un produit sur un axe *positif – négatif*.

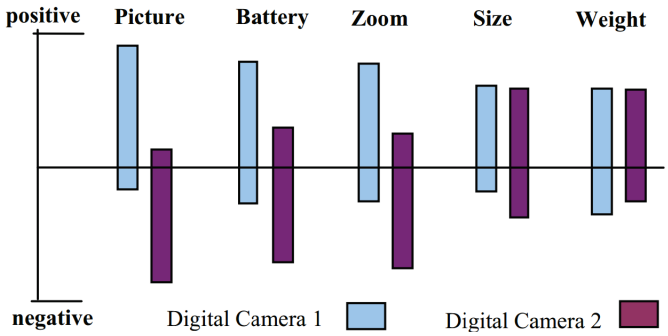


FIGURE 3.2 – Représentation visuelle du résumé d'opinion tel que présenté dans [Liu et al., 2005]

Sur ce résumé visuel, présenté en figure 3.2, l'accent est mis sur la comparaison entre deux produits similaires. D'autres types de résumés, tels que des résumés sous forme textuelle [Ku et al., 2005, Blair-Goldensohn and Hannan, 2008, Meng et al., 2012], permettent davantage de rendre compte des besoins et inquiétudes des utilisateurs en s'appuyant sur les propos émis et non sur des aspects attendus.

Toutefois, le but de ces productions reste avant tout de *constater*, or il est attendu des résumés d'opinion qu'ils soient en mesure d'*expliquer* des tendances ou même des phénomènes ponctuels. En utilisant une représentation telle que celle proposée par [Llosa, 1996] (représentée sur la figure 3.3), il est par exemple possible de présenter les sujets abordés par les utilisateurs selon leur potentiel impact sur leur satisfaction.

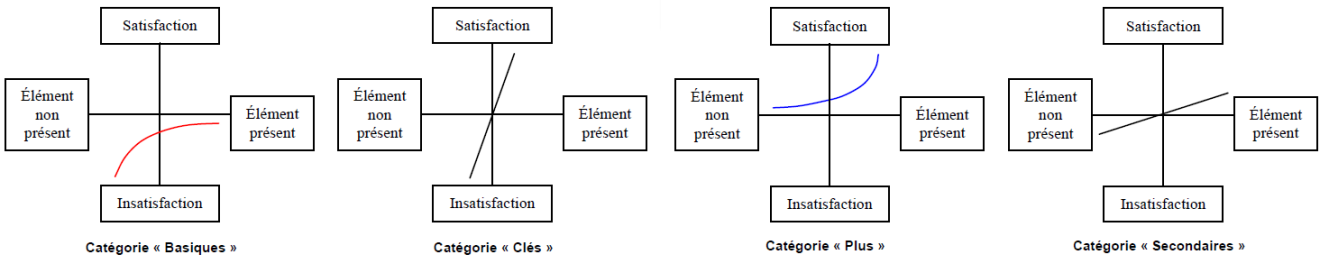


FIGURE 3.3 – Matrices de Llosa. Les éléments « basiques » sont les services que l'utilisateur s'attend à avoir et qui contribuent donc uniquement à une insatisfaction en cas d'absence ; les éléments « plus » sont les services auquel l'utilisateur ne s'attend pas et qui contribuent à une hausse de satisfaction seulement ; les éléments « clés » sont fortement associés à la satisfaction comme à l'insatisfaction des utilisateurs, et les éléments « secondaires » sont les services qui y contribuent peu.

3.2.3 La modélisation ABSA

Pour répondre à la fois au problème de l'identification des éléments constituant une opinion et à l'objectif de production d'un résumé clair, [Liu, 2012] propose une modélisation de l'opinion comme un tuple (e, a, s, h, t) associant :

- l'entité représentée par la cible du jugement (e pour *entity*) ;
- l'aspect sur lequel cette entité est jugée (a pour *aspect*) ;
- l'orientation sémantique de ce jugement (s pour *sentiment*) ;
- l'auteur de l'opinion (h pour *holder*) ;
- le moment où cette opinion est émise (t pour *time*).

Nous pouvons noter que ce tuple ne contient ni l'occurrence du sujet abordé, mais seulement l'entité représentée par ce terme à un niveau d'abstraction supérieur, ni le marqueur d'opinion employé mais seulement la polarité finale inférée sur le sujet. La notion centrale d'*aspect* (d'où la dénomination *aspect-based sentiment analysis* ou ABSA) correspond ici à un niveau de description du sujet visé plus précis que l'entité. Lorsqu'il n'y a pas lieu de décrire cet aspect, c'est-à-dire que l'entité est jugée dans son entièreté, la valeur de a n'est pas attribuée. Il s'agit d'une valeur qui, à l'instar de l'entité, ne représente pas le texte de l'occurrence du terme mais un concept pré-établi, auquel peuvent se rapporter plusieurs termes. La représentation abstraite de l'occurrence d'opinion ainsi formée permet une réalisation d'un résumé d'opinion instantanée. D'un point de vue théorique ce résumé est un arbre, tel qu'illustré en figure 3.4, décrivant les entités, aspects et polarités sur ces aspects.

Enfin, les notions d'auteur et de moment d'émission de l'avis, certes essentielles au résumé d'opinion (il est primordial de dissocier les clients ayant donné leur avis, ainsi que des avis éloignés dans le temps), ne font pas partie de notre périmètre d'étude. Dans le cas des corpus d'avis clients, ces informations sont en réalité données de façon structurée, autrement dit en tant que métadonnées associées au texte, il n'y a donc pas lieu de les extraire au moyen d'une méthode du traitement automatique du langage.

3.2.4 Limites de l'approche

Adaptabilité de la modélisation

Bien qu'intuitifs, des résumés tels que présentés sur les figures 3.4 et 3.2 présentent quelques limites d'utilisation. Ce type de résumé est particulièrement adapté aux avis concernant des produits de consommation ou services dont les sous-parties sont aisément identifiables – et parfois même listées sur le site marchand

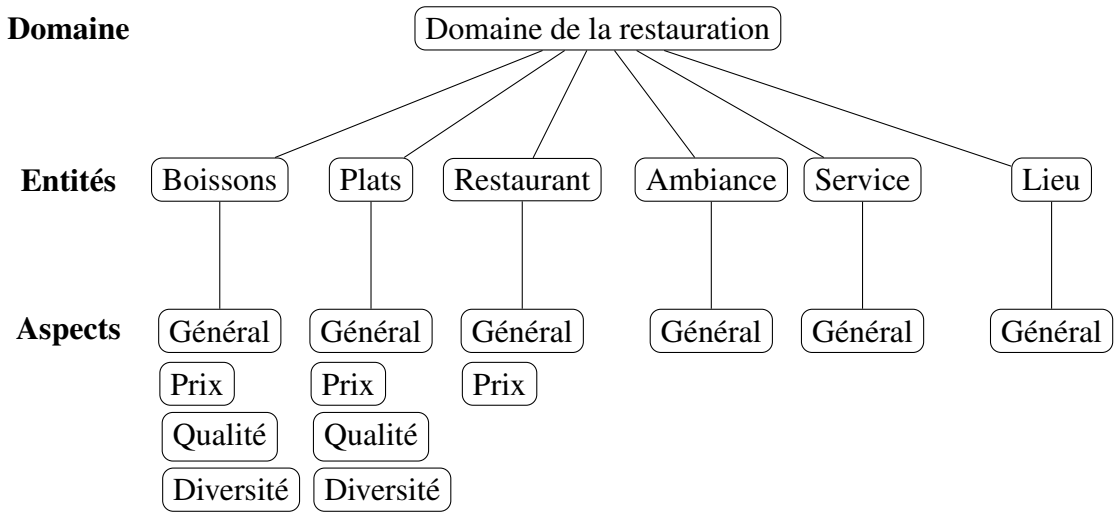


FIGURE 3.4 – Hiérarchie d'entités et d'aspects dans les corpus d'avis dans le domaine de la restauration, annotés pour l'atelier SEMEVAL ABSA 2016.

– et font intuitivement l'objet de critiques. Lorsque les avis recueillis décrivent une expérience vécue et consistent davantage en un récit émotionnel qu'en une énumération de points forts ou faibles, la structure du résumé d'opinion est bien moins prévisible. Or cela est, à notre connaissance, le cas de la majorité des avis car les utilisateurs se concentrent bien souvent sur cette expérience, qui peut différer d'une enseigne à l'autre, et non sur les caractéristiques d'un produit qui est identique quelle que soit l'enseigne et dont les utilisateurs sont parfois même tout à fait conscients *avant* l'achat.

Ceci nous renvoie à l'argument que nous évoquions au chapitre 2, soit « écouter plutôt que questionner ». Nous voyons ici qu'il est nécessaire de partir des expressions concrètes employées par les utilisateurs afin de bâtir des entités leur correspondant.

Attribution des entités

Afin de respecter une structure arborescente du résumé d'opinion tel que présentée sur la figure 3.4, les intitulés des entités et aspects doivent être établis préalablement à l'analyse. De cette façon, la modélisation ABSA propose une façon optimale de générer un résumé d'opinion sur l'ensemble du corpus à partir des résumés de chaque document. Cependant ce fonctionnement suppose deux étapes supplémentaires : la définition des entités et l'attribution des termes à une entité définie.

La définition des entités et des aspects du domaine du corpus est une étape peu détaillée dans la littérature. Dans les faits, les travaux sur ce sujet s'appuient sur des corpus annotés (ressources des ateliers SEMEVAL ABSA, corpus d'avis clients présenté dans [Hu and Liu, 2004]) et tiennent donc pour référence les ressources des travaux antérieurs. Ces corpus ne traitant que d'un certain nombre de domaines (appareils électroniques et restaurants principalement), la question de la définition d'entités et d'aspects pour de nouveaux domaines reste en suspens. Du point de vue de l'application de fouille d'opinion à laquelle nous contribuons, l'identification des entités suit la même logique que dans le cas des termes, à savoir que les entités émergent du corpus, puisque celles-ci sont fondées sur les termes. Plus précisément, la définition des entités résulte d'une segmentation manuelle des termes fréquents représentant un *concept commun*, ou des termes présentant un fort *impact* s'ils ne sont pas fréquents. La nature subjective de ces notions de concept commun et d'impact explique que cette étape soit réalisée manuellement.

La question de la classification des termes cibles d'opinion en entités est à notre sens une piste de recherche plus prometteuse car moins biaisée par l'aspect subjectif de l'analyse.

3.3 Fouille d'opinion à granularité intermédiaire

Afin de réduire le travail d'annotation nécessaire à l'analyse de nouveaux domaines, et pour rendre plus cohérente la tâche d'extraction d'opinion à granularité fine, nous proposons une modification de l'annotation des cibles d'opinion. Nous nous inspirons pour cela d'une autre tâche de recherche d'information, l'extraction d'entités nommées.

Les deux tâches paraissent en effet similaires aussi bien dans leur objectif que dans les méthodes employées. La reconnaissance d'entités nommées est une extraction de segments de phrases correspondant à des unités sémantiques indivisibles au moyen d'un étiquetage de séquence. L'historique des travaux de ce domaine est notablement plus substantiel que celui de l'extraction de termes cibles d'opinion, et nous supposons à ce titre que la modélisation des entités nommées, communément acceptée et répétitivement validée par de multiples campagnes d'évaluation, comme TREC[2] (*Text REtrieval Conference*) ou SEMEVAL, peut être source d'inspiration.

Le	**rayon DVD**	est	fourni	,	et	la	**vendeuse**	a	su	nous	aider	.
	LIEU						PERSONNE					

FIGURE 3.5 – Exemple d'annotation d'une phrase en entités nommées

Nous constatons que, malgré leur similitudes, les deux tâches ne partagent pas la même modélisation. Il n'est pas question en reconnaissance d'entités nommées d'identifier des expressions correspondant à une entité générique, puis de retrouver sa catégorie précise dans un deuxième temps, comme dans le cas de l'extraction de cibles d'opinion. Il est attendu que les catégories d'entités nommées soient directement extraites ; celles-ci sont à cet effet indiquées comme étiquettes d'apprentissage à l'image de la figure 3.5 dans les corpus annotés pour cette tâche pour des systèmes statistiques supervisés.

3.3.1 Fouille d'opinion au niveau entité

Principe

Nous suggérons une modélisation de la reconnaissance de cibles d'opinion utilisant les entités d'opinion [Lark et al., 2017] de façon analogue aux classes d'entités nommées, comme illustré en figure 3.6.

	Le	**rayon DVD**	est	fourni	,	et	la	**vendeuse**	a	su	nous	aider.
Annotation binaire		CIBLE						CIBLE				
Annotation par entité		PRODUITS						SERVICE				

FIGURE 3.6 – Exemple d'annotation binaire et par entité d'une phrase en cibles d'opinion

Cette approche est à notre sens plus adaptée à la production d'un résumé d'opinion puisque la notion d'entité est ici centrale à la fois dans la modélisation et dans la méthode d'extraction, à la différence de l'annotation à l'échelle du terme, où une étape supplémentaire est nécessaire.

Il parait plus cohérent d'annoter ainsi directement les éléments que nous recherchons dans le texte, à la manière de la reconnaissance d'entités nommées. Cette cohérence n'est pas uniquement théorique, mais rend compte du fait que les entités regroupent non seulement des termes mais également des marqueurs d'opinion, et d'une façon plus générale des contextes lexicaux similaires.

Du point de vue de la méthode utilisée pour l'extraction d'opinion, cette cohérence signifie une information plus riche et donc la possibilité d'exploiter plus efficacement les données annotées. Dans le contexte d'urgence d'acquisition de ressources pour la fouille d'opinion dans lequel nous nous plaçons cette approche est, de fait, une promesse fort intéressante.

[2] http://trec.nist.gov/

Observations

Nous proposons de mesurer cette cohérence sur les corpus annotés mis à disposition pour l'atelier SEMEVAL ABSA 2016 dans la mesure où ceux-ci font figure de référence pour la tâche d'extraction de cibles d'opinion en plusieurs langues. Afin de consolider nos observations, nous effectuons nos mesures sur un domaine d'avis clients disponible en deux langues, à savoir le domaine de la restauration en français et en anglais.

Pour chacune des entités identifiées pour ce domaine, nous mesurons la disparité des termes par entité en comptant le nombre de termes exclusifs à cette entité, et la disparité des marqueurs d'opinion par entité en comptant le nombre de marqueurs exclusifs cooccurrents d'un terme représentant cette entité.

Entité	#Termes cibles	#Cibles uniques	#Cibles excusives	#Marqueurs d'opinion	#Marqueurs uniques	#Marqueurs exclusifs
AMBIENCE	287	115 (40,07%)	99 (86,09%)	165	85 (51,52%)	33 (38,82%)
DRINKS	133	59 (44,36%)	58 (98,31%)	48	23 (47,92%)	4 (17,39%)
FOOD	1,311	541 (41,27%)	538 (99,45%)	776	193 (24,87%)	105 (54,40%)
LOCATION	32	16 (50,00%)	10 (62,50%)	18	15 (83,33%)	3 (20,00%)
RESTAURANT	343	119 (34,69%)	99 (83,19%)	214	90 (42,06%)	32 (35,56%)
SERVICE	432	62 (14,35%)	56 (90,32%)	250	122 (48,80%)	59 (48,36%)

TABLEAU 3.1 – Termes cibles et marqueurs d'opinion par entité dans le corpus SEMEVAL ABSA 2016 en anglais

Entité	#Termes cibles	#Cibles uniques	#Cibles excusives	#Marqueurs d'opinion	#Marqueurs uniques	#Marqueurs exclusifs
AMBIENCE	253	42 (16,60%)	32 (76,19%)	124	69 (55,65%)	29 (42,03%)
DRINKS	123	47 (38,21%)	46 (97,87%)	43	33 (76,74%)	4 (12,12%)
FOOD	1,248	400 (32,05%)	392 (98,00%)	540	231 (42,78%)	136 (58,87%)
LOCATION	56	28 (50,00%)	20 (71,43%)	25	19 (76,00%)	6 (31,58%)
RESTAURANT	247	42 (17,00%)	25 (59,52%)	127	75 (59,06%)	30 (40,00%)
SERVICE	536	49 (9,14%)	46 (93,88%)	321	150 (46,73%)	78 (52,00%)

TABLEAU 3.2 – Termes cibles et marqueurs d'opinion par entité dans le corpus SEMEVAL ABSA 2016 en français

Nous observons que ces mesures sont très proches pour les deux langues, au sens où l'ordre relatif des termes et des marqueurs d'opinion selon leur unicité et leur exclusivité est globalement conservé en dépit du fait que les deux corpus ont été annotés par des experts distincts. Il n'est pas évident de dégager d'autres conclusions de ces tableaux, étant donné que les mesures d'unicité et d'exclusivité ne semblent pas systématiquement liées. Afin d'apprendre plus de ces mesures, nous les croisons avec des courbes d'apprentissage par entité. Nous considérons qu'une entité est d'autant plus *cohérente* que son apprentissage est rapide et efficace, ce qui se traduit par une courbe d'apprentissage progressant rapidement vers une valeur de F-mesure haute.

Les courbes présentées en figure 3.7 ont été obtenues en réalisant un apprentissage pour chaque entité. Un modèle CRF (*Conditionnal Random Fields* [Lafferty et al., 2001]) est itérativement entraîné par ajouts successifs de sous-ensembles de 50 documents parmi l'ensemble des avis du domaine de la restauration. La sélection des documents est déterminée par une méthode d'apprentissage actif appelée échantillonnage de l'incertain (ou *uncertainty sampling [Lewis and Gale, 1994]*) : les 50 documents fournis au modèle à chaque nouvelle itération sont considérés les plus utiles pour l'apprentissage car ce sont ceux sur lesquels le modèle présente la plus faible probabilité de classification.

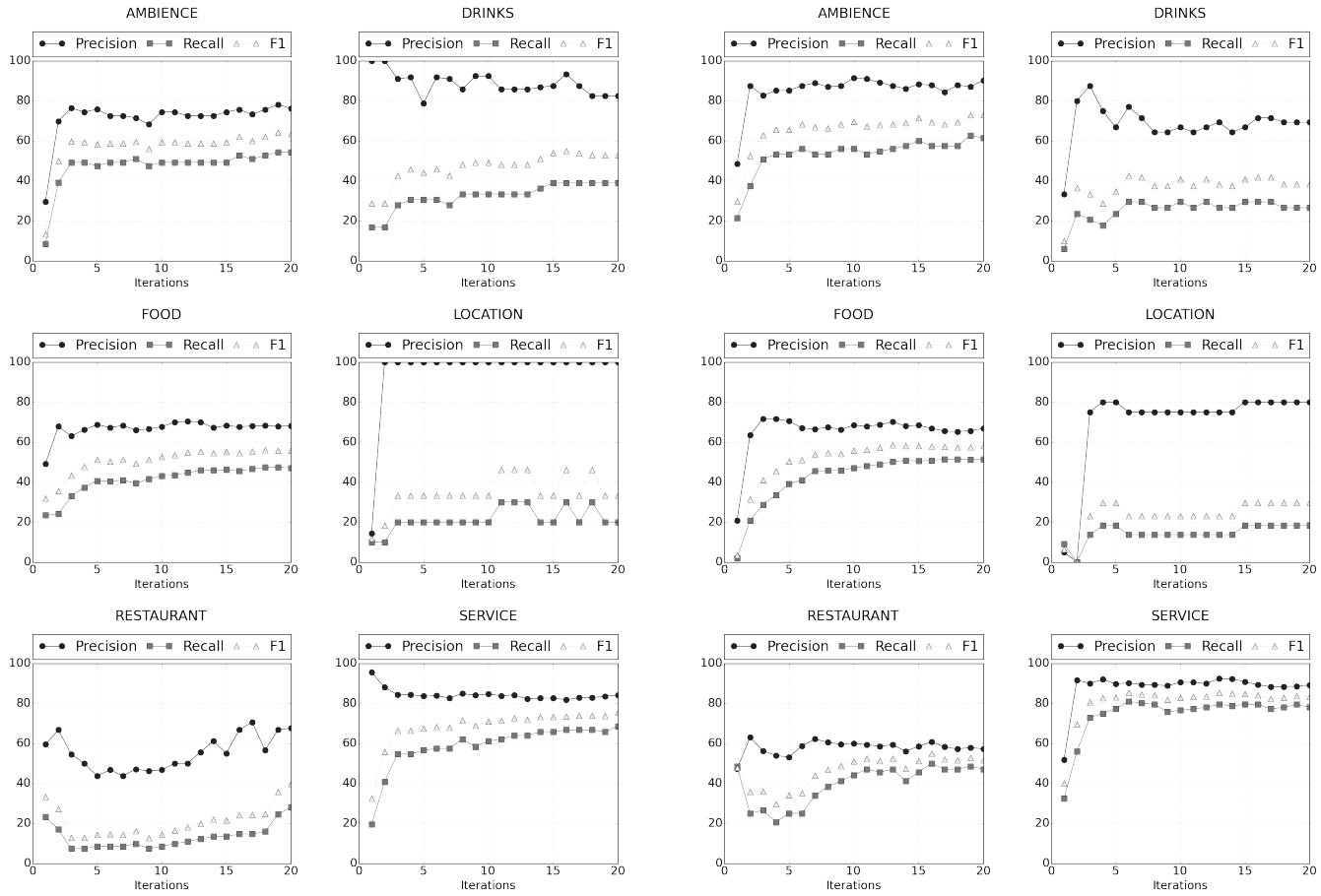


FIGURE 3.7 – Courbes d'apprentissage pour l'extraction d'entités dans le corpus en anglais (gauche) et dans le corpus en français (droite)

Nous remarquons en croisant les statistiques sur les éléments lexicaux montrés dans les tableaux 3.1 et 3.2 et les courbes d'apprentissage présentées en figure 3.7 que ce sont les marqueurs d'opinion qui affectent de façon plus importante l'apprentissage. En effet le facteur commun reliant les trois entités pour lesquelles le modèle arrive rapidement à une F-mesure constante, AMBIENCE (ambiance), FOOD (plats) et SERVICE est un taux élevé de marqueurs d'opinion exclusifs. Par ailleurs, le taux de termes exclusifs n'apparaît pas aussi crucial que ce à quoi nous nous attendions. À titre d'exemple le faible taux de termes exclusifs de l'entité AMBIENCE, partageant notamment le terme fréquent « restaurant » avec les entités RESTAURANT et LOCATION (lieu) n'a visiblement pas d'impact négatif critique sur le bon déroulé de son apprentissage.

3.3.2 Une approche inter-domaines

En addition à une cohérence accrue de la modélisation du discours, la fouille d'opinion au niveau entité offre également un nouveau point de vue sur la question de l'adaptation au domaine.

Entités et domaines

Nous l'avons vu, la notion d'opinion est fortement connexe à la notion de domaine. En ce qui concerne la fouille d'opinion à granularité fine, cette dépendance est matérialisée par une segmentation par domaine des corpus annotés. Il est considéré que ces ressources sont incompatibles, au sens où un modèle statistique entraîné sur un corpus d'un domaine montre des performances de classification insatisfaisantes sur des domaines différents. Cette incompatibilité est d'autant plus frappante que les cibles d'opinion sont annotées en utilisant le même libellé. Il est donc accepté que la notion de cible ne soit pas universelle, mais dépendante du périmètre du domaine qui, comme nous l'avons évoqué, est difficile à délimiter. À l'inverse, la notion

d'entité propose un périmètre restreint par un ensemble stable de termes et représente une unité sémantique aisément appréhendable.

Limiter l'effort d'adaptation au domaine

Cette approche offre en outre la possibilité de réutiliser un matériel annoté lors de l'analyse d'un nouveau domaine. Si nous reprenons l'exemple des corpus annotés pour les ateliers ABSA, les trois domaines (restaurants, hôtels et musées) partagent plusieurs entités, comme il est illustré en figure 3.8. En rapportant la notion de cible d'opinion à l'échelle de l'entité, nous pouvons donc construire de façon modulaire les corpus d'entraînement pour chaque domaine. Les exemples restant à annoter sont ceux qui ne correspondent à aucune entité déjà rencontrée.

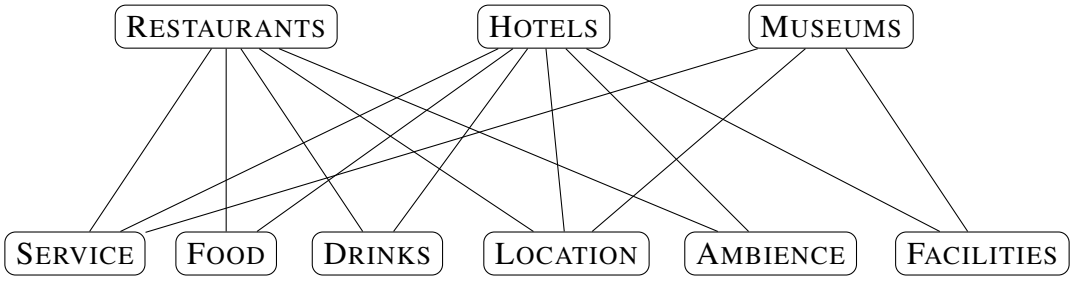


FIGURE 3.8 – Entités (cadres du bas) partagées entre domaines (cadres du haut)

Alors que l'approche par domaine impose un travail d'annotation constant, cette modélisation à l'échelle de l'entité représente non seulement un gain de temps par domaine, mais de surcroît ce gain de temps est croissant, la probabilité de rencontrer une entité inconnue allant en diminuant. Dans le chapitre 5, concernant les méthodes probabilistes d'extraction d'opinion, nous montrons que cette modélisation permet effectivement de réutiliser les données annotées à travers différents domaines.

3.4 Synthèse

Dans ce chapitre, nous avons présenté les différentes échelles auxquelles s'appliquent la fouille d'opinion parmi les avis clients :

- l'échelle du **document** ou de la **phrase**, permettant de dissocier des profils clients apparaissant dans le corpus et éventuellement d'alerter une entreprise de clients exprimant un besoin ou une inquiétude particulière ;
- l'échelle du **terme**, permettant de faire émerger les sujets abordés par les utilisateurs et d'en produire un résumé d'opinion représentatif.

Nous nous sommes ensuite interrogés sur la pertinence du modèle ABSA (*aspect-based sentiment analysis*) et en particulier sur l'annotation binaire de cette modélisation pour étiqueter les termes cibles d'opinion dans le texte. Nous suggérons en lieu et place de cette annotation une annotation multi-classe, centrée sur la notion d'entité. Cette modélisation permet notamment :

- de définir de façon plus **cohérente** la notion de cible d'opinion, en la définissant de façon universelle et non au cas par cas selon le domaine du corpus ;
- d'approcher l'épineux problème de l'**adaptation au domaine** d'une manière nouvelle, à l'aide de ressources modulaires. L'annotation des cibles ainsi définies au niveau entité est réutilisable pour différents domaines, ce qui réduit le travail manuel nécessaire à l'analyse de chaque nouveau domaine.

II

Méthodes d'extraction d'opinion

— Il y a des chasseurs, sur cette planète-là ?
— Non.
— Ça, c'est intéressant ! Et des poules ?
— Non.
— Rien n'est parfait, soupira le renard.

Antoine de Saint-Exupéry (*Le Petit Prince*)

4

Approche symbolique de la fouille d'opinion

Dans la partie précédente, nous avons étudié les différents éléments qui constituent la notion d'opinion, ainsi que leur interdépendance. Nous avons établi que produire un résumé d'opinion nécessite de détecter distinctement les sujets abordés par les utilisateurs et les marqueurs d'opinion employés pour qualifier ces sujets, puis d'inférer l'orientation sémantique de l'opinion que ces éléments forment. Le fait de discerner ces éléments pour notre application finale est avant tout le moyen de spécifier les types de ressources qu'il est préférable d'extraire pour appuyer cette application. Or le choix d'un type de ressource est intrinsèquement lié aux types de méthodes d'extraction que nous utilisons.

La présente partie décrit les paradigmes d'extraction de l'opinion parmi les avis clients employant chacun un type de ressource. Nous approchons tout d'abord ces méthodes par l'angle du paradigme symbolique, c'est-à-dire utilisant une description telle quelle des éléments linguistiques attendus dans les expressions à extraire.

4.1 Patrons de détection

La description des expressions d'opinion utilisant le paradigme symbolique repose sur le principe de *familles* d'expressions partageant un même canevas, ou *patron* de détection. Nous considérons selon ce paradigme qu'une occurrence d'opinion exprimée par un utilisateur respecte nécessairement un tel patron. Le corollaire de cette hypothèse est que si une expression d'opinion ne correspond à aucun patron répertorié, il est nécessaire d'en définir un nouveau. Dans cette section nous détaillons la définition de ces patrons, ainsi que la représentation employée permettant leur utilisation pour la fouille d'opinion.

4.1.1 Travaux existants

Les patrons de détection constituent un moyen intuitif de représenter des expressions particulières du langage, ce qui peut expliquer leur emploi parmi les premiers travaux en extraction d'opinion.

[Riloff et al., 2003] présente une méthode d'extraction de patrons représentant des expressions subjectives, reposant sur un algorithme précédemment défini pour la recherche d'information [Thelen and Riloff, 2002]. Les patrons, composés d'un emplacement pour le sujet de la phrase et d'un contexte syntaxique, sont associés à un score de pertinence correspondant à leur capacité à extraire des expressions subjectives, pondéré en fonction de leur apparition au sein de phrases subjectives. Cette méthode a été source d'inspiration pour notre travail, et nous la présentons plus en détail au chapitre 8, dans la mesure où celle-ci concerne davantage l'extraction des patrons que la fouille d'opinion à proprement parler.

[Yi et al., 2003] effectue une extraction d'opinion à l'aide d'un lexique affectif et d'un ensemble de patrons syntaxiques, dans lesquels l'emplacement d'un marqueur d'opinion est spécifié et projette son orientation sémantique sur le sujet de l'opinion, de manière très similaire à la méthode que nous présentons dans ce chapitre. En suivant les mêmes étapes fondamentales (identifier les sujets de l'opinion, puis leurs occurrences au sein d'expressions subjectives et enfin leur orientation sémantique) [Popescu and Etzioni, 2005] présente une application de fouille d'opinion, OPINE, dont l'objectif est de produire un résumé d'opinion. L'idée mise en avant dans ce travail est d'utiliser les dépendances syntaxiques d'une phrase afin d'identifier avec précision le lien entre un sujet abordé par un utilisateur et un marqueur d'opinion, puis d'appliquer une règle logique d'inférence sur les éléments identifiés afin de retrouver l'orientation sémantique de l'opinion. [Qiu et al., 2011] exploite également les dépendances syntaxiques jointes à des règles d'inférence dans le but non seulement d'extraire les sujets d'opinion mais également d'enrichir le lexique affectif, en projetant les structures de dépendances sur des segments de phrases où un marqueur d'opinion n'a pas encore été détecté. Le principe d'extraction jointe a également permis d'orienter nos travaux en extraction de patrons, c'est pourquoi nous nous intéresserons plus en détail à ce travail au chapitre 8.

4.1.2 Propriétés des patrons

Définition d'un patron

Un patron lexico-syntaxique est une séquence composée de mots représentés par leur forme lemmatisée ou leur catégorie grammaticale, tel qu'illustré sur la figure 4.1. Afin de limiter les cas d'ambiguïté lors de la définition des patrons, un élément peut être représenté par plusieurs formes. Cela permet entre autres de séparer les usages des homographes ne partageant pas la même catégorie grammaticale.

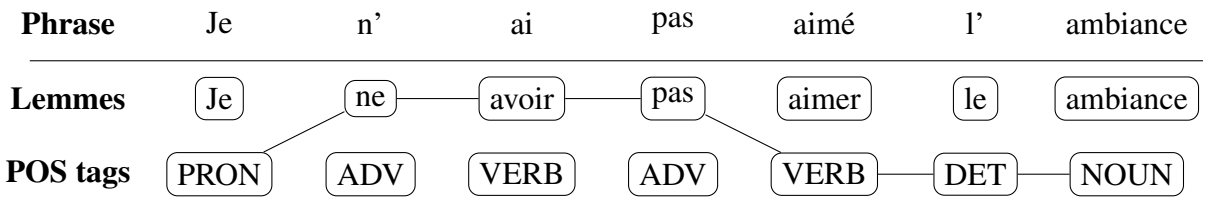


FIGURE 4.1 – Présentation graphique du patron PRON *ne pas avoir* VERB DET NOUN, utilisant le jeu d'étiquettes grammaticales *Universal POS tagset*[1]

L'approche symbolique de la fouille d'opinion à granularité fine requiert par conséquent une ressource listant de tels patrons pouvant couvrir autant d'expressions d'opinion qu'il est possible, mais uniquement ce type d'expression. Les deux difficultés que pose cette acquisition découlent logiquement de cette opposition.

D'une part, la diversité des expressions employées par les utilisateurs, dont nous avons donné quelques exemples aux chapitres précédents, fait de cette recherche une quête sans fin. En effet, il existe une multitude de ces patrons possibles pour chaque combinaison de terme et de subjectivème. D'autre part, semblablement à d'autres problèmes d'extraction d'information, cette quête de couverture croise une quête de pertinence. Dans notre cas, celle-ci correspond au fait de s'assurer que les patrons définis ne couvrent pas des expressions qui ne sont pas des opinions. À titre d'exemple, si le patron [NOUN] [ADJ] peut effectivement constituer un patron de détection d'opinion dans un grand nombre de cas (« *menu excellent* », « *livraison correcte* »), tous les segments de textes couverts par ce patron ne représentent pas des expressions d'opinion (« *menu traditionnel* », « *vin blanc* »), ni même des syntagmes nominaux valides (« *menu* ***enfant coûteux*** », « ***livraison petit*** *colis* »).

Expressivité des patrons

Une première solution au problème de pertinence des patrons est d'augmenter leur expressivité, afin de spécifier les cas où ceux-ci peuvent déclencher une reconnaissance. Pour cette raison, nous définissons des patrons non pas sur deux plans, lexical et syntaxique, mais sur trois plans. Le troisième plan que nous utilisons décrit le rôle propre à l'opinion de l'élément dans la phrase, et est essentiel à la spécification de l'activation d'un patron. Nous distinguons quatre rôles possibles : chaque élément d'un patron peut être un terme, un subjectivème, une négation ou tout *autre* mot.

Les termes sont reconnus par une première phase de reconnaissance (dans notre cas, au moyen de patrons lexico-syntaxiques), tandis que les subjectivèmes et négations sont étiquetés depuis un lexique. La notion d'*autre* rôle – qui peut aussi se traduire par *aucun* – sert avant tout à décrire l'absence d'un des autres rôles pour un emplacement donné du patron. Dans la mesure où les patrons ne sont plus seulement lexico-syntaxiques par l'ajout de ce plan, nous utilisons par la suite la dénomination « patron de détection ».

Si nous reprenons l'exemple précédent du patron [PRON] [ne] [avoir] [pas] [VERB] [DET] [NOUN], nous nous apercevons que celui-ci est ambigu, car ce patron peut tout aussi bien correspondre à une proposition dénuée d'opinion, telle que « *Je n'ai pas visité la cathédrale* ». La spécification de ce patron en utilisant un patron à trois niveaux, comme illustré sur la figure 4.2, laisse en revanche peu d'erreur possible.

[1] http://universaldependencies.org

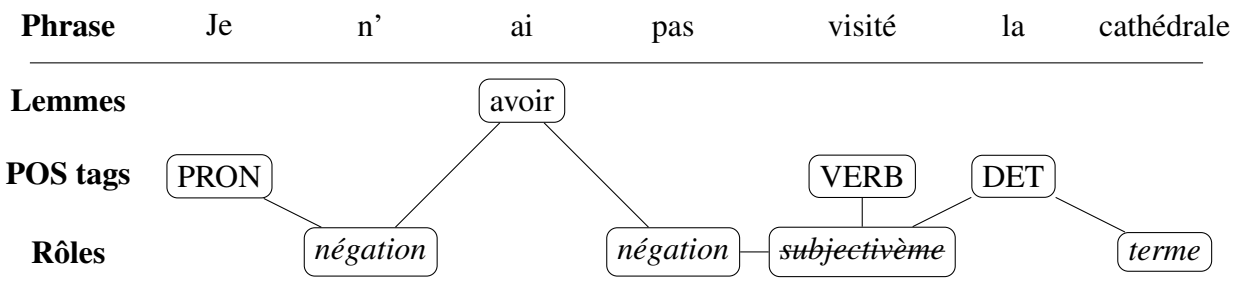


FIGURE 4.2 – Présentation graphique du patron PRON *ne pas avoir* VERB–*subjectivème* DET *terme* et de son alignement avec la phrase « *Je n'ai pas visité la cathédrale* »

Dans le cas de ce patron, la spécification de l'élément correspondant à un verbe en ajoutant la condition que ce verbe soit reconnu comme subjectivème permet d'invalider le segment de phrase analysé en tant qu'expression d'opinion car le verbe « *visiter* » n'est pas considéré comme subjectivème. Par ailleurs, cette représentation permet une plus large couverture puisque l'occurrence du sujet respectant l'emplacement du terme n'est pas limité à la seule étiquette grammaticale NOM mais couvre l'ensemble des mots et syntagmes correspondant à l'énonciation d'un sujet que nous avons rencontré en chapitre 2.

L'importance du prétraitement

Chacune des étapes d'un système d'extraction d'opinion employant l'approche symbolique est extrêmement dépendante de la qualité du prétraitement, qui est un sujet majeur pour l'analyse de contenu généré par des utilisateurs. La correspondance entre patrons à trois niveaux et les séquences de mots au sein d'une phrase doit être totale, notamment pour éviter les cas d'ambiguïtés. L'ajout, le retrait ou la substitution d'un seul élément peut en effet avoir un impact critique sur la sémantique d'une expression reconnue par un patron de détection. En outre, la représentation du texte sur ces trois niveaux d'abstraction (lemmes, étiquettes grammaticales et rôles) est, comme nous l'avons vu, subséquente à une reconnaissance des termes et de subjectivèmes qui dépend elle-même de la phase de prétraitement. En effet, les termes sont extraits au moyen de patrons syntaxiques et les subjectivèmes sont identifiés à partir de leur forme lemmatisée ainsi que leur étiquette grammaticale, afin de réduire les ambiguïtés lexicales présentées en section 2.4.2.

4.1.3 Représentation et reconnaissance

La définition des patrons et leur robustesse au bruit dans le texte ne sont pas les seuls défis que propose l'approche symbolique. La mise en correspondance des patrons avec le texte suppose en effet une représentation déterministe de ces motifs et un algorithme de séquençage permettant de délimiter les expressions cohérentes.

Pour répondre à ces besoins, nous utilisons une machine à état finis (ou, par la suite, automate) dont les chemins de l'état initial à un état final représentent un patron d'opinion. Une ressource contenant les patrons sous forme de règles respectant un langage formel permet de répertorier les types d'expressions reconnues et d'initialiser la génération de cet automate. Nous illustrons ce fonctionnement par l'exemple simple d'un automate formé de trois règles, décrits sur la figure 4.3 – dans la réalité, une telle ressource contient plusieurs dizaines de règles, ce qui en rend la maintenance ardue.

terme	ADJ, *subjectivème*		(« *livraison rapide* »)
terme	ADV, *autre*	ADJ, *subjectivème*	(« *livraison très rapide* »)
terme	ADV, *négation*	ADJ, *subjectivème*	(« *livraison pas rapide* »)

FIGURE 4.3 – Trois patrons de détection tels que nous les définissons manuellement, associés à des exemples de détection

La génération est réalisée en trois parties. À partir des règles, une version naïve de l'automate, c'est-à-dire sous la forme d'un graphe étoilé, est construite (figure 4.4). Chaque règle forme une branche différente à partir d'un état initial commun, ce qui peut créer de nombreuses redondances. Les transitions sur ces branches correspondent à chacun des éléments du patron de détection.

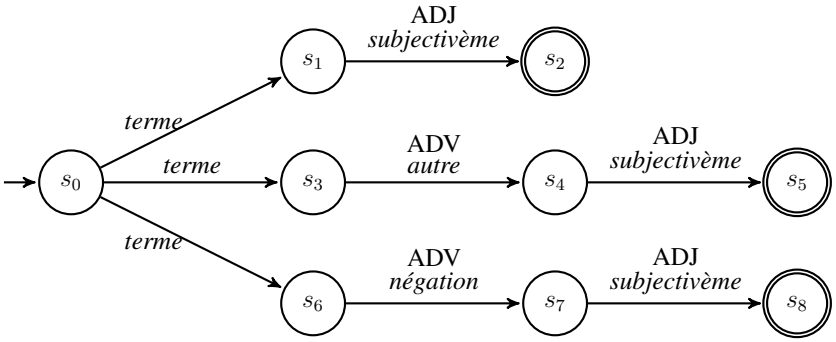


FIGURE 4.4 – Automate non déterministe représentant trois patrons de détection

L'automate est ensuite déterminisé (figure 4.5). Il est essentiel pour cette étape que les règles définies manuellement ne comportent pas d'éléments ambigus car cette déterminisation ne peut résoudre des conflits sémantiques. En revanche, une logique de priorité est appliquée sur les transitions comportant des éléments similaires lors du parcours d'automate. Ainsi, pour éviter un cas de non-déterminisme, une branche plus spécifique sera préférée à une branche plus globale. L'ordre de spécificité correspond au niveau d'abstraction de la transition : une étiquette grammaticale est moins spécifique qu'un rôle, qui est moins spécifique qu'une forme lemmatisée.

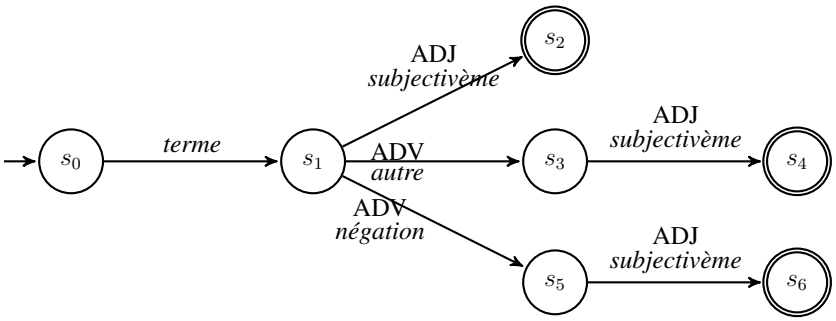


FIGURE 4.5 – Automate déterministe représentant trois patrons de détection

L'automate est finalement minimisé (figure 4.6), afin de réduire le temps de calcul nécessaire à la reconnaissance. Si cette dernière étape n'a pas d'impact sur la capacité du système à extraire les opinions dans le texte, elle est néanmoins essentielle dans le cadre d'une application commerciale de fouille d'opinion, de par le gain de temps de calcul conséquent lors de l'analyse.

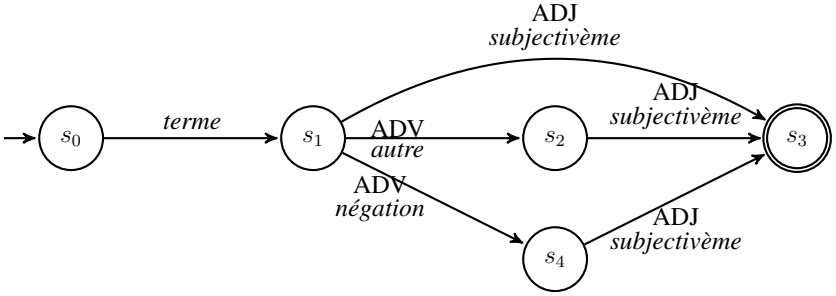


FIGURE 4.6 – Automate minimal représentant trois patrons de détection

4.2 Fouille d'opinion à l'aide de patrons

Les propriétés des patrons de détection maintenant présentées, nous détaillons dans cette section l'utilisation de ces patrons à trois niveaux d'expressivité pour extraire les expressions d'opinion. Nous expliquons premièrement le principe de reconnaissance de ces expressions au moyen d'un automate, puis nous discutons des intérêts et limites de cette approche.

4.2.1 Principe

Comme nous l'avons évoqué, la fouille d'opinion suivant le paradigme symbolique que nous développons ici repose sur une correspondance d'une représentation abstraite d'un type d'expression avec la réalité d'une phrase écrite par un utilisateur. Cette correspondance est réalisée en transformant chaque phrase en une séquence d'éléments selon la même représentation abstraite, de manière à pouvoir la comparer avec chaque patron de détection.

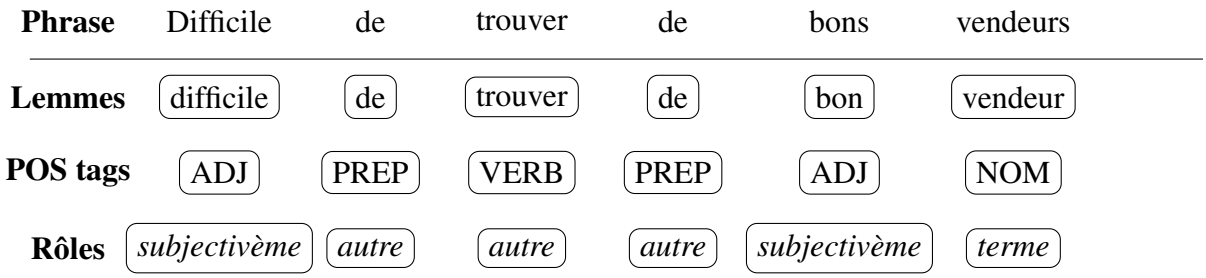


FIGURE 4.7 – Présentation graphique de la phrase « *Difficile de trouver de bons vendeurs* » sur trois niveaux d'abstraction

Parcours d'automate et recouvrement

Le texte des avis, transformé en un flux d'éléments linguistiques à trois niveaux (lemme, étiquette grammaticale et rôle), est lu au travers de l'automate et soumis à un algorithme de parcours délimitant la sélection de l'expression d'opinion. Le cas trivial de cette reconnaissance est un parcours depuis l'état initial jusqu'à l'état final conduisant à l'extraction d'un unique segment de phrase. Un cas bien plus fréquent est celui où plusieurs correspondances sont possibles. De multiples stratégies peuvent être envisagées pour ce parcours. Nous listons ici les principales :

1. extraire *toutes* les expressions, ce qui implique par conséquent un traitement post-extraction afin de choisir les expressions pertinentes ;
2. extraire les expressions qui ne sont *pas totalement recouvertes* par d'autres, afin d'éviter les redondances d'information ;
3. extraire les expressions qui ne sont pas totalement recouvertes par d'autres, et déterminer la pertinence des expressions se *recouvrant partiellement* selon un paramètre qui peut être la longueur (si nous considérons une expression plus longue comme plus informative), la précédence (si nous considérons une expression commençant plus tôt dans le texte comme plus à même de contenir l'information essentielle) ou simplement un score de priorité associé au patron.

Chacun de ces points présente des inconvénients pour la fouille d'opinion, par conséquent le choix d'une stratégie est guidé par certaines concessions que nous considérons acceptables pour notre application. Extraire les expressions se chevauchant (stratégies 1 et 2) permet bien évidemment de maximiser la couverture des sujets abordés, mais rend l'analyse de ces sujets difficilement lisible, car fortement sujette

à la présence de doublons. D'un autre côté, les conditions que nous pouvons imposer à l'algorithme de reconnaissance (à la manière de la stratégie 3) sont fortement dépendantes de connaissances linguistiques a priori, ce qui contredit dans une certaine mesure le principe « d'écouter et non questionner », que nous avons défini précédemment et que nous cherchons à respecter dans ce travail.

Afin d'apporter une analyse la plus juste possible aux utilisateurs de l'application, nous donnons priorité à la précision de l'extraction, c'est pourquoi la stratégie retenue impose une sélection conditionnelle – reposant sur la longueur et la précédence – des segments de phrase lorsqu'ils se chevauchent partiellement (stratégie 3).

Lier termes et subjectivèmes

Une fois la correspondance établie entre le patron d'opinion et un segment de phrase, une règle associée au patron est appliquée afin de lier termes et subjectivèmes formant une ou plusieurs opinions. En effet une expression d'opinion comprend généralement un terme et un subjectivème, cependant notre système de reconnaissance doit être en mesure de traiter les combinaisons de relations entre un à plusieurs termes avec un à plusieurs subjectivèmes, notamment pour prendre en compte les expressions de comparaison (« *Je préfère acheter en ligne plutôt qu'en magasin* »), d'énumération de termes (« *Bons conseils, service et réactivité* »), d'énumération de subjectivèmes (« *Le colis est mal emballé, sale et endommagé* ») et mise en opposition (« *Le service est un peu froid mais efficace* »).

Inférence de polarité

La dernière étape de cette extraction consiste à inférer la polarité de l'opinion. Selon son contexte, la polarité du marqueur d'opinion, projetée sur le sujet cible d'opinion, peut être conservée, inversée ou annulée. Dans le cas des opinions sans subjectivème, un patron peut également détecter des expressions d'opinion avec une polarité fixe.

Les éléments linguistiques guidant cette inférence de polarité sont appelés modifieurs (ou *shifters*) [Hatzivassiloglou and McKeown, 1997, Wilson et al., 2005, Taboada et al., 2011, Boubel, 2012] et doivent eux-mêmes faire l'objet d'une extraction préalable à l'identification d'opinion. En français, les modifieurs sont en grande partie couverts par les marqueurs de négation (« *ne...pas* », « *pas* », « *jamais* », « *sans* »), mais d'autres mots ou expressions peuvent également remplir cette fonction. C'est le cas des mots exprimant un manque ou au contraire une surabondance (« *pas assez de choix* », « *trop d'attente* »), des expressions rhétoriques dont la polarité dépend fortement du contexte de l'occurrence (« *Je n'ai jamais vu pareil accueil* »), ou encore des formulations ironiques (« *Personne ne nous aide, bravo le personnel!* » , « *Génial, mon colis est arrivé en trois morceaux...* »). Enfin, un dernier type de modifieur est le cas des formes conditionnelles et interrogatives, qui provoquent une annulation de la polarité.

Du point de vue de leur application dans des règles d'inférence de polarité, la prise en compte de ces modifieurs n'est pas simple, au sens où leur effet sur l'orientation sémantique ne dépend pas de leur seule présence mais suit une logique propre à chaque patron. Il est ainsi fréquent qu'un marqueur de négation n'entraîne pas d'inversion de polarité, ou que deux marqueurs de négation ne s'inversent pas l'un l'autre. L'utilisation de l'adverbe « trop », en particulier, est sujet à cette définition de règle spécifique, car celui-ci est amplificateur lorsqu'il est associé à un subjectivème (« *trop bien* », « *trop mauvais* ») et porteur d'opinion négative avec un sujet qui n'est pas a priori associé à une polarité (« *trop d'attente* »).

Les éléments linguistiques modifieurs de polarité ne sont pas la seule façon d'inverser l'orientation sémantique d'un subjectivème, et subséquemment de l'opinion. Nous avons vu précédemment qu'une expression d'opinion peut comprendre plusieurs subjectivèmes qualifiant un à plusieurs sujets. Or dans certains cas, tous les subjectivèmes de l'expression ne portent pas un jugement sur le sujet mais agissent comme des modifieurs pour ceux qui y contribuent. La phrase « *Difficile de trouver un bon produit* » illustre un tel cas, où « *difficile* » influe sur la polarité finale du marqueur d'opinion « *bon* » pour le sujet « *produit* ». Il est alors nécessaire de décrire dans le patron de détection correspondant quel subjectivème porte le jugement et lequel le modifie.

Chaîne de traitement globale

Afin de clarifier l'ensemble des processus intervenant dans la détection des opinions à l'aide de l'approche symbolique, nous détaillons la chaîne de traitement globale sur un schéma (figure 4.8). L'ordre dans lequel les différentes détections ou projections sont appelées est crucial, dans la mesure où chacune de ces « briques » dépend de la précédente. Il est par exemple impératif de tenir compte des négations afin de ne pas extraire des termes incohérents, ou encore d'identifier les subjectivèmes avant les négations puisque les premiers peuvent parfois contenir les seconds.

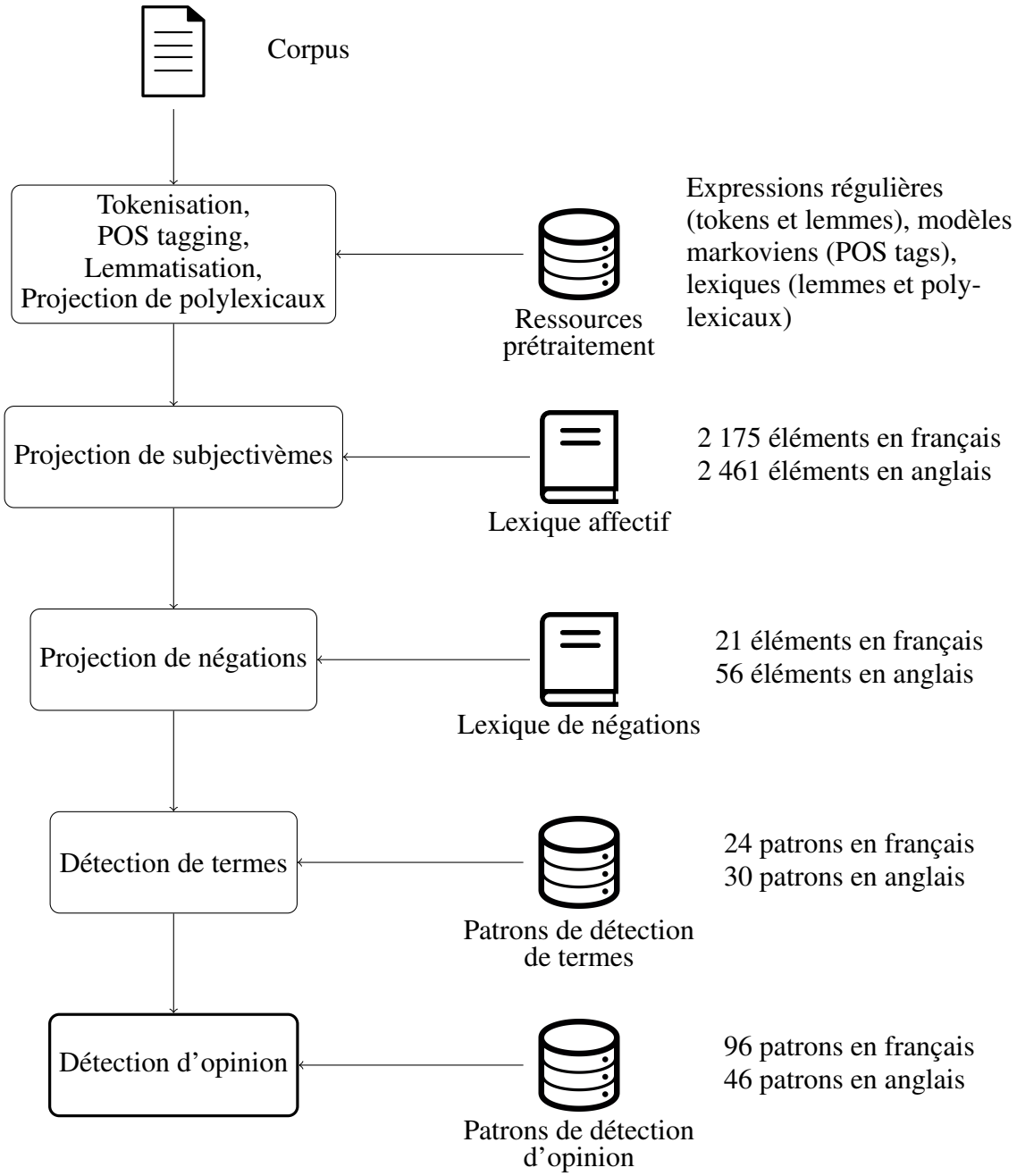


FIGURE 4.8 – Schéma de la chaîne complète de détection d'opinion au moyen de patrons

L'annotation de chaque métadonnée (tokens, POS tags, lemmes, polylexicaux, subjectivèmes, négations, termes et opinions) est associée à un segment du texte en utilisant le *framework* UIMA[2], qui offre en outre la possibilité de les consulter ou de les enrichir par la suite.

[2] https://uima.apache.org/

4.2.2 Intérêts et limites

Le choix d'utiliser une approche symbolique de la fouille d'opinion peut être, à notre sens, motivé par deux avantages : la robustesse des patrons de détection aux différents contextes lexicaux et l'explicabilité de leur déterminisme. En revanche, cette approche demande un effort de maintenance croissant.

Robustesse des patrons

Le fait de composer des patrons de détection d'opinion sur des niveaux plus abstraits que la forme lexicale, à l'aide des étiquettes grammaticales et des rôles pour l'opinion notamment, n'est pas seulement un moyen d'augmenter la couverture de ces patrons pour un corpus donné, mais rend également ces patrons robustes au changement de contextes lexicaux à travers les corpus d'avis sur différents domaines. En effet, nous avons vu que la dépendance au domaine des opinions est fortement liée aux subjectivèmes employés pour les exprimer. Or l'abstraction des formes lexicales de ces marqueurs d'opinion à un emplacement générique [*subjectivème*] rend la détection peu sensible au changement de domaine.

Par ailleurs cette robustesse s'étend, dans une certaine mesure, au multilinguisme. Nous n'affirmons pas ici que l'ensemble des patrons définis sont applicables d'une langue à une autre mais, pour les cas de langues grammaticalement similaires, certains patrons ne comportant pas d'élément lexical peuvent être adaptés simplement afin de fournir une première version d'une ressource de détection d'opinion.

Patron	Français	Anglais	Chinois
[ADJ, *subjectivème*] [*terme*]	bon produit	good product	好产品
[*terme*] [*subjectivème*]	conception défectueuse	design defect	设计缺陷
[*terme*] [*négation*] [ADJ, *subjectivème*]	conseiller pas enthousiaste	consultant isn't enthusiastic	顾问不热情大方

FIGURE 4.9 – Exemples de patrons applicables dans plusieurs langues

Les patrons définis portent en ce sens une promesse forte, qui est de formuler explicitement le phénomène d'expression d'opinion dans le langage. Tel le Saint Graal de notre champ de recherche, une ressource listant l'ensemble des structures du langage qui décrivent l'énonciation d'une opinion serait un atout inestimable. Tout comme l'objet de la légende arthurienne, construire cette ressource est bien sûr utopique et représente davantage un idéal auquel nous faisons référence lors de l'ajout d'un patron dans une ressource plus modeste : ce patron explique-t-il un phénomène global ou local ? Pourra-t-il être utilisé dans un autre domaine ? Représente-t-il dans chacune de ses occurrences une expression d'opinion ?

Déterminisme des patrons

Si la promesse d'une explication transparente de la notion d'opinion dépasse notre cadre de recherche, le caractère explicite des patrons de détection a cependant une conséquence directe très importante dans notre travail, et plus particulièrement dans le cadre industriel de celui-ci, qui est de *savoir expliquer l'erreur*.

Les deux erreurs possibles sont qu'une expression correspondant à une opinion ne soit pas détectée (nous parlons alors de faux négatif), ou qu'une expression détectée ne soit pas une opinion (nous parlons alors de faux positif). Dans les deux cas, cela amène les utilisateurs de l'application de fouille d'opinion à se questionner sur sa qualité et il est alors nécessaire pour l'entreprise de savoir expliquer ses défauts.

L'avantage d'une méthode symbolique déterministe dans ce cas est la possibilité de mettre en parallèle le texte relevé par l'utilisateur et le parcours de l'automate, et ainsi d'expliquer le défaut de détection, puis de corriger un patron fautif ou d'ajouter un patron manquant.

Amélioration continue et dérive des patrons

Les avantages que nous venons de citer sont toutefois ternis par la difficulté de maintenance de la liste de patrons de détection. Nous avons en effet montré que cette ressource nécessite une amélioration continue, puisque son exhaustivité comme sa précision peuvent être remises en question à chaque analyse. Rendre plus automatique l'extraction de nouveaux patrons de détection est un problème que nous avons considéré comme central dans notre travail, et nous le détaillons dans le chapitre 8.

Le manque d'exhaustivité vient principalement du fait que nous privilégions la précision de l'analyse à la quantité d'expressions détectées, ce qui se traduit lors de la définition des patrons par une préférence donnée aux éléments descripteurs moins abstraits et donc moins couvrants.

Par ailleurs, quelque soit le niveau d'abstraction d'un patron ajouté, il est attendu qu'une certaine *dérive* de son utilisation soit constatée, c'est-à-dire le fait que le patron ne détecte pas seulement des expressions d'opinion. Un élément de haut niveau d'abstraction sur le mot, par exemple une étiquette grammaticale, peut ainsi correspondre à une multitude de cas pour lesquels le mot est pertinent pour la détection, et une multitude de cas pour lesquels il ne l'est pas. L'équilibre entre le gain en couverture du patron et cette dérive constitue la principale difficulté de l'amélioration continue.

Enfin, le fonctionnement par patrons offre certes un cadre permettant de modifier directement une détection incorrecte, ce qui est grandement bénéfique, mais ce pouvoir de répercussion sur l'analyse vient au prix d'une vérification méticuleuse lorsque des modifications sont opérées. En particulier, il est impossible de s'assurer d'une dérive minimale en étudiant uniquement un ensemble restreint de phrases d'évaluation, nous utilisons donc pour cela un ou plusieurs corpus entiers sur lesquels nous mesurons les écarts de détection. Bien que nous ayons œuvré pour rendre cette phase plus automatique et plus ergonomique, celle-ci reste fastidieuse et de surcroît s'intensifie avec le temps puisque l'effort de vérification est croissant en fonction du nombre de patrons.

4.3 Synthèse

Dans ce chapitre, nous avons décrit en quoi consiste l'approche symbolique de la fouille d'opinion et nous avons expliqué plus particulièrement le fonctionnement de patrons de détection d'opinion tels que nous les utilisons.

- Les patrons sont formés d'éléments représentant un ou plusieurs mots sur trois niveaux d'abstraction : leur **forme lemmatisée**, leur **étiquette grammaticale** ou leur **rôle** dans la construction de l'opinion ;
- Nous utilisons un **automate** dont chaque patron correspond à un chemin de l'état initial à l'état final pour identifier les opinions dans le texte, qui doit par conséquent être représenté sur ces trois niveaux également ;
- Une **règle d'inférence** de l'orientation sémantique est associée à chaque patron, permettant de tenir compte des éléments linguistiques **modifiant la polarité** d'un marqueur d'opinion.

La souplesse de l'approche symbolique est un avantage notable pour une application commerciale utilisant un système de fouille d'opinion, cependant il convient de contrôler les changements contre-productifs que cette souplesse permet.

- Les patrons de détection sont **peu sensibles au domaine** du corpus ;
- Il est relativement aisé d'en constituer un ensemble pour un **nouveau langage**, *ex nihilo* ou en adaptant les patrons d'un langage existant ;
- De par leur déterminisme, les **erreurs** de détection peuvent être **expliquées** et **corrigées**...
- ... cependant les **effets de bord** de ces corrections, au même titre que les ajouts de patrons, sont fastidieux à contrôler, et le coût en **effort manuel** de cette maintenance croît à chaque modification.

5

Approche probabiliste de la fouille d'opinion

Le second paradigme que nous avons expérimenté pour la fouille d'opinion est l'apprentissage automatique et plus spécifiquement l'apprentissage supervisé. Dans ce chapitre, nous rappelons brièvement les principes de l'apprentissage supervisé, puis nous présentons en quoi cette approche demande de construire des ressources différentes de l'approche précédente pour la fouille d'opinion. Nous nous intéressons ensuite à l'application du paradigme probabiliste pour la fouille d'opinion à granularité fine, et nous comparons cette application avec celle de la modélisation par entité que nous avons proposé au chapitre 3. Enfin, nous identifions les avantages et inconvénients de cette approche dans notre cadre de travail.

5.1 Classification de documents pour la fouille d'opinion

La classification de documents par des méthodes d'apprentissage automatique est l'objet d'un large éventail de travaux en recherche d'information. Son application principale concerne la catégorisation de textes en groupes sémantiquement cohérents, tels que des articles journalistiques à propos d'un même sujet. L'idée d'adapter cette méthode à la fouille d'opinion est de considérer les documents de même orientation sémantique comme un groupe sémantiquement cohérent.

5.1.1 Classification

Le principe de toute classification supervisée est composé de trois étapes. La première est d'assigner à chaque élément d'un ensemble de départ une étiquette correspondant à une classe (étape d'annotation). Dans notre cas, les éléments sont les documents, et les classes les différentes orientations sémantiques que nous considérons, par exemple trois classes *positif*, *négatif* et *neutre*. La seconde étape (étape d'apprentissage) consiste à transformer ces éléments en une représentation vectorielle de leur propriétés, appelées traits de classification ou *features*. Dans le cas des textes, ces traits peuvent être les mots ou n-grammes de mots du document (nous parlons alors de classification de « sac de mots »), ou bien des propriétés moins dépendantes du contenu du texte telles que le nombre de phrases ou les types de ponctuations, voire totalement décorrélées du texte si nous disposons de métadonnées telles que la date d'émission ou l'auteur du document.

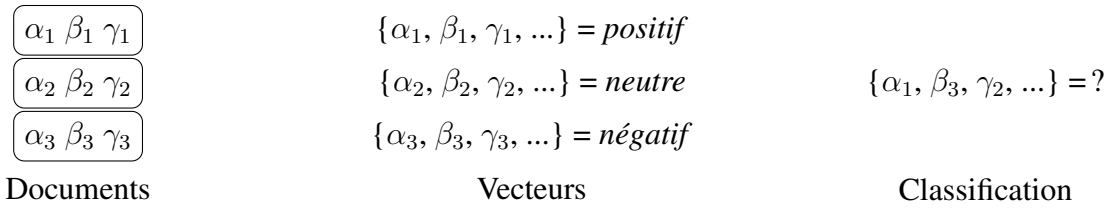


FIGURE 5.1 – Schématisation du principe de classification supervisée

À partir de cette représentation normalisée, un modèle statistique associe à chacune de ces propriétés un certain degré de dépendance vis-à-vis de la classe de l'élément. Ainsi à la dernière étape (étape de classification), lorsque des éléments inconnus – ou plus précisément leur représentation vectorielle – sont présentés à ce modèle, celui-ci peut leur attribuer la classe la plus probable.

Comme nous l'avons évoqué au chapitre 3, la classification en polarité de documents est une application très répandue, cependant nous questionnons son intérêt du fait du peu d'enseignements qu'offre cette méthode et de la redondance avec les notes attribuées par les utilisateurs dans le cas des avis clients. La catégorisation de textes en émotions, en revanche, peut être une source d'information intéressante et comme nous allons le voir dans cette section, reste un défi pour les méthodes employant un apprentissage automatique.

5.1.2 Catégorisation en émotions

Nous faisons état des travaux en catégorisation d'émotions – qui ne concernent pas nécessairement la fouille d'opinion parmi des avis clients – puis nous présentons les expériences que nous avons eu l'occasion de mener dans le cadre d'un atelier sur ce sujet.

Travaux existants

La catégorisation en émotions est un sujet bien moins couvert que la classification en polarité, ce qui peut être expliqué en partie par une demande moins forte d'application dans le milieu industriel. [Alm et al., 2005] identifient les émotions que communiquent les phrases de plusieurs œuvres de la littérature jeunesse en utilisant un classifieur linéaire. Les classes considérées sont les émotions primaires définies dans [Ekman and Oster, 1979] et les traits de classification sont choisis selon l'approche « sac de mots », restreinte à certaines étiquettes grammaticales (noms, verbes, adjectifs et adverbes), ainsi que les mots apparaissant dans un lexique affectif. [Wu et al., 2006] utilise un modèle SMM (*Separable Mixture Model*) d'une manière similaire sur un ensemble de phrases issues de dialogues entre étudiants et psychologues, et ne considèrent que trois classes (*heureux*, *malheureux* et *neutre*), ce qui rapproche ce travail d'une classification en polarité. [Yang et al., 2007] scinde les classes *positif* et *négatif* en quatre émotions (*bonheur*, *joie*, *tristesse* et *colère*) et compare des modèles linéaire SVM (*Support Vector Machine*) et séquentiel CRF (*Conditionnal Random Field* [Lafferty et al., 2001]), pour la catégorisation d'articles de blogs parmi ces quatre émotions.

Expériences

La classification de documents ou de phrases en émotions n'est pas actuellement une piste de recherche active pour notre application de fouille d'opinion, cependant dans le cadre de l'atelier DEFT[1] (Défi en Fouille de Texte) nous avons pu expérimenter ce type de méthode [Hernandez et al., 2015]. En particulier, nous avons pu comparer l'impact de différents lexiques affectifs lors de la reconnaissance d'émotions.

Le corpus d'évaluation de cet atelier est formé de 6 670 tweets en français sur le sujet de l'écologie répartis en 18 catégories sémantiques fines, auxquelles s'ajoute une classe « neutre » correspondant à l'énoncé d'une information objective. Quelques exemples de ces textes sont présentés dans le tableau 5.1. Ces catégories détaillées impliquent des variations relativement sensibles entre les classes pouvant entraîner des ambiguïtés, ce qui constitue la difficulté majeure de la tâche. En particulier, la résolution de ces ambiguïtés est complexe dans le cas où deux catégories sémantiquement proches ne sont pas représentées de façon équilibrée, car il peut exister un biais en faveur de la classe la plus présente.

Exemple de tweet	Classe sémantique
Les #océans, plus grand #écosystème de la planète, vont de plus en plus #mal http ://t.co/OfCnAgNtlH	DEPLAISIR
J'irai même plus loin : Que l'homme survive à l'effondrement de ses écosystèmes n'est qu'une hypothèse... http ://t.co/fCFWTDskd0 #URGENCE	INFORMATION
Les mini-maisons, moins chères et plus écologiques à la mode au #Québec http ://t.co/PbQXAeCKKl #polQC #construction	VALORISATION
Écologistes surveillés : une plainte dérange le SCRS http ://t.co/QPphdbvjOm	DERANGEMENT

TABLEAU 5.1 – Exemple de tweets du corpus DEFT 2015, associés à une catégorie sémantique fine

[1] https://deft.limsi.fr/2015/

Afin de catégoriser les tweets, nous avons employé un modèle SVM (plus précisément, l'implémentation LIBSHORTTEXT de [Yu et al., 2013] qui exploite la bibliothèque LIBLINEAR[2] [Fan et al., 2008]) dont les traits ont tout d'abord consisté en une représentation en sac de bigrammes de mots. Nous avons par la suite amélioré ces premiers résultats au moyen de lexiques existants ou construits de façon externe au corpus fourni, à savoir le lexique des affects LIDILEM [Augustyn et al., 2006], une traduction du lexique ANEW [Bradley and Lang, 1999], un lexique d'émoticônes issu de Wikipedia[3] et un lexique que nous avons extrait semi-automatiquement à partir d'un grand corpus externe de tweets, puis dans un deuxième temps par l'acquisition semi-automatique de mots sémantiquement liés aux catégories définies au sein du corpus d'entraînement.

Dans la mesure où ces lexiques présentent un recouvrement, nous avons jugé plus pertinent de comparer une méthode les utilisant tous (approche « Tous lexiques »), puis une méthode n'utilisant que les marqueurs d'opinions extraits du corpus fourni (approche « Lexiques endogènes »). Les résultats de ces deux approches ainsi que l'approche *baseline* sont indiqués dans le tableau 5.2.

		Baseline			Tous lexiques			Lexiques endogènes		
Classe	#Tweets	P	R	F1	P	R	F1	P	R	F1
Accord	153	65,96	20,26	31,00	61,54	26,14	**36,70**	54,28	24,83	34,08
Amour	8	0	0	0	0	0	0	0	0	0
Apaisement	9	0	0	0	0	0	0	66,66	22,22	**33,33**
Colère	206	63,21	32,52	42,95	67,29	34,95	**46,01**	72,22	25,24	37,41
Déplaisir	47	0	0	0	0	0	0	0	0	0
Dérangement	12	100	41,67	58,82	100	33,33	**50,00**	100	25,00	40
Désaccord	212	73,48	45,75	**56,40**	71,64	45,28	55,49	64,41	49,52	56,00
Dévalorisation	394	52,48	26,90	35,57	58,52	26,14	36,14	47,12	39,59	**43,03**
Ennui	4	0	0	0	0	0	0	0	0	0
Information	3531	68,20	91,31	78,08	69,03	94,00	**79,60**	67,57	85,43	75,46
Insatisfaction	9	66,67	22,22	33,33	100	22,22	**36,36**	0	0	0
Mépris	173	19,05	4,62	7,44	38,30	10,40	16,36	40,44	20,93	**27,58**
Peur	269	81,15	57,62	67,39	79,13	67,66	**72,95**	74,03	71,00	72,48
Plaisir	34	50,00	8,82	15,00	50,00	5,88	10,53	43,75	20,58	**28,00**
Satisfaction	73	64,86	32,88	**43,64**	66,67	24,66	36,00	61,53	22,22	32,65
Surprise négative	10	100	10,00	18,18	100	10,00	18,18	100	30	**46,15**
Surprise positive	4	0	0	0	0	0	0	0	0	0
Tristesse	34	0	0	0	57,14	11,76	19,51	50	17,64	**26,08**
Valorisation	1 491	60,51	47,48	53,21	64,97	46,14	**53,96**	52,79	40,57	45,88

TABLEAU 5.2 – Détail des résultats pour la tâche d'identification des catégories sémantiques fines sur le corpus DEFT 2015.

Nous constatons pour cette classification fine des tweets que l'approche de base utilisant une représentation en sac de bigrammes de mots fournit des résultats satisfaisants sur les classes les plus présentes, mais ne parvient pas à désambiguïser avec justesse les classes moins fréquentes. Les lexiques affectifs que nous avons utilisés permettent d'affiner dans une certaine mesure cette classification. Nous observons cependant deux tendances selon le type de lexique employé. Dans le cas des lexiques affectifs construits indépendamment du corpus fourni, la plupart des classes bénéficient d'un gain modéré tandis qu'en utilisant les lexiques issus du corpus de ce défi, quelques classes sémantiques sont particulièrement bien identifiées, au détriment des autres.

[2] http://www.csie.ntu.edu.tw/~cjlin/liblinear/

[3] https://en.wikipedia.org/wiki/List_of_emoticons

5.2 Fouille d'opinion à l'aide d'un étiquetage de séquence

Les méthodes d'apprentissage automatique de séquences offrent deux types d'application : dans un cas l'objet que nous cherchons à qualifier est la séquence entière, dans l'autre nous qualifions chacun des éléments de la séquence. Dans ce travail nous nous intéressons à l'application de l'étiquetage de séquence pour la fouille d'opinion à granularité fine, c'est pourquoi nous traitons uniquement du deuxième cas.

5.2.1 Étiquetage de séquence

Le paradigme de cette approche, dans le cas de l'apprentissage supervisé, reste similaire à celui de la classification de documents. Chaque élément est associé à un ensemble de traits que le modèle corrèle de manière probabiliste avec la classe de l'élément. La spécificité de l'étiquetage de séquence réside dans le fait que chaque élément n'est pas considéré indépendamment mais est reconnu parmi un voisinage d'autres éléments. Dans notre cas, c'est-à-dire la classification de mots au sein d'une phrase, il s'agit de prendre en compte une fenêtre autour de chaque mot pour en déduire le rôle dans la phrase. Or nous avons vu jusqu'ici à de multiples reprises l'importance du contexte pour la fouille d'opinion à granularité fine, ce qui explique l'attrait de cette méthode.

Principe

Parmi les techniques existantes en étiquetage de séquence, celle que nous avons le plus expérimenté dans nos travaux est l'apprentissage d'un modèle CRF (*Conditionnal Random Fields*, [Lafferty et al., 2001]). Le modèle CRF est une généralisation des modèles de Markov à états cachés (*Hidden Markov Models*, ou HMM), dans le sens où la probabilité de classification d'un élément dans un HMM ne dépend que des propriétés de cet élément et de la classe de l'élément précédent de la séquence, tandis que le modèle CRF offre la possibilité de faire influencer toutes les propriétés possibles de la séquence (propriétés d'éléments précédents, suivants, ou globales de la séquence) sur cette probabilité. En ce sens, [Lafferty et al., 2001] montrent que l'usage d'un HMM est subsumé par celui d'un modèle CRF.

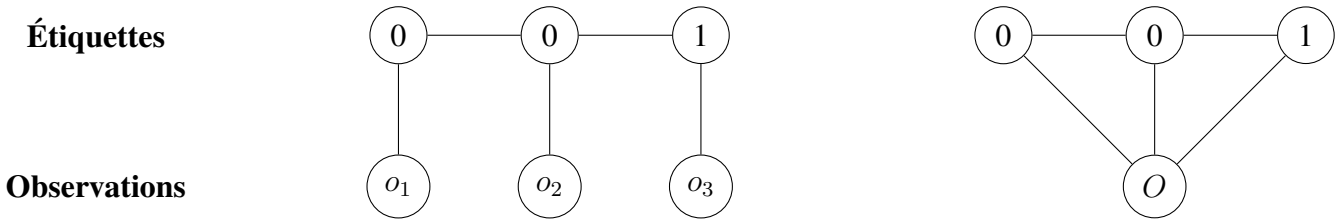


FIGURE 5.2 – Schéma simplifié d'un modèle HMM (gauche) et d'un modèle CRF (droite).

Ces dernières années, d'autres méthodes d'étiquetage de séquence ont été développées – ou plus largement expérimentées – dans le courant de la recherche sur l'apprentissage automatique, et ont montré des performances supérieures au modèle CRF, tels que les modèles LSTM (*Long Short-Term Memory* [Hochreiter and Schmidhuber, 1997]) sur les tâches d'étiquetage grammatical, d'extraction de syntagmes et de reconnaissance d'entités nommées [Huang et al., 2015]. Cependant, nous n'avons pas constaté lors de nos expérimentations de gain notable de performance sur les données d'évaluation dont nous disposons, mais seulement un temps d'apprentissage décuplé, c'est pourquoi nous avons fait le choix de ne pas nous concentrer sur ces méthodes reposant sur l'entraînement de réseaux de neurones.

Travaux existants

L'étiquetage de séquence par un modèle CRF est l'objet de nombreux travaux en fouille d'opinion à granularité fine. Nous notons ici les contributions principales et liées à notre travail.

[Choi et al., 2005] entraînent un modèle CRF pour réaliser une extraction des cibles d'opinion sur le corpus MPQA (*Multi-Perspective Question Answering*). Nous reparlons de ce travail plus en profondeur dans le chapitre 6, dans la mesure où il s'agit ici d'une approche hybride symbolique et probabiliste. Nous notons toutefois que les auteurs relèvent, lors de l'analyse des erreurs, que la structure complexe ou inattendue de certaines phrases ainsi que le manque d'exhaustivité du lexique affectif limitent les performances du modèle CRF, du moins pour les traits de classification utilisés. [Breck et al., 2007] effectuent sur le même corpus une classification des marqueurs d'opinion à l'aide d'un modèle CRF. Ces deux expériences sont donc complémentaires et montrent que le modèle CRF est davantage adapté à l'extraction des cibles.

[Li et al., 2010] identifient également les marqueurs d'opinion au moyen d'un modèle CRF mais infèrent de plus la polarité de ces marqueurs. Les traits de classification diffèrent de travaux précédents en incluant des informations issues d'une analyse en dépendances syntaxiques de la phrase. Le but final de ce travail est, comme dans notre cas, la production d'un résumé d'opinion à partir d'avis clients. [Jakob and Gurevych, 2010] réalisent une extraction des cibles d'opinion sur des corpus d'avis clients et mettent en avant la difficulté d'adapter un modèle CRF unique pour analyser plusieurs domaines. [Li et al., 2012] proposent pour cela de s'appuyer sur les dépendances syntaxiques afin d'extraire conjointement les cibles et marqueurs d'opinion dans un nouveau domaine, à partir d'un modèle entraîné sur un domaine connu, au sens où ces informations sont annotées. Enfin, de nombreuses participations aux ateliers SEMEVAL ABSA [Pontiki et al., 2014, Pontiki et al., 2015, Pontiki et al., 2016] utilisent cette méthode pour la tâche d'extraction de cibles d'opinion, que ce soit parmi des tweets ou des avis clients [Zhiqiang and Wenting, 2014, Hamdan et al., 2015, Brun et al., 2016].

5.2.2 Étiquetage de séquence d'opinion à granularité fine

De la même façon que nous avons décrit les patrons de détection en tant que ressource symbolique pour la fouille d'opinion à granularité fine, nous décrivons l'annotation de cibles d'opinion en tant que moyen de créer une ressource pour l'approche probabiliste du problème. Comparativement à la construction de patrons, l'inertie de l'annotation est extrême. La reconnaissance des cibles d'opinion nécessite d'annoter des centaines voire des milliers de phrases, ce qui représente un travail fastidieux. De ce constat, il découle que le choix de la modélisation est crucial avant d'investir le temps conséquent qu'implique cette annotation. Nous identifions trois facteurs influant sur ce choix : les éléments à annoter, la granularité de la séquence et la nature de l'annotation.

Éléments annotés

La première question est bien entendu celle des éléments que nous cherchons à extraire et, concomitamment, à annoter. Il s'agit en réalité d'un problème à deux facettes car deux objectifs sont à concilier. Nous souhaitons d'une part que notre modèle soit en mesure de reconnaître des mots ou segments de phrases pertinents pour notre analyse, et d'autre part que ceux-ci puissent être résumés à grande échelle. Pour cet aspect, la difficulté est de choisir entre l'échelle des termes, dont les occurrences sont plus simples à consolider, et des expressions d'opinion complètes, plus informatives, tel qu'illustré sur la figure 5.3.

Leurs	**tarifs sont compétitifs**			et	le	**service client est hors du commun**				
0	**1**			0	0	**1**				
Leurs	**tarifs**	sont	compétitifs	et	le	**service client**	est	hors	du	commun
0	**1**	0	0	0	0	**1**	0	0	0	0

FIGURE 5.3 – Exemple d'annotation par expression et par cible d'opinion

Granularité de l'annotation

Deuxièmement, puisque nous nous situons d'ores et déjà à un niveau intra-phrastique, la question de la granularité ne se pose que sur le fait de regrouper des mots préalablement à l'annotation, ou de considérer chaque mot indépendamment, comme présenté sur la figure 5.4. Dans notre application, la phase de prétraitement inclut en effet un certain nombre de regroupements polylexicaux ; à bas niveau tout d'abord, c'est-à-dire lors de la tokenisation, où des expressions polylexicales de toutes catégories grammaticales sont considérées comme un seul mot (« *en dépit de* », « *ni plus ni moins* », « *une fois pour toutes* », etc.), puis à un niveau supérieur, où des regroupements sont effectués entre mots, comme dans le cas des termes et des marqueurs d'opinion.

Leurs	**tarifs**	sont	compétitifs	et	le	**service client**	est	hors du commun
0	**1**	0	0	0	0	**1**	0	0

Leurs	**tarifs**	sont	compétitifs	et	le	**service**	**client**	est	hors	du	commun
0	**1**	0	0	0	0	**1**	**1**	0	0	0	0

FIGURE 5.4 – Exemple d'annotation par mots et par tokens

Dans notre cas, l'annotation est fortement guidée par la nature bruitée des documents, ou plus spécifiquement la qualité du prétraitement qui en résulte. S'il paraît intuitivement plus cohérent d'annoter les mots ou syntagmes par rôle, puisque ce sont bien les éléments de différents rôles (termes, subjectivèmes, négation) que nous cherchons à lier, les expériences d'extraction sur les données que nous traitons ont globalement montré qu'il est préférable de conserver une annotation au mot, et ainsi permettre au modèle probabiliste de « corriger », au sens de délimiter selon d'autres bornes, certaines erreurs commises lors des étapes préalables comme l'identification des termes.

Nature de l'annotation

Enfin, la nature de l'annotation doit également faire l'objet d'une étude pré-annotation, afin d'utiliser une modélisation correspondant au mieux aux données à extraire. Nous avons exposé au chapitre 3 la possibilité d'annoter les cibles de façon binaire, ou bien en utilisant une modélisation multi-classe spécifiant l'entité à laquelle la cible fait référence. À cela s'ajoute, dans les deux cas, la possibilité de tenir compte de la position des éléments de la séquence, notamment pour différencier les éléments aux extremums des segments relatifs à une opinion. À titre d'exemple le schéma d'annotation BIO (*Begin, Inside, Outside*) ne différencie que le premier élément, tandis que le schéma BILOU (*Begin, Inside, Last, Outside, Unit*) différencie le premier et dernier élément des expressions polylexicales ainsi que les éléments monolexicaux. Un exemple de chacun de ces schémas est indiqué sur la figure 5.5.

Leurs	**tarifs**	sont	compétitifs	et	le	**service client**	est	hors	du	commun
O	**I**	O	O	O	O	**I**	**I**	O	O	O	O
Leurs	**tarifs**	sont	compétitifs	et	le	**service**	**client**	est	hors	du	commun
O	**B**	O	O	O	O	**B**	**I**	O	O	O	O
Leurs	**tarifs**	sont	compétitifs	et	le	**service**	**client**	est	hors	du	commun
0	**U**	0	0	0	0	**B**	**L**	0	0	0	0

FIGURE 5.5 – Exemple d'annotation selon les schémas IO, BIO et BILOU

L'intérêt principal d'une telle approche est de désambiguïser les cas d'entités adjacentes. Au cours de nos expériences, nous n'avons pas constaté que ces distinctions améliorent notablement les performances de l'extraction, voire les dégradent. [Breck et al., 2007] font une observation similaire et indiquent que le faible taux d'expressions adjacentes peut expliquer le fait qu'un format binaire IO (*Inside, Outside*) modélise plus efficacement les données annotées.

Chaîne de traitement globale

Nous représentons le processus global de détection d'opinions au moyen d'un étiquetage de séquence sur la figure 5.6. Les étapes I, II et III correspondent respectivement à l'extraction des traits de classification, à l'entraînement du modèle puis à son utilisation pour la détection.

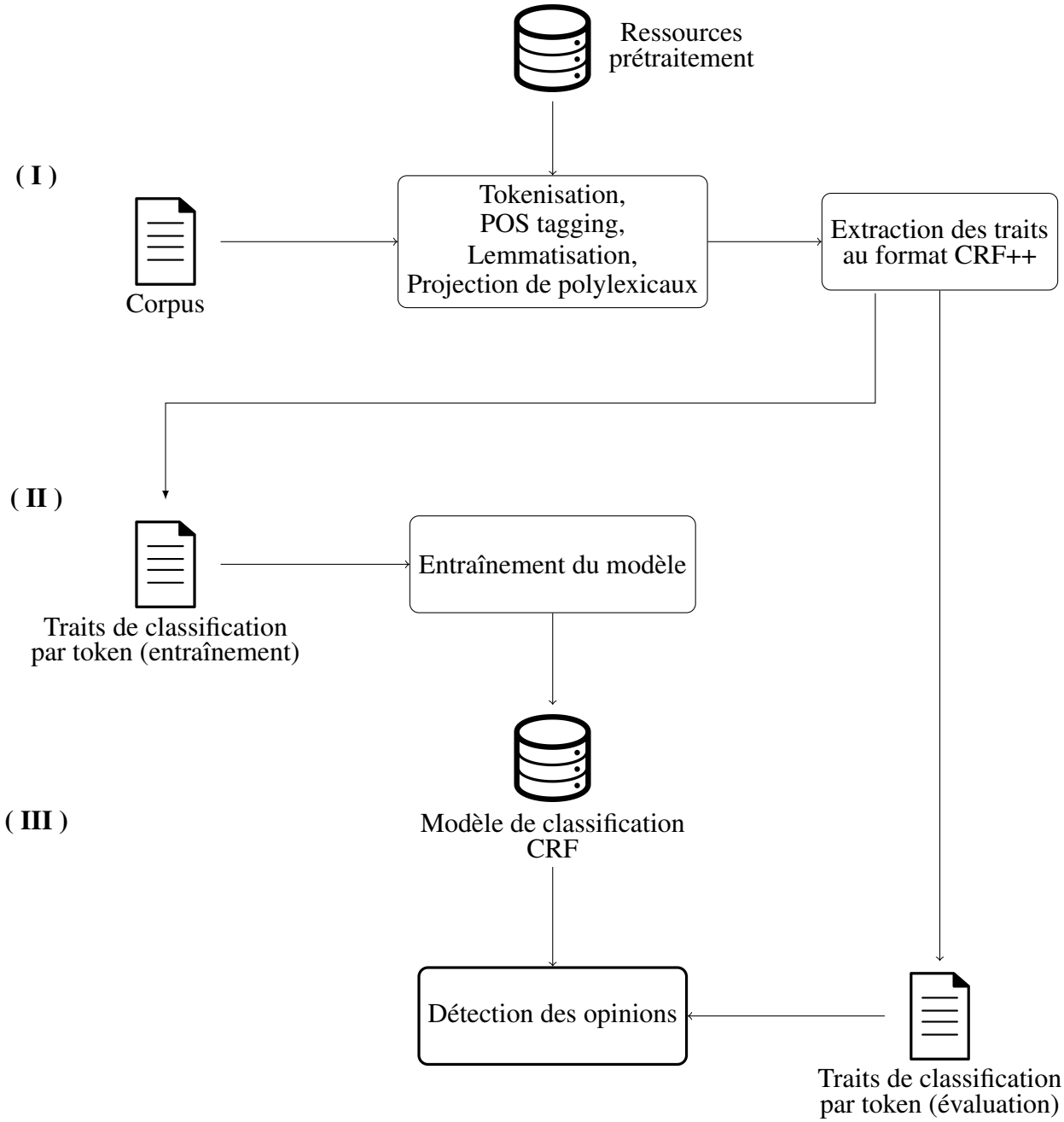


FIGURE 5.6 – Schéma de la chaîne complète de la méthode probabiliste

Tout comme dans le cas de la méthode symbolique, les métadonnées (ici tokens, POS tags et lemmes) sont annotés et manipulés à l'aide de UIMA, dont l'implémentation est en langage Java, ce qui n'est pas le cas de la librairie CRF++ [Kudo, 2005], librement disponible[4], que nous avons jugé la plus souple pour nos expériences. Ce schéma représente le processus suivi dans notre travail, mais ne traduit donc pas l'intégration du modèle probabiliste dans l'application finale, qui s'appuie sur des outils Java.

[4] https://taku910.github.io/crfpp/

5.2.3 Comparaison avec l'extraction de cibles

Comme nous l'avons décrit en section 3.3.1, une façon possible d'annoter les cibles d'opinion est de préciser l'entité à laquelle cette cible réfère (par la suite, « annotation par entité »), plutôt que d'indiquer simplement si l'empan de texte est une cible d'opinion ou non (par la suite, « annotation binaire »). Nous souhaitons ici valider deux hypothèses sur l'intérêt de cette annotation par entité, c'est-à-dire 1) une plus grande précision des cibles extraites et 2) le fait de pouvoir utiliser des corpus de plusieurs domaines pour entraîner le modèle d'étiquetage. Cette réutilisation permettrait de limiter le travail de maintenance et d'annotation, dans la mesure où un modèle unique est entraîné, et non un modèle par domaine.

Afin de comparer l'annotation binaire et par entité des cibles d'opinion, nous entraînons dans un premier temps les deux modèles sur le même ensemble d'avis, à savoir le corpus du domaine de la restauration issu des ateliers SEMEVAL – seuls les libellés associés aux cibles diffèrent pour les deux extractions. Nous réalisons cette expérience dans deux langues pour lesquelles ce corpus est disponible, soit le français et l'anglais, au moyen de l'outil CRF++. Les traits de classification pour chaque mot comprennent la forme lexicale, l'étiquette grammaticale et le lemme des mots compris dans une fenêtre de trois mots.

Annotation	Précision	Rappel	F1
Binaire	74,27	59,25	65,92
Par entité	**74,64**	**60,03**	**66,55**
p-value	$3,31\text{e}{-1}$	$3,29\text{e}{-1}$	

TABLEAU 5.3 – Extraction de cibles d'opinion utilisant l'annotation binaire et par entité parmi des avis clients sur des restaurants en français.

Annotation	Précision	Rappel	F1
Binaire	67,60	62,23	64,81
Par entité	**68,43**	**62,88**	**65,54**
p-value	$2,48\text{e}{-3}$	$3,91\text{e}{-3}$	

TABLEAU 5.4 – Extraction de cibles d'opinion utilisant l'annotation binaire et par entité parmi des avis clients sur des restaurants en anglais.

Les résultats, indiqués sur les tableaux 5.3 et 5.4, montrent que l'annotation par entité améliore non seulement la précision (+0,83 points de pourcentage (pp) pour l'anglais, +0,37 pp pour le français) mais également le rappel (+0,65 pp en anglais, +0,78 pp en français) pour l'extraction des cibles d'opinion. Le test de significativité, réalisé au moyen d'un test de Student, est positif pour une valeur de seuil standard de 0,05 dans le cas de l'extraction en anglais, mais ne permet pas de conclure sur le cas du français. Toutefois, indépendamment du degré d'amélioration observé, le fait que l'extraction par entité ne dégrade pas les résultats d'extraction reste un constat intéressant car cette annotation présente par ailleurs l'avantage de fournir un libellé plus informatif.

En reprenant un cadre d'expérimentation identique, nous souhaitons ensuite comparer les deux types d'annotation pour une application inter-domaine, comme nous l'évoquions en section 3.3.2. Pour cela, nous concaténons aux données annotées du domaine de la restauration des données provenant d'autres domaines dans chacune des deux langues, soit des avis à propos d'hôtels en anglais et des avis à propos de musées pour le français, et nous entraînons les modèles d'étiquetage sur cette concaténation. Il est à noter que cette concaténation n'est pas équilibrée, les données de ces domaines supplémentaires étant relativement restreintes. Les résultats sont ici toujours observés sur le corpus d'évaluation du domaine de la restauration.

Annotation	Précision	Rappel	F1
Binaire (R)	**74,27**	**59,25**	**65,92**
Binaire (R+M)	72,69	58,33	64,72
Par entité (R)	**74,64**	60,03	66,55
Par entité (R+M)	73,34	**61,57**	**66,94**

TABLEAU 5.5 – Extraction des cibles d'opinion inter-domaines parmi des avis clients en français sur des restaurants (R) et des musées (M).

Annotation	Précision	Rappel	F1
Binaire (R)	**67,60**	**62,23**	**64,81**
Binaire (R+H)	67,25	61,91	64,47
Par entité (R)	73,84	51,70	60,81
Par entité (R+H)	**74,37**	**52,67**	**61,66**

TABLEAU 5.6 – Extraction des cibles d'opinion inter-domaines parmi des avis clients en anglais sur des restaurants (R) et des hôtels (H).

Comme nous pouvons le lire sur les tableaux 5.5 et 5.6, l'annotation par entité est préférable dans le cas d'un apprentissage inter-domaine, et c'est à notre sens particulièrement dans ce cadre que ce type d'annotation se révèle être un atout. Nous pouvons voir en effet sur les deux premières lignes des tableaux que dans le cas de l'annotation binaire, ajouter des données d'un autre domaine au corpus d'entraînement dégrade les performances de l'étiquetage. Cela confirme d'ailleurs les observations de travaux précédents sur la spécificité du domaine en extraction de cibles d'opinion [Jakob and Gurevych, 2010]. Or ajouter des données de domaines différents annotés par entité produit des résultats opposés, comme l'indiquent les deux dernières lignes des tableaux.

Ces résultats tendent à confirmer notre hypothèse de partage des entités entre domaines, ou tout du moins que la spécification de cibles d'opinion par entité aide la désambiguïsation de contextes lors de l'étiquetage. Seule la mesure de précision dans le cas du français est dégradée dans ce cadre inter-domaine, ce qui peut être du, d'après nos observations des données, au fait que le domaine des musées partage des entités de façon moins directe avec le domaine de la restauration que celui des hôtels. Notamment, la notion de « service » ne couvre pas tout à fait les mêmes usages dans un restaurant ou dans un musée, ce qui peut expliquer des différences de contexte lexical non négligeables.

Afin d'analyser plus en détails ces résultats, nous réalisons une extraction entité par entité sur le corpus d'avis d'un domaine (restaurants), puis multi-domaine, dont les résultats sont indiqués sur les tableaux 5.7 et 5.8.

	Restaurants			Restaurants + Musées			
Entité	Précision	Rappel	F1	Précision	Rappel	F1	Gain (F1)
AMBIENCE	86,21	66,67	75,19	92,00	61,33	73,60	-1,59
DRINKS	73,33	32,35	44,90	72,22	38,23	50,00	+5,10
FOOD	65,37	53,00	58,54	72,96	53,62	61,81	+3,27
LOCATION	60,00	13,64	22,22	57,14	18,18	27,58	+5,36
RESTAURANT	56,45	51,47	53,85	62,50	51,47	56,45	+2,60
SERVICE	87,90	80,15	83,85	92,37	80,14	85,82	+1,97

TABLEAU 5.7 – Extraction d'entités d'opinion sur un corpus mono-domaine (restaurants), puis multi-domaine (restaurants et musées) en français

Cette expérience met en lumière le fait que les entités partagées, SERVICE et LOCATION, sont mieux reconnues (en français, leurs gains en mesure F1 respectifs sont de +1,97 et +5,36 et en anglais, +3,18 et +6,50). L'évolution des entités exclusives à un domaine sont moins évidentes. Non seulement les résultats de l'extraction multi-domaine peuvent améliorer ou détériorer l'extraction du domaine seul, mais ces différences ne sont également pas homogènes entre les deux langues. À titre d'exemple, le modèle entraîné sur deux domaines en français améliore l'identification de l'entité RESTAURANT (+2,6 en mesure F1), tandis que celle-ci est moins bien reconnus dans le cas du corpus en anglais (-1,05 en mesure F1).

	Restaurants			Restaurants + Hôtels			
Entité	Précision	Rappel	F1	Précision	Rappel	F1	Gain (F1)
AMBIENCE	76,60	61,02	67,92	80,43	62,71	70,48	+2,56
DRINKS	78,95	41,67	54,55	82,35	38,89	52,83	-1,72
FOOD	67,10	47,69	55,76	68,42	48,00	56,42	+0,66
LOCATION	100,00	40,00	57,14	58,33	70,00	63,64	+6,50
RESTAURANT	58,97	28,05	38,02	59,46	26,83	36,97	-1,05
SERVICE	78,95	69,44	73,89	81,44	73,15	77,07	+3,18

TABLEAU 5.8 – Extraction d'entités d'opinion sur un corpus mono-domaine (restaurants), puis multi-domaine (restaurants et hôtels) en anglais

En résumé, ces investigations sur des données réelles ont confirmé dans une certaine mesure que la nature complexe des expressions d'opinion nécessite une modélisation plus riche que l'annotation binaire actuellement utilisée. À notre sens, les observations que nous avons pu faire sur les entités définies pour ces corpus d'avis ne sont que les signes particuliers d'un besoin plus général d'une représentation cohérente des types de cibles d'opinion par catégories sémantiques. Une telle modélisation multi-classe de la fouille d'opinion à granularité fine dépasse quelque peu notre cadre de travail, et pourrait raisonnablement constituer un sujet de recherche à part entière.

5.2.4 Intérêts et limites

Du point de vue des ressources nécessaires, la mise en place d'une méthode d'extraction d'opinion probabiliste entraîne une situation radicalement différente de celle de l'approche symbolique, car l'abstraction que nous opérons à travers les patrons de détection, et qui constitue la grande force de l'approche symbolique, est ici réalisée automatiquement par le modèle probabiliste de l'apprentissage. Nous observons ici en quoi cela peut impacter positivement ou négativement notre travail de construction de ressources pour la fouille d'opinion.

Maintenance et non-dérive

L'extraction probabiliste apparaît tout d'abord comme une manière de répondre au problème de maintenance de la ressource sur laquelle notre application de fouille d'opinion s'appuie.

Premièrement, l'ajout de documents annotés à un corpus d'entraînement semble un moindre travail que l'ajout d'un patron à une liste de règles. D'une part, l'effort cognitif demandé lors de l'annotation des cibles d'opinion, que nous pouvons assimiler à une lecture de chaque phrase, est de moins grande ampleur qu'imaginer la représentation d'un patron de détection sur plusieurs niveaux d'abstraction ainsi que sa projection sur d'éventuelles phrases de contre-exemples. D'autre part, le besoin de validation est ici inexistant dans la mesure où celui-ci est confondu avec l'action d'annoter, là où nous avons vu que chaque nouveau patron de détection requiert une validation sur un grand ensemble de phrase.

Deuxièmement, il est raisonnable de supposer que ce travail d'annotation est de moins en moins important, la courbe d'apprentissage convergeant vers une asymptote représentant la performance maximum théorique du modèle d'extraction, là où nous avons constaté le besoin croissant de maintenance du modèle symbolique.

Par ailleurs, indépendamment du fait que l'action d'annoter soit considérée plus ou moins « difficile », il n'est pas demandé aux contributeurs une même rigueur lors de l'ajout de documents annotés que lors de la définition de nouveaux patrons de détection, les erreurs ponctuelles d'annotation pouvant être lissées par le nombre d'exemple fournis. Dans le cadre industriel dans lequel nous nous plaçons, cette souplesse signifie une plus grande capacité à délivrer des améliorations de l'application, et la possibilité pour des collaborateurs moins experts et par conséquent plus nombreux de contribuer à ces améliorations, à la différence de l'approche symbolique.

Dépendance au domaine

Il est toutefois une propriété fondamentale que ce changement de paradigme n'a pas conservé, à savoir l'indépendance au domaine du corpus analysé. En effet l'annotation étant réalisée au niveau lexical, nous ne bénéficions pas ici de l'abstraction des marqueurs d'opinion permettant une détection à travers différents domaines. Les avantages d'utiliser un modèle probabiliste que nous venons d'énumérer ne s'appliquent donc que dans le cadre d'un domaine particulier.

La construction d'une telle ressource, certes facilitée, est donc à répéter pour chaque nouveau domaine à traiter, correspondant dans notre travail à un secteur d'activité et à un type de corpus. Outre les contraintes techniques que cette multiplicité des modèles implique, telles que l'attribution d'un modèle à un nouveau corpus, ou la sélection d'une ressource existante pour l'ajout de données annotées, le fait de créer

régulièrement une nouvelle ressource est problématique dans la mesure où la phase initiale est justement le point faible de la construction d'un modèle probabiliste. Le problème de ce « démarrage à froid », ou « *cold start* » tel que nous pouvons le retrouver dans la littérature [Moghaddam and Ester, 2013], provient du fait qu'une ressource rassemblant des documents annotés ne peut être exploitée qu'à partir d'une certaine masse critique. Atteindre cette masse critique pour chaque domaine est une tâche fastidieuse et réduire cet effort constitue à notre sens le vrai défi de l'approche probabiliste.

5.3 Synthèse

Dans ce chapitre, nous avons étudié comment une approche probabiliste de la fouille d'opinion pouvait être mise en place, et en particulier ce que ce type de méthode implique sur les ressources employées pour l'analyse à granularité fine. La méthode état de l'art que nous avons adoptée est un étiquetage de séquence au moyen d'un modèle markovien CRF.

- Cette méthode permet d'annoter des **segments de texte**, et la classification de chaque élément tient compte du **contexte global** de la séquence ;
- Le choix des **éléments annotés**, de la **granularité** de la séquence et du **format** de l'annotation est crucial pour les performances de l'étiquetage.

Nous avons constaté que les avantages et obstacles de la construction d'une ressource spécifique à cette approche, c'est-à-dire l'annotation, sont exactement à l'opposé de ceux de l'approche symbolique.

- Le travail manuel nécessaire lors de l'**étape initiale** est conséquent, et doit être **répété** pour chaque nouveau domaine ;
- Une fois cette étape réalisée, la **maintenance** de la ressource est relativement **aisée**, permettant des améliorations plus fréquentes et auxquelles des collaborateurs non experts peuvent contribuer.

6

Comparaison et combinaison de méthodes

Au cours des deux chapitres précédents, nous avons étudié le fonctionnement des méthodes symboliques et probabilistes pour la fouille d'opinion à granularité fine. Cette étude nous a permis de définir les types de ressources nécessaires à leur mise en place, soit une liste de patrons à plusieurs niveaux d'abstraction dans le cas de l'approche symbolique, et des corpus de différents domaines annotés à l'échelle du terme pour l'approche probabiliste. En sus de ces ressources spécifiques, nous devons disposer dans les deux cas d'un lexique affectif et d'un lexique de modifieurs de polarité. Dans ce chapitre, nous souhaitons déterminer si l'une de ces deux approches subsume l'autre, ou s'il est préférable de tirer parti des deux paradigmes dans notre cadre de travail, ce qui impliquerait d'extraire les deux types de ressources.

Après une brève revue des comparaisons existantes et des systèmes hybrides dans la littérature, nous analysons les performances de deux systèmes employant chacun une approche, puis nous expérimentons plusieurs combinaisons de méthodes. Enfin, nous observons ce que chaque approche peut apporter, et en quoi une méthode hybride peut impacter notre travail de construction de ressources pour la fouille d'opinion.

6.1 Le meilleur des deux mondes

L'histoire des approches symboliques et probabilistes dans le domaine du traitement des langues est tumultueuse. Tout d'abord vues comme un moyen d'améliorer les résultats des méthodes probabilistes initiées dans les années 1950 et 1960, les méthodes symboliques sont à leur tour critiquées pour leur capacités de représentation et d'analyse du langage limitées à partir des années 1980, comparativement à l'approche probabiliste qui connaît alors une renaissance, portée notamment par des machines aux capacités de calcul grandissantes. Ce va-et-vient entre des visions opposées, mais orientées vers un même but, a donné lieu à nombre de travaux cherchant à comparer et à combiner ces méthodes. L'ouvrage *The Balancing Act : Combining Symbolic and Statistical Approaches to Language* [Klavans and Resnik, 1996], édité à la suite d'un atelier sur le sujet des méthodes hybrides en extraction d'information, répertorie certains de ces travaux et témoigne d'une volonté de rapprocher les communautés œuvrant sur chaque approche [Abney, 1996, Kapur and Clark, 1996, Price, 1994].

Nous relevons principalement trois pistes possibles pour la conception d'un système hybride : utiliser des éléments issus d'une méthode symbolique comme traits d'un modèle probabiliste, utiliser des méthodes probabilistes pour l'acquisition de ressources symboliques, ou employer les deux approches conjointement sans en fusionner le fonctionnement.

6.1.1 Éléments symboliques pour l'approche probabiliste

Une première façon de bénéficier des avantages des deux approches est d'enrichir les vecteurs d'un modèle probabiliste au moyen de structures lexico-syntaxiques identifiées par une analyse symbolique préalable. Ces structures peuvent alors être pondérées en fonction de leur représentativité pour les classes recherchées, et des relations supplémentaires peuvent être tissées entre les éléments différenciant chacune des classes.

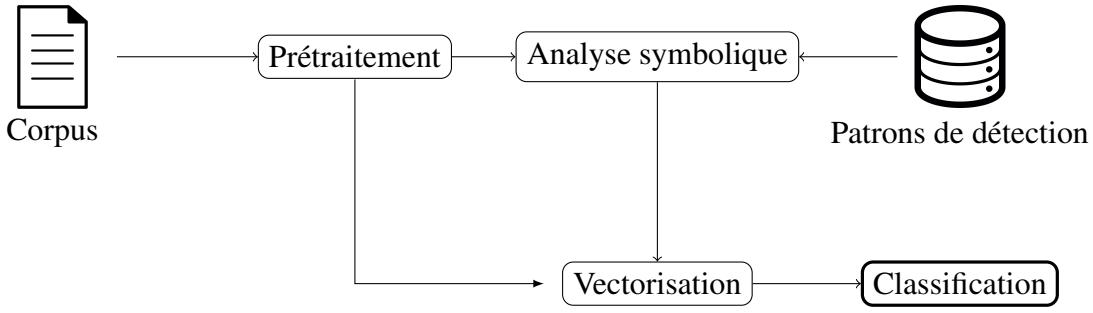


FIGURE 6.1 – Schéma d'une méthode hybride probabiliste utilisant des traits symboliques

[Hatzivassiloglou, 1996] montre l'apport de ce type de traits pour la classification d'adjectifs sémantiquement proches. Cinq niveaux d'information provenant d'une analyse linguistique symbolique sont comparés en tant que traits d'un modèle probabiliste. Ces informations vont de la simple étiquette grammaticale jusqu'à des patrons de détection conditionnant l'extraction d'adjectifs à des expressions particulières. Ces travaux montrent que chacune de ces informations améliore les performances globales du modèle probabiliste. [Alshawi, 1994] définit un système probabiliste de traduction de la parole en s'inspirant d'un premier système symbolique, et plus précisément en conservant les fonctionnalités fortes de celui-ci, telles que l'analyse de la structure du discours. Afin de parer les structures agrammaticales du discours pour un système de traduction de la parole, [Rosé and Waibel, 1994] identifient des syntagmes au moyen d'une méthode reposant sur des règles pour contraindre le champ des résultats possibles pour le modèle probabiliste. [Rosé et al., 2003] montrent dans le cas d'une application de questions – réponses qu'une approche hybride est préférable à une approche « sac de mots » ou une approche purement symbolique. Le système proposé est un arbre de décision prenant en compte d'une part les informations provenant de l'analyse en dépendances syntaxiques et d'autre part la prédiction d'un modèle bayesien sur une représentation en sacs de mots. [Pazienza et al., 2005] proposent une étude approfondie des forces et faiblesses de chaque approche pour l'extraction terminologique et analysent l'apport de l'utilisation préalable de patrons syntaxiques pour plusieurs méthodes d'extraction statistiques reposant sur une mesure d'association. [Choi et al., 2005] montrent qu'utiliser les traits de classification issus d'une première reconnaissance par des patron de détection améliore les résultats d'un étiquetage CRF seul pour l'identification de l'émetteur d'une opinion. [Abacha and Zweigenbaum, 2011] s'appuient sur un même principe pour la reconnaissance d'entités nommées dans le domaine médical, et fournissent à un modèle CRF les traits obtenus par une méthode symbolique qui, seule, présente des résultats moyens.

6.1.2 Méthodes probabilistes pour l'approche symbolique

Le second type d'application hybride que nous rencontrons dans la littérature concerne les systèmes à dominante symbolique s'appuyant sur des méthodes probabilistes pour créer et en compléter les ressources nécessaires.

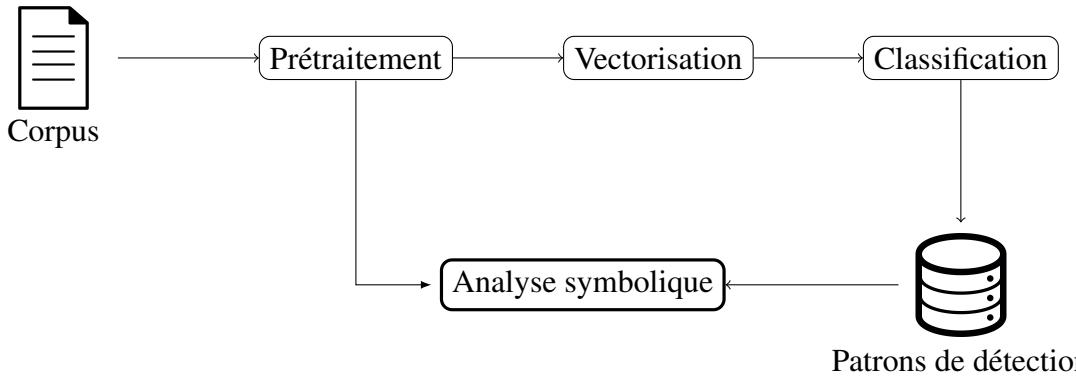


FIGURE 6.2 – Schéma d'une méthode hybride exploitant l'extraction de ressources symboliques au moyen d'une méthode probabiliste

L'algorithme BASILISK [Thelen and Riloff, 2002] associe un score de pertinence à des nouveaux patrons de détection extraits d'un corpus à partir de quelques patrons définis manuellement. Tout d'abord présenté pour l'extraction de mots sémantiquement proches, ce processus a été utilisé pour l'extraction de noms subjectifs [Riloff et al., 2003] et d'expressions relatives à une opinion [Riloff and Wiebe, 2003]. [Deng and Wiebe, 2015] proposent un système de fouille d'opinion à granularité fine dont le cœur est un modèle de logique sur des prédicats sémantiques, tels que $ami(x, y) \wedge insulte(z, y) \Rightarrow \neg ami(x, z)$ (« si x est l'ami de y, z insultant y n'est pas l'ami de z »), préalablement extraits par un classifieur SVM.

6.1.3 Utilisation conjointe des méthodes

Passage de relais

Le fonctionnement que nous identifions comme un « passage de relais » concerne les applications en deux étapes dont chacune fait appel à des méthodes d'approches différentes.

[Daille, 1996] propose une extraction terminologique sur des corpus en français reposant tout d'abord sur une sélection de termes candidats par des filtres linguistiques, puis sur un classement de ces candidats au moyen d'une mesure statistique d'association. [Westerhout and Monachesi, 2008] constatent qu'une approche purement symbolique d'extraction de définitions n'est pas suffisamment précise pour la construction semi-automatique d'un glossaire et emploient un classifieur bayesien dans le but de filtrer ces premiers résultats. Afin de réduire le désavantage de l'approche probabiliste que représente l'annotation, [Wiegand and Klakow, 2010] étudient dans quelle mesure une extraction reposant sur des patrons peut amorcer cette étape. Le système de classification en polarité de documents proposé se montre plus performant que l'unique application des patrons de détection et rivalise avec un système supervisé équivalent.

Traitement parallèle

Enfin, il se peut que les deux approches offrent tout simplement des résultats complémentaires, et soient utilisées de manière dissociée.

C'est ce qu'ont observé [Piao et al., 2005] en comparant l'extraction de termes complexes au moyen d'un étiqueteur symbolique et d'un algorithme probabiliste reposant sur la distribution des colocations. Les auteurs remarquent notamment une plus grande précision de l'approche symbolique et une plus large couverture des termes du domaine à l'aide des colocations. Plus spécifiquement, ils observent que les termes extraits par les deux approches se recouvrent relativement peu et que chaque approche semble plus adaptée pour un certain périmètre d'extraction. Du fait de cette complémentarité, le système hybride mis en avant ici consiste en réalité en une application parallèle des deux méthodes afin de bénéficier du meilleur des deux mondes. Nous avons également pu constater cette complémentarité lors d'un travail portant sur l'extraction des expressions d'opinion parmi un corpus de tweets [Lark et al., 2015], dont nous avons par ailleurs diffusé librement l'annotation[1].

6.2 Comparaison des approches

Il est indéniable, au vu des différentes contributions présentées précédemment, qu'une approche mêlant méthodes symboliques et probabilistes est une piste prometteuse pour de nombreux domaines de l'extraction d'information. Cependant, s'il est un enseignement que nous devons retenir de cette étude, c'est que chaque application nécessite une conception d'un système hybride particulier.

Afin de déterminer de quelle manière nous pouvons bénéficier de ce croisement de méthodes, nous commençons par comparer notre approche symbolique utilisant les patrons de détection avec l'étiquetage de séquence au moyen d'un modèle CRF pour l'extraction d'opinion sur un corpus commun.

6.2.1 Cadre expérimental

Nous effectuons cette comparaison sur les corpus disponibles d'avis clients annotés en cibles d'opinion pour l'atelier SEMEVAL ABSA 2016 en français et en anglais, mentionnés aux chapitres 3 et 5, traitant du domaine de la restauration. Les chaînes de traitement implémentées pour les deux approches sont celles illustrées précédemment, en section 4.2.1 pour la méthode symbolique et en section 5.2.2 pour l'étiquetage de séquence.

[1] github.com/ressources-tal/canephore

6.2.2 Évaluation

Tâche d'extraction

Les méthodes sont comparées sur deux tâches, afin de mettre en perspective leurs différences. La première est l'extraction des expressions cibles d'opinion, et la deuxième la spécifie en tenant compte de leur polarité. L'inférence de polarité est en effet un problème décorrélé de l'identification des cibles, c'est pourquoi rien n'indique qu'un système performant sur la première tâche l'est sur la seconde.

Résultats

Les performances de chacune des méthodes sur les deux tâches sont mesurées selon deux types de correspondance entre les éléments extraits et attendus, afin de pallier l'aspect subjectif de l'évaluation de la fouille d'opinion à granularité fine. Comme nous l'avons vu au chapitre 2, délimiter les termes portant une information optimale n'est pas évident et plusieurs solutions peuvent paraître acceptables. Nous proposons donc une mesure « exacte », comparant les cibles retrouvées avec celles annotées pour l'atelier SEMEVAL, et une mesure « permissive », comptant comme vrais positifs les cas de cibles chevauchant ou incluant une cible de référence. Les tableaux 6.1 et 6.2 indiquent respectivement les résultats obtenus sur les corpus d'avis en français et en anglais.

Information extraite	Méthode	Correspondance	Précision	Rappel	F1
Cibles	Patrons	Exacte	50,70	30,04	37,73
		Permissive	65,02	38,53	48,38
	CRF	Exacte	**74,27**	**59,25**	**65,92**
		Permissive	**77,75**	**67,43**	**72,23**
Cibles et polarité	Patrons	Exacte	44,13	26,15	32,84
		Permissive	57,04	33,80	42,45
	CRF	Exacte	**53,04**	**32,78**	**40,59**
		Permissive	**58,83**	**36,21**	**44,83**

TABLEAU 6.1 – Comparaison des méthodes d'extraction de cibles d'opinion en français

Information extraite	Méthode	Correspondance	Précision	Rappel	F1
Cibles	Patrons	Exacte	61,78	39,24	47,99
		Permissive	73,80	46,87	57,33
	CRF	Exacte	**67,60**	**62,23**	**64,81**
		Permissive	**78,78**	**71,63**	**75,04**
Cibles et polarité	Patrons	Exacte	**54,33**	34,50	**42,20**
		Permissive	**64,42**	40,92	50,05
	CRF	Exacte	44,38	**39,64**	41,88
		Permissive	56,68	**50,63**	**53,49**

TABLEAU 6.2 – Comparaison des méthodes d'extraction de cibles d'opinion en anglais

Le constat global de ces expériences est une supériorité quasi-systématique de l'approche probabiliste pour l'identification des cibles d'opinion. Seul le cas de l'extraction des cibles polarisées en anglais témoigne d'un avantage des patrons, tout en montrant des résultats modestes. Cet « anomalie » est en fait un signe particulier d'un phénomène constant sur l'ensemble des résultats. Nous observons que la différence entre l'extraction de cibles seules et polarisées est bien moindre pour l'approche symbolique (le modèle CRF affiche une perte de 25,33 points de pourcentage (pp) en mesure F1 « exacte » en français et de 22,93 pp en anglais, tandis que cette perte dans le cas des patrons est de 4,89 pp en français et de 5,79 pp en anglais), autrement dit l'inférence de polarité au moyen des patrons de détection semble plus fiable.

6.2.3 Analyse fine des erreurs

Malgré l'apparente dominance du modèle probabiliste sur le jeu d'expériences présenté, plusieurs observations portent à croire qu'un système hybride reste une piste prometteuse. D'une part, la robustesse des patrons en ce qui concerne l'inférence de polarité semble un levier intéressant pour renforcer le modèle CRF. D'autre part, les performances de ce modèle pour l'extraction des cibles polarisées restent moyennes, or c'est bien cette information qui est utilisée dans notre application *in fine*. Enfin, les résultats obtenus sur l'ensemble du corpus ne permettent pas de juger de la complémentarité des méthodes. Par conséquent, nous analysons ici les erreurs fréquentes et les expressions exclusivement identifiées par chaque approche, dans le but de concevoir un système hybride bénéficiant au mieux des apports de chacune.

Analyse de l'approche symbolique

En observant un échantillon de 500 phrases extraites du corpus évalué, nous discernons trois types d'erreurs majoritaires pouvant expliquer les faux négatifs parmi les phrases incorrectement analysées par notre méthode symbolique : les erreurs dues à un marqueur d'opinion inconnu, à un patron inconnu, ou à l'incapacité des patrons à décrire une opinion implicite. Des exemples de chacune de ces erreurs sont présentés dans les tableaux 6.3 et 6.4. Les erreurs dues à un prétraitement incorrect, bien que non négligeables, ne peuvent être considérées comme majeures.

Exemple de phrase	Erreur identifiée
*Quand il fait un peu frisquet, la **terrasse** est chauffée.*	« *chauffée* » n'est pas un marqueur d'opinion connu
*Nous avons découvert avec par hasard ce **cabanon de plage** et ce fut une excellente surprise.*	Aucun patron connu ne lie le terme « *cabanon de plage* » et le marqueur d'opinion « *excellente surprise* »
*Un **repas** qui n'a pas nécessité toute une implication du début à la fin.*	L'opinion est implicite

TABLEAU 6.3 – Exemples faux négatifs caractéristiques de la méthode d'extraction symbolique en français

La frontière entre ces types d'erreur est toutefois difficile à tracer. La notion d'opinion implicite peut recouvrir certains cas de patrons non connus, en particulier lorsqu'un marqueur d'opinion connu est situé loin du terme cible dans la phrase, mais peut également être assimilée à un manque d'un marqueur d'opinion dans le lexique, si nous considérons une grande partie de la phrase comme un long marqueur d'opinion polylexical. Parmi les cas de marqueurs d'opinion monolexicaux inconnus la cause d'erreur la plus récurrente est l'emploi d'adjectifs dont la polarité est ambigüe, comme les adjectifs indiquant une grandeur (« petit », « grand », « chaud », « froid », etc.) comme nous l'avons évoqué au chapitre 2.

Exemple de phrase	Erreur identifiée
*The **food** is not what it once was.*	« *not what it once was* » n'est pas un marqueur d'opinion connu
*The **subwoofer to the sound system** was located under my seat, which became annoying.*	Aucun patron connu ne lie le terme « *subwoofer to the sound system* » et le marqueur d'opinion « *annoying* »
*Overall, I would go back and eat at the **restaurant** again.*	L'opinion est implicite

TABLEAU 6.4 – Exemples faux négatifs caractéristiques de la méthode d'extraction symbolique en anglais

En ce qui concerne les faux positifs, les explications des échecs de cette méthode sont moins évidentes. Comme le montrent les résultats des correspondances « exactes » et « permissives » des tableaux 6.1 et 6.2, une partie importante des erreurs de reconnaissance sont des erreurs de délimitation. Par ailleurs, nous notons un grand nombre de cas où les cibles d'opinion extraites par la méthode symbolique ne semblent pas linguistiquement incorrectes, mais ne respectent simplement pas les critères de l'atelier SEMEVAL. En effet les guides d'annotation fournis pour l'atelier demandent de considérer certains termes non pas comme des cibles d'opinion mais comme des marqueurs d'une entité exprimée dans l'avis de façon globale. Ceci explique les exemples de faux positifs tels que « *présentations* », indiqué dans le tableau 6.5, et « *location* » indiqué dans le tableau 6.6, marquant respectivement, selon l'annotation de référence, une opinion sur l'aspect des plats et sur la localisation du restaurant. Enfin, notre projection de patrons ne tient pas compte du contexte de l'expression d'opinion, ce qui provoque des détections qui n'ont pas lieu d'être.

Exemple de phrase	Annotation de référence
Un restaurant pas très grand par la taille mais vraiment remarquable par la ***qualité de sa cuisine****.*	La cible attendue était « *cuisine* »
Fraîcheur des produits, de belles ***présentations****.*	« *présentations* » n'est pas une cible attendue
Sympa pour ***prendre un verre*** *après le boulot.*	Aucune cible attendue

TABLEAU 6.5 – Exemples de faux positifs caractéristiques de la méthode symbolique en français

Exemple de phrase	Annotation de référence
The best ***calamari in Seattle*** *!*	La cible attendue était « *calamari* »
Not the greatest ***location***	« *location* » n'est pas une cible attendue
This is sad for what once was one of the best ***places*** *you could ever eat.*	Aucune cible attendue

TABLEAU 6.6 – Exemples de faux positifs caractéristiques de la méthode symbolique en anglais

Analyse de l'approche probabiliste

De par la nature latente des mécanismes de détection de la méthode probabiliste, expliquer les causes d'échec de détection du modèle CRF sur un corpus en particulier paraît être une tâche vaine, c'est pourquoi nous préférons comparer les expressions extraites par cette approche avec celles obtenues par la méthode symbolique. Les différences entre celles-ci qui sont en défaveur de l'approche probabiliste (du point de vue de l'annotation de référence) montrent deux catégories d'erreurs fréquentes, illustrées dans les tableaux 6.7 et 6.8, à savoir les délimitations incohérentes des cibles d'opinion et les cas de polarité triviale non trouvée.

Les erreurs de délimitation incohérente se caractérisent par une cible détectée chevauchant la cible de référence ou incluse dans celle-ci et ne représentant pas une unité linguistiquement correcte, ou du moins pas aussi informative que la cible attendue. Cela constitue une première différence entre les cibles incorrectes identifiées par le modèle CRF et les syntagmes identifiés par les patrons pouvant certes dépasser les annotations de référence, mais proposant un élément exploitable pour l'analyse.

Lorsque le modèle d'étiquetage ne parvient pas à retrouver une opinion correctement identifiée par un patron de détection, que ce soit par une absence de reconnaissance ou un élément extrait de façon incorrecte, nous considérons cette erreur comme une « polarité triviale non trouvée ». Cette dénomination, certes peu flatteuse pour la méthode symbolique, traduit surtout le fait que les éléments de l'opinion attendue sont explicites et se trouvent dans une fenêtre de mots restreinte.

Exemple de phrase	Erreur détectée
*ils utilisent des **produits** de mauvaise qualité*	« *produits* » est extrait avec une polarité positive
*Nous ne nous attendions pas à de la grande cuisine dans cet **endroit** très touristique*	« *endroit* » est extrait avec une polarité positive
*C'était un très bon **déjeuner***	Aucune cible extraite
***vins de sicile** agréables...*	La cible détectée est « *vins* »
*île flottante et **tarte aux pralines** [...] que du bonheur*	La cible détectée est « *tarte aux* »
*La **tarte sablée aux agrumes** était également parfaite*	« *tarte sablée* » et « *agrumes* » sont détectés comme deux cibles distinctes, « *agrumes* » avec une polarité négative

TABLEAU 6.7 – Exemples de cibles d'opinion incorrectement extraites par le modèle CRF et correctement extraites par les patrons en français

Du point de vue d'une application commerciale, le fait que des erreurs sur des expressions simples peuvent apparaître est un frein évident à l'adoption de l'approche probabiliste, car celles-ci sont les plus visibles lors de l'analyse et sont les plus difficilement explicables. Dans les faits, ces erreurs peuvent être dues à des éléments de contextes rares, tel que le marqueur d'opinion « *chintzy* » (tableau 6.8), ou ambigus, comme le marqueur d'opinion « *touristique* » (tableau 6.7) dont la polarité dépend fortement du reste de la phrase. Un sur-apprentissage de contextes d'une polarité en particulier peut également être la cause d'une mauvaise inférence, à l'image du terme « *produits* » (tableau 6.7), présent dans 27 opinions positives sur 33 opinions dans le corpus d'entraînement en français.

Extrait de phrase	Erreur détectée
*the **service** is the worst I have experienced anywhere*	« *service* » est détecté avec une polarité positive
*Chintzy **portions***	« *portions* » est détecté avec une polarité positive
*I appreciate their **delivery** too.*	« *delivery* » est détecté avec une polarité négative
*the best summertime **deck** experience*	La cible détectée est « *summertime* »
*we splurged on the **Shilshole Sampler**...*	La cible détectée est « *Shilshole* »
*a beautiful assortment of enormous **white gulf prawns***	La cible détectée est « *gulf* »

TABLEAU 6.8 – Exemples de cibles d'opinion incorrectement extraites par le modèle CRF et correctement extraites par les patrons en anglais

D'un autre côté, nous gardons à l'esprit que ces écarts sont essentiellement des effets de bords d'une méthode permettant justement de prendre en compte des contextes ambigus et inconnus de la méthode symbolique, ce qui explique par ailleurs la supériorité globale du modèle d'étiquetage pour la tâche d'extraction de cibles d'opinion.

Cette comparaison renforce en réalité notre intuition qu'aucune des deux approches n'est idéale pour notre cadre d'application. Nous souhaiterions bénéficier des performances en extraction de cibles du modèle CRF, et de la stabilité des patrons pour l'inférence de polarité. Dans la section suivante, nous évaluons dans quelle mesure ce mariage peut être réalisé.

6.3 Combinaison des approches

De par les résultats obtenus au cours des expériences précédentes, il apparaît clairement qu'une combinaison des approches est une piste intéressante pour notre cadre d'application. Comme nous l'avons vu en première partie de ce chapitre, il existe cependant de nombreuses façons de mêler les deux approches. Nous présentons ici plusieurs méthodes hybrides, que nous évaluons sur l'extraction des cibles polarisées et une correspondance exacte avec l'annotation de référence.

6.3.1 Intersection et union des résultats

Nous réalisons en premier lieu une mise en parallèle naïve des résultats de chaque approche, en considérant leur intersection et leur union, indiquées dans le tableaux 6.9 et 6.10.

La méthode « CRF et Patrons » correspond simplement à une restriction des cibles identiquement extraites par les deux approches. L'union des résultats est mesurée de deux manières afin de tenir compte des conflits entre les éléments détectés par les patrons ou le modèle CRF. Ainsi pour chaque mot, la méthode « CRF ou Patrons » ne considère les résultats de l'approche symbolique que dans les cas où le modèle probabiliste montre une absence de reconnaissance, et la méthode « Patrons ou CRF » effectue l'inverse.

Méthode	Précision	Rappel	F1
CRF	53,04	32,78	**40,59**
Patrons	44,13	26,15	32,84
CRF et Patrons	**82,86**	17,29	28,61
CRF ou Patrons	42,52	**38,15**	40,22
Patrons ou CRF	19,87	36,07	25,62

TABLEAU 6.9 – Union et intersection des cibles d'opinion extraites en français

Méthode	Précision	Rappel	F1
CRF	44,38	39,64	41,88
Patrons	54,33	34,50	42,20
CRF et Patrons	**67,77**	22,77	34,09
CRF ou Patrons	40,85	45,86	**43,21**
Patrons ou CRF	21,84	**47,29**	29,88

TABLEAU 6.10 – Union et intersection des cibles d'opinion extraites en anglais

L'intersection montre une précision élevée mais un rappel très faible, traduisant le fait que les cibles identiques dans les deux jeux de résultats sont globalement correctes, mais ne concernent qu'une partie très limitée des éléments extraits, c'est-à-dire les opinions très peu ambigües. L'union des résultats présente, de façon attendue, une précision plus faible et un rappel plus élevé que chacune des approches. Toutefois, nous remarquons une nette baisse de précision de l'extraction lorsque les patrons sont appelés en premier (« Patrons ou CRF »), tandis que les résultats de la méthode « CRF ou Patrons » sont relativement similaires à ceux du modèle CRF seul. Cette différence indique que lorsqu'une cible est détectée à la fois par le modèle d'étiquetage et par un patron avec une polarité opposée, le premier est plus probablement correct.

6.3.2 Séparation des tâches

Nous avons observé lors de la comparaison des méthodes que le modèle d'étiquetage semble plus adapté à l'extraction des cibles, et les patrons à l'inférence de polarité. Il nous semble donc logique d'expérimenter une séparation des deux tâches.

Méthode	Précision	Rappel	F1
CRF	53,04	32,78	40,59
Patrons	44,13	26,15	32,84
Séparation	**59,32**	**36,51**	**45,20**

TABLEAU 6.11 – Utilisation séparée des méthodes d'extraction en français

Méthode	Précision	Rappel	F1
CRF	44,38	39,64	41,88
Patrons	**54,33**	34,50	42,20
Séparation	49,20	**43,95**	**46,43**

TABLEAU 6.12 – Utilisation séparée des méthodes d'extraction en anglais

Nous remplaçons pour cela la polarité des cibles d'opinion trouvées pour le modèle d'étiquetage par celle inférée par un patron, si celle-ci existe et si elle est différente. Les résultats de cette méthode (« Séparation »), indiqués dans les tableaux 6.11 et 6.12, sont très encourageants dans la mesure où la mesure F1 dépasse de plus de 4 points la meilleure approche pour chacune des langues (+4,61 pp en français et +4,23 pp en anglais). Seule la précision dans le cas de l'anglais est plus faible que celle de l'approche symbolique, ce qui est compensé par une augmentation de rappel substantielle (+9,45 pp).

6.3.3 Traits de classification provenant de patrons

Nous retenons de notre courte étude sur les méthodes hybrides pour l'extraction d'information que l'utilisation de traits de classification extraits par une approche symbolique peut améliorer les performances d'un modèle probabiliste dans plusieurs applications, y compris la fouille d'opinion. Nous souhaitons par conséquent évaluer l'ajout d'information provenant des patrons de détection pour notre modèle CRF.

Nous procédons de manière itérative, en incluant tout d'abord la notion de terme (« CRF + T »), puis de marqueur d'opinion (« CRF + S »), et la combinaison des deux (« CRF + TS »). La notion de négation est ensuite introduite. Dans la mesure où les marqueurs de négation sont utilisés conjointement aux marqueurs d'opinion, nous expérimentons un système considérant les deux notions (« CRF + NS »), puis l'ensemble de ces trois informations (« CRF + TNS »). Enfin, nous comparons ces systèmes avec un modèle employant non pas chacun des éléments mais l'expression d'opinion dans sont entièreté (« CRF + O ») lorsqu'une opinion est détectée par un patron. L'annotation pour chacune de ces méthodes est illustrée par une phrase d'exemple sur la figure 6.3.

Token mot	Lemme	POS	T	S	TS	NS	TNS	O	Annotation
Le	le	DET	-	-	-	-	-	-	O
personnel	personnel	NOUN	terme	-	terme	-	terme	opinion+	I+
reste	rester	VERB	-	-	-	-	-	opinion+	O
très	très	ADV	-	-	-	-	-	opinion+	O
sympathique	sympathique	ADJ	-	subj+	subj+	subj+	subj+	opinion+	O
et	et	CONJ	-	-	-	-	-	-	O
l'	le	DET	-	-	-	-	-	-	O
attente	attente	NOUN	terme	-	terme	-	terme	opinion+	I+
n'	ne	ADV	-	-	-	neg	neg	opinion+	O
est	être	VERB	-	-	-	-	-	opinion+	O
pas	pas	ADV	-	-	-	neg	neg	opinion+	O
excessive	excessif	ADJ	-	subj-	subj-	subj-	subj-	opinion+	O

FIGURE 6.3 – Détails des traits de classification utilisés pour chaque méthode ; toutes exploitent la forme lexicale, le lemme et l'étiquette grammaticale, auxquels sont ajoutées les informations de la colonne correspondante.

Le premier constat que nous pouvons faire à la lecture de l'ensemble des résultats (tableaux 6.13 et 6.14) est que l'ajout de chacun des traits est bénéfique, si nous les comparons à la *baseline* que constitue le modèle CRF entraîné sur la forme lexicale, le lemme et l'étiquette grammaticale des mots. Le comportement du modèle n'est pas exactement le même pour les deux langues lors de l'ajout d'une nouvelle information, cependant il est intéressant de voir que trois tendances se répètent dans les deux cas. Premièrement, les notions de terme et de marqueur d'opinion sont pour le modèle des appuis complémentaires au sens où les résultats des méthodes utilisant leur apport conjoint (« CRF + TS » et « CRF + TNS ») sont supérieurs à ceux des méthodes employant soit l'un soit l'autre. Deuxièmement, nous remarquons qu'un indicateur de la présence d'un marqueur de négation représente, contre-intuitivement, davantage un obstacle qu'une aide pour le modèle CRF. Cela renforce notre vision des négations comme des éléments particulièrement difficiles à intégrer au processus d'extraction car très ambigus, comme nous l'évoquions au chapitre 2.

Méthode	Précision	Rappel	F1
CRF	53,04	32,78	40,59
CRF + T	53,81	33,68	41,43
CRF + S	53,92	34,87	42,35
CRF + TS	55,15	35,92	43,50
CRF + NS	53,69	34,72	42,17
CRF + TNS	53,81	35,77	42,97
CRF + O	**61,24**	**38,15**	**47,02**

TABLEAU 6.13 – Intégration de traits symboliques pour l'extraction de cibles en français

Méthode	Précision	Rappel	F1
CRF	44,38	39,64	41,88
CRF + T	45,90	41,88	43,80
CRF + S	46,28	41,56	43,79
CRF + TS	47,00	42,36	44,56
CRF + NS	46,13	41,72	43,81
CRF + TNS	47,10	42,68	44,78
CRF + O	**52,04**	**46,66**	**49,20**

TABLEAU 6.14 – Intégration de traits symboliques pour l'extraction de cibles en anglais

Enfin, il est surprenant de constater que l'unique trait marquant une détection d'opinion (« CRF + O ») améliore les performances de l'étiquetage bien au-delà de toutes les autres combinaisons. Il n'y a en apparence qu'une différence minime entre l'information apportée par ce trait et ceux des méthodes « CRF + TS » ou « CRF + TNS », puisqu'une opinion est au sens de l'approche symbolique l'ensemble cohérent d'un terme et d'un marqueur d'opinion (et éventuellement d'une négation). Par ailleurs ces dernières proposent une information sous une forme plus détaillée. Ces résultats très positifs (+8,2 pp en précision et +5,37 pp en rappel sur le corpus français, +7,66 pp en précision et +7,02 pp en rappel pour l'anglais) traduisent à notre sens l'importance des patrons de détection, en ce qu'ils permettent de distinguer un ensemble d'indicateurs en simple cooccurrence d'une structure d'opinion cohérente.

6.4 Synthèse

Dans ce chapitre, nous avons étudié différentes conceptions de systèmes hybrides, c'est-à-dire employant une approche à la fois symbolique et probabiliste, pour des applications d'extraction d'information. Nous avons identifié trois paradigmes, dont nous nous sommes inspirés pour plusieurs aspects de notre travail.

1. Une première méthode consiste à **enrichir les traits de classification** d'un modèle probabiliste par des éléments issus d'une extraction symbolique ;
2. Dans le cas où l'approche dominante du système est symbolique, un modèle probabiliste peut être utilisé afin de **constituer des ressources plus efficacement** ;
3. Enfin, pour certaines applications les deux approches sont **complémentaires**, par conséquent la simple **combinaison des éléments extraits** se révèle plus pertinente que les résultats de l'une ou l'autre seule.

Dans un second temps, nous avons comparé les patrons de détection et l'étiquetage de séquence pour l'extraction de cibles d'opinion associées à une polarité. Nous avons ensuite cherché à établir un système hybride selon les paradigmes 1 et 3. Le paradigme 2 concerne davantage la construction de ressources, c'est pourquoi ce sujet n'est pas abordé ici mais au chapitre 8, « Extraction de patrons ».

- Sur les corpus étudiés, l'étiquetage de séquence se révèle bien plus performant que les patrons de détection pour l'extraction de cibles d'opinion, mais cette différence est atténuée lorsque leur **polarité est prise en compte** ;
- Un premier système hybride s'appuyant sur le **modèle CRF pour l'identification de cibles**, puis sur les **patrons pour l'inférence de polarité** améliore significativement les résultats des deux approches ;
- Un second système hybride incorporant la détection d'**expressions d'opinion de l'approche symbolique comme traits de classification** du modèle CRF présente des performances meilleures encore.

III

Expériences et contributions

— Il vaut mieux parfois regarder les choses plutôt que les posséder. Posséder, c'est s'inquiéter et beaucoup de paquets à transporter.

Tove Jansson (*Moomin et la Comète*)

7

Extraction de subjectivèmes

Le lexique affectif, regroupant les subjectivèmes d'une langue et parfois d'un domaine, est la pierre angulaire de tout système de fouille d'opinion. Comme nous l'avons vu au chapitre 3, cette ressource est utilisée quelle que soit la granularité de l'analyse. Les méthodes symboliques d'extraction d'opinion ne peuvent tout simplement pas fonctionner sans cet appui (*cf* chapitre 4), et, comme nous avons pu le vérifier au chapitre 6, celui-ci permet d'améliorer grandement les méthodes statistiques. Il est très rare qu'un système de fouille d'opinion proposé pour l'anglais dans la littérature n'ait pas recours à un des lexiques affectifs de référence – ou bien cela est fait précisément pour en tester la pertinence [Otmakhova and Shin, 2015] – tel que le lexique de subjectivité MPQA[1] [Wiebe et al., 2005], le lexique SENTIWORDNET[2] [Baccianella et al., 2010] ou le « lexique d'opinion »[3] présenté dans [Hu and Liu, 2004]. Pour ce qui est des autres langues, la majorité des travaux que nous rencontrons font appel à des ressources propriétaires ou spécifiques à un cadre d'application, en l'absence de références communément admises.

Du fait de l'importance primordiale de cette ressource, l'extraction de subjectivèmes fut le premier objet de nos efforts au cours de ce travail. Dans ce chapitre, nous indiquons les paradigmes existants, puis nous expliquons en quoi la création de lexiques affectifs requiert une évaluation particulière. Enfin nous présentons les résultats de plusieurs méthodes d'identification de marqueurs d'opinion et de leur polarité en français.

7.1 Paradigmes

La construction d'un lexique affectif est généralement motivée par une application finale, dont les objectifs imposent certaines contraintes qui lui sont propres. Pour cette raison, il n'est pas déraisonnable de penser qu'il existe autant de méthodes d'extraction que de lexiques. À travers l'étude des procédés existants, nous distinguons cependant deux familles de méthodes : celles reposant sur un corpus, au sein duquel des marqueurs d'opinion sont extraits, et celles dont l'objectif est davantage d'étendre un lexique existant, en recherchant dans une ressource externe de nouveaux éléments par similarité sémantique ou orthographique avec des marqueurs connus.

7.1.1 Construction de lexique à partir d'un corpus

La recherche de subjectivèmes dans un corpus repose sur l'idée que ceux-ci se trouvent au sein de structures particulières du discours et peuvent donc être identifiés par leur contexte. Ces structures sont représentées soit par des patrons lexico-syntaxiques, soit par un modèle de classification supervisée. Dans les deux cas, il est souvent nécessaire de définir préalablement quelques marqueurs initiaux, dont il convient de valider la pertinence [Vincze and Bestgen, 2011], à partir desquels les structures caractéristiques sont reconnues.

Extraction par patrons lexico-syntaxiques

[Hatzivassiloglou and McKeown, 1997] définissent quelques patrons de conjonction liant deux adjectifs afin de créer des paires d'adjectifs dans un corpus journalistique, selon l'hypothèse qu'une énumération en conserve la polarité (« *pratique, rapide et économique* ») et qu'une mise en opposition l'inverse (« *beau mais cher* »). Les paires sont ensuite regroupées par classification linéaire. La polarité de quelques adjectifs connus est alors propagée dans chaque groupe, produisant rapidement un lexique affectif (d'adjectifs uniquement) avec une précision satisfaisante. [Kanayama and Nasukawa, 2006] ont appliqué une méthode similaire pour l'extraction de subjectivèmes en japonais. [Ding et al., 2008] étendent ce fonctionnement à l'extraction de verbes et de noms, et l'appliquent au sein de corpus d'avis clients afin de détecter plus particulièrement les marqueurs d'opinion non triviaux, tels que les adjectifs ambigus ou les formes verbales dont la subjectivité dépend du terme qu'elles qualifient. [Riloff et al., 2003] utilisent l'algorithme BASILISK

[1] mpqa.cs.pitt.edu/lexicons/subj_lexicon/

[2] sentiwordnet.isti.cnr.it/

[3] www.cs.uic.edu/~liub/FBS/sentiment-analysis.html#lexicon

[Thelen and Riloff, 2002], capable d'identifier des patrons lexico-syntaxiques à partir de quelques mots de départ, ou « graines », afin d'extraire distinctement des noms faiblement et fortement subjectifs d'un corpus journalistique. Les patrons sont pondérés en fonction de leur spécificité aux contextes des noms « graines », et ceux présentant un fort potentiel d'extraction sont projetés sur le corpus afin de découvrir des nouveaux noms. [Huang et al., 2014] expérimentent un procédé similaire pour la construction d'un lexique affectif en chinois depuis un corpus d'interactions sur le réseau social Weibo[4]. Tout comme dans [Riloff et al., 2003], la méthode d'extraction par patrons proposée dans ce travail n'est pas utilisée pour l'inférence de polarité des nouveaux mots, qui est effectuée par classification. À partir des relations de dépendances associant un sujet à un marqueur d'opinion, [Qiu et al., 2009] identifient à la fois de nouveaux sujets et de nouveaux marqueurs dans un corpus d'avis clients. Des règles supplémentaires s'appliquant aux éléments extraits sont cependant nécessaires pour inférer la polarité des marqueurs, notamment pour prendre en compte l'effet des négations.

Classification supervisée

[Li et al., 2012] étendent un lexique de marqueurs d'opinion « stables », c'est-à-dire fréquents dans plusieurs domaines, de manière distincte pour chaque domaine d'application. L'extension est réalisée par classification en considérant les premiers éléments stables comme exemples positifs d'un modèle d'amorçage RAP (*Relational Adaptive bootstraPping*), puis en retenant par itérations les marqueurs extraits avec la plus grande probabilité. [Yang et al., 2014] utilisent un modèle LDA (*Latent Dirichlet Allocation*[Blei et al., 2003]) pour rechercher des mots sémantiquement proches de graines définies manuellement exprimant une émotion dans le corpus SEMEVAL 2007. [Staiano and Guerini, 2014] exploitent un corpus journalistique annoté par les internautes[5] selon l'émotion suscitée par chaque article afin d'en extraire des marqueurs statistiquement caractéristiques de chaque émotion. [Hamilton et al., 2016] utilisent la structure en domaines du réseau social Reddit[6] pour faire émerger des lexiques affectifs spécifiques à chacun. La méthode proposée construit tout d'abord un réseau sémantique sur la base d'une vectorisation des contextes de subjectivèmes graines, puis infère la polarité des nouveaux mots selon leur distance aux graines positives et négatives, de façon similaire à [Turney, 2002].

Classification non supervisée

[Choi and Cardie, 2009] adaptent un lexique affectif général à un domaine spécifique afin d'améliorer l'extraction d'opinion dans ce domaine. Bien qu'il ne s'agisse pas à proprement parler de la construction d'un lexique, il est intéressant de constater leur utilisation d'un calcul par contraintes pour la ré-attribution de polarité des marqueurs d'opinion ambigus. [Jijkoun et al., 2010] réalisent une extraction de mots subjectifs par différence statistique en supposant que des expressions caractéristiques de l'opinion sur un sujet sont plus fréquentes dans un corpus du domaine spécifique à ce sujet que dans un corpus général. Un processus similaire est employé par [Maks and Vossen, 2012] pour la constitution d'un lexique affectif en néerlandais. [Pak and Paroubek, 2010] relèvent les cooccurrents d'émoticônes parmi un grand nombre de tweets afin de construire un lexique affectif en français. Les émoticônes sont en effet des marqueurs d'opinion particulièrement fiables car peu ambigus autant d'un point de vue de la subjectivité que de la polarité. Par ailleurs, ces symboles sont internationalement compris et utilisés, par conséquent cette méthode est potentiellement applicable à différents langages. [Feng et al., 2011] proposent de ne pas limiter le lexique affectif à des marqueurs qualifiant directement une cible d'opinion, et comparent plusieurs méthodes non supervisées pour assigner une valeur subjective à un vaste ensemble de mots considérés habituellement comme neutres ou objectifs.

[4] tw.weibo.com
[5] www.rappler.com
[6] www.reddit.com

7.1.2 Extension et adaptation de lexique

Les méthodes d'extension et d'adaptation de lexique diffèrent de celles exploitant un corpus par le fait que la similarité entre les éléments recherchés et ceux de référence ne dépend pas d'un contexte mais de propriétés propres aux unités lexicales, le plus souvent indiquées dans une ressource externe.

Liens sémantiques

Le principe fondamental de l'extension de lexique reposant sur les liens sémantiques est une conservation de la polarité parmi les éléments partageant une relation de synonymie, et une inversion dans le cas des éléments antonymes. Cette méthode de construction de lexique affectif n'est donc possible que s'il existe une ressource répertoriant ces liens sémantiques, tel que le réseau WORDNET [Fellbaum, 1998] en anglais.

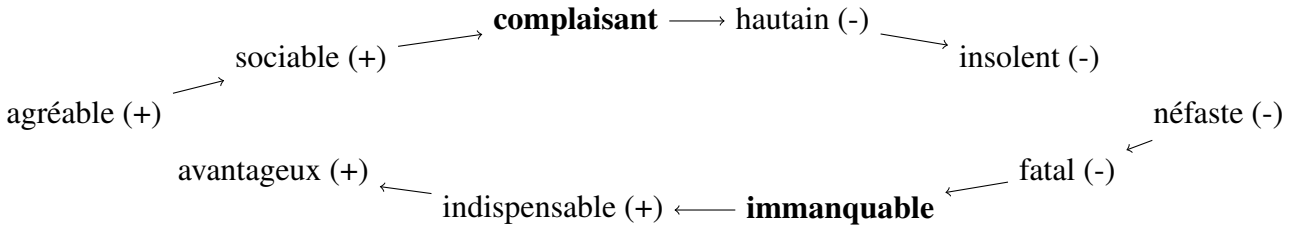


FIGURE 7.1 – Schéma de la propagation de polarité dans un réseau de synonymes

[Hu and Liu, 2004] infèrent ainsi à l'aide de WORDNET la polarité d'adjectifs proches de termes mais inconnus d'un lexique affectif. [Kim and Hovy, 2004] proposent de retrouver la polarité globale d'une phrase en utilisant également ces liens sur les adjectifs, verbes et noms apparaissant dans celle-ci. La ressource SENTIWORDNET [Esuli, 2008], aujourd'hui très régulièrement utilisée pour les travaux de fouille d'opinion en anglais, est née de ce besoin d'étendre les lexiques initiaux constitués manuellement. Ce lexique est en quelque sorte une version parallèle de WORDNET, incluant une valeur de polarité pour chaque entrée, détaillée en deux scores normalisés (compris entre 0 et 1) qui représentent la connotation positive et négative du mot. Ce lexique a été par la suite enrichi [Baccianella et al., 2010], concomitamment à la mise à jour de la ressource WORDNET.

Une telle ressource n'est malheureusement pas disponible en français actuellement et, à notre connaissance, les adaptations existantes de WORDNET librement disponibles sont incomplètes. Nous constatons par exemple que les ressources WOLF[7] (WordNet Libre du Français) ou WoNeF[8] (WordNet Français) comportent de nombreuses entrées en anglais non traduites et ne proposent pas certaines des relations sémantiques de WORDNET, comme les antonymes. Par ailleurs, les expériences que nous avons pu conduire en recherche de marqueurs d'opinion par cette méthode ont rapidement montré des limites en précision, du fait de la dérive sémantique des synonymes. Les liens de synonymie retrouvés ainsi dans le thesaurus OPENOFFICE[9] révèlent parfois des marqueurs ambigus inversant la polarité propagée (en gras sur la figure 7.1), ou bien des mots ne correspondant pas au sens porté dans les avis clients, voire de polarité inverse, tels que « *implorer* », synonyme de « *adorer* », ou « *aventure* », associé à « *catastrophe* » (*cf* tableau 7.1).

Marqueur	Synonymes
adorer	glorifier, idolâtrer, déifier, implorer, supplier
génial	ingénieux, habile, imaginatif, astucieux, pharamineux
catastrophe	apocalypse, fin, bérézina, aventure, bide

TABLEAU 7.1 – Exemples de synonymes de marqueurs d'opinion connus

[7] alpage.inria.fr/~sagot/wolf.html
[8] wonef.fr
[9] www.openoffice.org/fr

Traduction

Dans l'optique d'acquérir semi-automatiquement des ressources pour la fouille d'opinion dans une langue inconnue, nous nous intéressons aux travaux de traduction de lexiques existants. Il peut paraître en effet plus efficace de traduire les mots déjà sélectionnés pour leur pertinence dans une autre langue que de réitérer le processus global d'extraction de marqueurs d'opinion dans la langue d'étude.

Une première façon de réaliser cette traduction est d'utiliser un simple dictionnaire bilingue et de reporter les valeurs de polarité des éléments dans la langue source sur leurs correspondants dans la langue cible [Das and Bandyopadhyay, 2010]. Lorsque des ressources linguistiques supplémentaires sont disponibles dans la langue cible, tels que des adaptations de WORDNET, ceux-ci peuvent être alignés avec les équivalents en langue source afin d'améliorer la précision des traductions, ou d'étendre la couverture de l'inférence de polarité [Das and Bandyopadhyay, 2010, Perez-Rosas et al., 2012, Chen et al., 2014]. Nous pouvons cependant constater que les traductions peuvent mener, de la même façon que les liens de synonymie, à une dérive du sens des éléments initiaux. Dans le cas de [Chen and Skiena, 2014] notamment, ayant mis à disposition des lexiques affectifs pour 136 langues à partir d'alignements des entrées du Wiktionnaire[10] et de traductions automatisées par l'interface de Google[11] dédiée, les éléments extraits sont quelque peu discutables (en français par exemple, « *tout* », « *retraité* » ou « *aptitude* » sont considérés comme marqueurs d'opinions négatifs, et « *sans* » , « *capitale* » ou « *traitement* » comme marqueurs positifs).

Une manière certes naïve de filtrer les résultats d'une traduction automatique est de vérifier la cohérence entre l'élément en langue source et la traduction inverse de l'élément en langue cible, ce que nous avons expérimenté pour le prototypage de systèmes de fouille d'opinion en plusieurs langues. Le tableau 7.2 illustre ceci sur des exemples de marqueurs d'opinion traduits du français en allemand.

Marqueur d'opinion	Traduction (source → cible)	Traduction inverse (cible → source)
amical	*freundlich*	*amical*
coincer	*Marmelade*	*confiture*
détester	*Hass*	*haine*

TABLEAU 7.2 – Exemples de traductions de marqueurs d'opinion connus du français à l'allemand par une approche naïve

Morphologie

Enfin, l'ajout de mots et d'expressions au lexique affectif peut être justifié par une similarité morphologique avec les entrées de la ressource existante. Nous distinguons cependant parmi les méthodes reposant sur cette similarité deux objectifs distincts, à savoir la prise en compte des erreurs orthographiques et la génération de marqueurs d'opinions à l'aide d'affixes.

Les erreurs orthographiques sont relativement fréquentes parmi les documents générés par des utilisateurs. Celles-ci sont des barrières pour la construction de lexique car elles multiplient le problème de la couverture d'une part, et d'autre part bloquent l'inférence de la polarité au moyen de réseaux de connaissances puisque les mots incorrectement orthographiés ne sont pas présents dans les ressources externes. Il s'agit donc soit de corriger ces erreurs afin d'obtenir une forme reconnue [Dey and Haque, 2009], soit de répertorier des variantes morphologiques incorrectes [Vernier, 2011].

[Brun and Gala, 2012] diffusent à l'ensemble d'une famille morphologique la polarité d'un adjectif connu. Une famille morphologique est définie dans ce cas comme l'ensemble des mots pouvant être construits par dérivation, c'est-à-dire par l'ajout d'affixes à partir d'une même base. L'ajout de préfixes caractéristiques de l'antonymie (« a- », « in- », « dis- », etc.) permet également de générer des marqueurs d'opinion de polarité inverse à partir d'éléments initiaux [Godbole et al., 2007, Das and Bandyopadhyay, 2010, Mohammad et al., 2009].

[10] fr.wiktionary.org

[11] cloud.google.com/translate

7.2 Évaluation

Dans la suite de ce chapitre, nous proposons d'expérimenter certains de ces paradigmes pour notre cadre d'application et par conséquent de les évaluer. Or l'évaluation de ces méthodes d'extraction présente une difficulté majeure en ce qu'elle ne peut être réalisée de manière absolue, comme dans le cas d'autres tâches d'extraction d'information, à l'image de la détection d'opinion évaluée dans les chapitres 5 et 6.

La première question que cette évaluation soulève est celle de l'objet étudié. Il est en effet tout aussi envisageable d'évaluer chaque élément retrouvé lors de l'extraction, ou bien la ressource construite dans son ensemble. Cela nous amène à nous interroger sur le contexte de chaque élément étudié : comme le présentent [Nazarenko et al., 2009] dans le cas de l'extraction de termes, nous procédons soit à une évaluation d'un lexique en jugeant chacune de ses entrées indépendamment des phrases dans lesquelles il apparaît, soit à la mesure de la différence perçue dans l'application finale lors de l'emploi de la ressource.

7.2.1 Évaluation per se

L'évaluation intrinsèque des subjectivèmes consiste à en valider la valeur d'indicateur d'opinion sur la seule base de leur forme lexicale, ce qui peut donc être effectué lors de leur extraction ou en parcourant le lexique. Les cas limites de cette approche concernent bien entendu les marqueurs d'opinion les plus ambigus, pour lesquels il est nécessaire de projeter (mentalement ou réellement) l'élément sur un grand nombre de phrases afin d'en vérifier la pertinence, ce qui est coûteux en temps en sus de ne pas être nécessairement fiable. De manière plus générale, cette évaluation *per se* requiert un évaluateur disposant d'un certain degré d'affinité avec la langue, et éventuellement le domaine, de la ressource en construction. Cependant, il s'agit d'un moyen rapide (factuellement, quelques secondes tout au plus sont nécessaires à la validation d'un élément explicite) de filtrer une liste précise de marqueurs d'opinion que nous souhaitons considérer pour une application.

7.2.2 Évaluation in situ

À l'inverse, l'évaluation de la ressource au travers de celle de l'application finale – la détection d'opinion dans notre cas – permet de mesurer l'intérêt d'un lexique dans son ensemble, et ce de manière automatique. Dans ce cas, une connaissance étendue des notions de subjectivité et d'opinion dans le langage ne sont pas nécessaires, ni même la compréhension de la langue dans laquelle la détection d'opinion est réalisée. Toutefois, cette approche ne peut garantir l'évaluation de chacun des éléments du lexique, mais uniquement ceux présents dans le corpus employé. De surcroît, il n'est pas aisé par ce biais d'identifier au sein d'une nouvelle ressource les marqueurs conduisant à une meilleure détection de l'opinion et ceux la dégradant.

7.2.3 Évaluation choisie

La construction d'un lexique affectif est à notre sens le résultat d'un ajout incrémental de mots et d'expressions soigneusement choisis, c'est pourquoi une maîtrise de l'ensemble du contenu de cette ressource nous paraît tout aussi essentielle que l'évaluation de son apport réel lors de son utilisation.

Dans notre cadre de travail, nous jugeons les nouveaux ajouts au lexique affectif par une approche *per se*, en validant manuellement les candidats. Afin de tenir compte de la capacité des méthodes d'extraction de marqueurs d'opinion à proposer des éléments pertinents, nous évaluons ces éléments sur différentes précisions à N (dans les expériences suivantes, p@5, p@10, p@50 et p@100, lorsqu'il y a lieu d'ordonner les résultats). En parallèle de ce filtrage manuel, les ressources employées sont régulièrement testées pour la détection d'opinion sur un même jeu de corpus afin de garantir une non-régression des performances. Enfin, nous conservons la possibilité d'éditer manuellement le contenu du lexique affectif, notamment lorsque des erreurs ou des omissions sont signalées par les utilisateurs de l'application, qui en sont, d'une certaine façon, les véritables évaluateurs.

7.3 Expériences

L'étude que nous avons mené sur les paradigmes de construction ou d'extension de lexiques affectifs nous ont orienté vers des méthodes d'extraction reposant sur des corpus. Ce choix est justifié d'un côté par les effets de bord néfastes que nous avons constaté lors d'expériences employant des ressources linguistiques externes, c'est-à-dire la dérive sémantique des synonymes et des traductions, et de l'autre par la plus grande pertinence que présentent les éléments émergeant de corpus des domaines que nous traitons.

Nous menons dès lors plusieurs expériences visant à identifier des subjectivèmes dans des corpus d'avis clients, premièrement au moyen de structures linguistiques, puis par classification supervisée.

7.3.1 Cadre expérimental

Nous cherchons à extraire de nouveaux marqueurs d'opinion parmi quatre corpus en français relativement homogènes en taille et, autant qu'il est possible de le mesurer, homogènes en qualité d'écriture (phrases longues et construites, pas d'écriture abrégée, cependant orthographe très variable, ponctuation parfois non respectée). Chacun traite néanmoins d'un domaine très différent, comme indiqué dans le tableau 7.3.

Le lexique de subjectivèmes utilisé, lorsque c'est le cas, est le lexique affectif français que nous avons décrit en section 4.2.1, c'est-à-dire une liste de 916 marqueurs positifs et 1 220 négatifs.

Nom	# Documents	Description
« Banques »	2 043	Avis spontanés sur plusieurs services de banques en ligne
« Burgers »	5 217	Avis spontanés sur des restaurants de « hamburgers gastronomiques »
« Chaussures »	6 536	Commentaires suite à l'achat de chaussures sur deux sites d'e-commerce
« Parcs »	4 994	Commentaires suite à la visite de deux parcs d'attractions français

TABLEAU 7.3 – Description des corpus utilisés pour évaluer l'extraction de marqueurs d'opinion

7.3.2 Structures linguistiques

Nous entendons par structure linguistique une connaissance a priori du contexte des marqueurs recherchés. Les méthodes d'extraction présentées ici s'appuient sur une hypothèse de cohérence du discours, concrétisée soit par la proximité des indicateurs, soit par une structure caractéristique de l'expression d'une opinion.

Cooccurrences

La première expérience que nous présentons est inspirée des travaux de [Hu and Liu, 2004] et constitue avant tout une approche basique pouvant être réalisée sans lexique affectif existant. Les candidats subjectivèmes sont ici les mots apparaissant à moins de trois mots d'un terme dans une phrase, pondérés par l'importance de cette proximité le selon le score 7.1, où $occ(w)$ est le nombre d'occurrences du mot w et $occ(t, w)$ son nombre de cooccurrences avec un terme.

$$score(w) = \frac{log_2(occ(t, w))}{occ(w)} \tag{7.1}$$

L'évaluation en précision à 5, 10, 50 et 100 éléments pour chaque corpus est reportée dans le tableau 7.4. Dans un deuxième temps, nous ne considérons que les adjectifs extraits par cette méthode.

Corpus	p@5	p@10	p@50	p@100
Banques	20,00	10,00	22,00	16,00
Banques (adjectifs seulement)	60,00	50,00	40,00	36,00
Burger	0,00	0,00	6,00	3,00
Burger (adjectifs seulement)	80,00	60,00	38,00	33,00
Chaussures	60,00	40,00	20,00	15,00
Chaussures (adjectifs seulement)	100	60,00	38,00	35,00
Parcs	40,00	40,00	16,00	15,00
Parcs (adjectifs seulement)	60,00	50,00	40,00	39,00

TABLEAU 7.4 – Extraction de subjectivèmes par cooccurrence avec les termes du corpus

Nous vérifions quantitativement par cette expérience la supposition de [Hu and Liu, 2004], à savoir que les adjectifs sont les seuls éléments pertinents émergeant de ces cooccurrences. Le cas des commentaires post-achat du corpus Chaussures, dont la précision à 5 éléments est satisfaisante sans cette limitation n'est en réalité pas en contradiction avec ce constat puisque les 3 marqueurs validés sont des adjectifs.

Corpus	Banques	Burgers	Chaussures	Parcs
Top 5 Subjectivèmes	*incompréhensibles*	*réussie*	*nulle*	***hivernale***
	opérationnel	***confite***	*phénoménal*	*talentueux*
	rude	*croustillants*	*meilleures*	*étonnants*
	pointues	*passables*	*appréciés*	***certaines***
	horrible	*exellente*	*lamentable*	*décevantes*

TABLEAU 7.5 – Top 5 des adjectifs subjectivèmes extraits par corpus (erreurs en gras)

Cette première méthode peut donc être vue comme un moyen d'amorcer un lexique d'adjectifs subjectifs dont il reste à attribuer la polarité, ce qui peut être fait lors de la validation manuelle. Il est intéressant de noter que les marqueurs extraits avec les plus hauts scores, listés dans le tableau 7.5, sont par ailleurs relativement spécifiques au domaine du corpus.

Extraction par patrons

Nous réalisons une extraction de marqueurs d'opinion en nous inspirant de la méthode utilisée dans [Riloff et al., 2003]. Nous recherchons les segments de phrase correspondant à un des patrons de détection à notre disposition, en retirant toutefois la contrainte de la présence d'un marqueur connu.

Lorsqu'un segment est reconnu par une règle syntaxique, deux cas de figure sont donc possibles. Dans le cas où l'emplacement du marqueur d'opinion correspond bel et bien à un élément du lexique affectif, le segment de phrase représente une opinion, que nous désignons comme « connue ». Dans le cas contraire, le segment de phrase est désigné comme « segment candidat » et celui-ci est conservé en tant que potentielle expression d'opinion dans une base de connaissance à part.

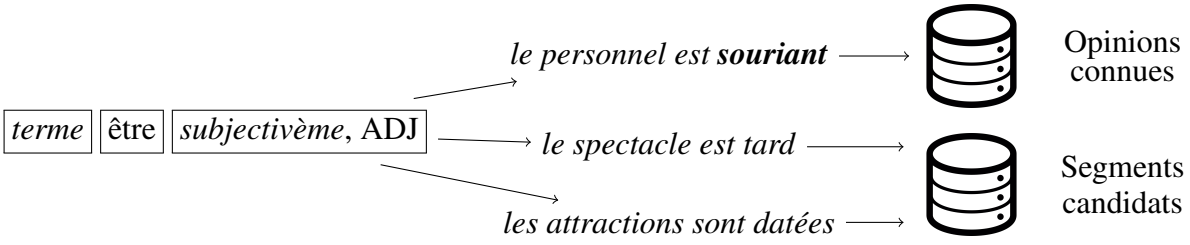


FIGURE 7.2 – Sélection d'expressions d'opinion et de segments candidats pour un même patron

Nous disposons donc à la suite de cette projection de patrons d'une liste d'opinions connues (par la suite O, pour opinions), et d'une liste de segments candidats linguistiquement cohérents (par la suite C, pour candidats) mais dont nous ignorons s'il s'agit ou non d'expressions d'opinions. Nous associons ensuite un score à chaque segment candidat, dérivé de celui défini pour l'algorithme BASILISK [Thelen and Riloff, 2002], représentant leur pertinence du point de vue de l'opinion. L'attribution du score se déroule selon les étapes suivantes :

1. Pour chacune des occurrences des opinions de O ou des segments candidats de C, nous retenons les lemmes des 2 mots précédents et suivants, dans la limite de la phrase.
2. Un premier score S_1 est attribué aux fenêtres de mots ainsi extraites, selon la formule (7.2).
3. Seules les fenêtres dont le score S_1 dépasse un certain seuil sont conservées (par la suite, F). Nous fixons empiriquement ce seuil au quart supérieur du classement.
4. Les segments candidats de C dont une occurrence au moins est entourée par une fenêtre de F sont associés à un second score S_2, selon la formule (7.3).

$$S_1(f) = \frac{o_f}{t_f} \times log_2(o_f) \qquad (7.2)$$

où :

- f est une fenêtre de mots
- o_f est le nombre d'occurrences de f autour d'une opinion connue
- t_f le nombre total d'occurrences de f

$$S_2(c) = \frac{\sum_{i \in N_c} log_2(1 + o_i)}{N_c} \times \frac{r_o}{r_{total}} \qquad (7.3)$$

où :

- c est un segment candidat
- N_c est le nombre de fenêtres contenant c
- o_i est le nombre d'occurrences de la i-ème fenêtre autour d'une opinion connue
- r_o est le nombre de fois où la règle ayant extrait c a extrait une opinion connue
- r_{total} est le nombre total de fois où la règle ayant extrait c a extrait un segment

La position du terme cible étant connue dans chaque patron, nous tronquons lors d'une dernière étape les segments candidats de manière à ne disposer que de la partie contenant le marqueur, éventuellement multi-mots (« *ma carte a été avalée par un distributeur* » est par exemple remplacé par « *a été avalée par un distributeur* »). Les marqueurs ainsi associés au score (S_2) sont évalués en précision à N éléments.

Corpus	p@5	p@10	p@50	p@100
Banques	60,00	80,00	64,00	60,00
Burger	40,00	50,00	54,00	43,00
Chaussures	80,00	70,00	70,00	70,00
Parcs	100	80,00	58,00	48,00

TABLEAU 7.6 – Extraction de subjectivèmes par identification de patrons lexico-syntaxiques

Les résultats de cette extraction, présentés dans le tableau 7.6, sont très encourageants dans la mesure où les 5 à 10 premiers marqueurs sont extraits avec précision, d'autant plus que pour cette expérience, ceux-ci ne se limitent pas aux adjectifs. En outre, nous observons qu'une grande partie des marqueurs pertinents sont des expressions multi-mots que nous avons précédemment considérées ambigües car composées d'éléments qui n'indiquent pas intrinsèquement une opinion, tel que « *pas résolu à ce jour* » (corpus Banques), « *fait dans l'urgence* » (corpus Burgers), « *resté en allemagne* » (corpus Chaussures) ou « *oublier le boulot* » (corpus Parcs).

Connecteurs logiques

La limite commune des deux précédentes méthodes d'extraction de marqueurs est le manque d'automatisation de l'inférence de polarité. Nous retenons de notre étude des paradigmes d'extraction qu'un moyen fiable de réaliser cette inférence par des structures linguistiques est de propager la polarité de subjectivèmes connus à des éléments dont ils partagent une conjonction, tel que présenté dans le travail de [Hatzivassiloglou and McKeown, 1997]. Nous proposons d'appliquer ce principe et d'étendre les règles lexico-syntaxiques capturant de nouveaux marqueurs d'opinion.

En sus des conjonctions *et* et *mais* utilisées dans [Hatzivassiloglou and McKeown, 1997], nous considérons la virgule comme unité liant deux adjectifs de même polarité et nous permettons la présence d'un adverbe au sein de chaque structure. Lorsque cet adverbe est un marqueur de négation, l'adjectif juxtaposé est préfixé par une notation « non- », afin d'en désambiguïser la polarité. Les règles employées sont les suivantes :

- x et y (« *rapide et pratique* »)
- x et ADV y (« *rapide et vraiment pratique* »)
- non-x et non-y (« *pas cher et pas mauvais* »)
- x mais non-y (« *cher mais pas mauvais* »)
- non-x mais y (« *pas cher mais mauvais* »)

} x et y partagent la même polarité

- x et non-y (« *cher et pas pratique* »)
- non-x mais non-y (« *pas cher mais pas pratique* »)
- non-x et y (« *pas rapide et cher* »)
- x mais y (« *rapide mais cher* »)
- x mais ADV y (« *rapide mais vraiment cher* »)

} x et y ont une polarité inverse

Contrairement aux expériences précédentes il n'y a ici aucun ordre parmi les candidats extraits, c'est pourquoi nous en évaluons l'ensemble. Nous validons tout d'abord tous ceux retrouvés pour juger de leur subjectivité, puis en distinguant les éléments de chaque polarité lorsqu'une valeur a pu leur être propagée.

	Tous marqueurs		Positifs		Négatifs	
Corpus	#Élements	Précision	#Élements	Précision	#Élements	Précision
Banques	176	17,04	47	21,27	36	38,88
Burger	514	31,12	171	38,01	45	62,22
Chaussures	132	32,57	72	23,61	5	60,00
Parcs	425	27,76	151	26,49	47	31,91

TABLEAU 7.7 – Extraction de subjectivèmes par relations de conjonction

L'ensemble de ces résultats, indiqués dans le tableau 7.7, est assez moyen pour les quatre corpus, que ce soit du point de vue de la subjectivité ou de la polarité. Seuls les marqueurs négatifs extraits (en particulier dans les corpus Burgers et Chaussures) présentent une précision satisfaisante.

Cet échec s'explique, d'après nos observations, par trois phénomènes. Premièrement, le fait de restreindre les marqueurs recherchés aux unités monolexicales rend l'évaluation ambigüe. Ainsi, les adjectifs associés à une polarité qui ne semble pas explicite pour le domaine concerné ont été écartés des résultats pertinents. Deuxièmement, certains de ces éléments portant justement une polarité ambigüe ou neutre peuvent être en conjonction avec des subjectivèmes connus des deux polarités, et propager en conséquence une valeur sans en être de réels indicateurs. Enfin, les erreurs d'étiquetage grammatical font que les mots retenus ne sont pas toujours des adjectifs et parfois même des unités sans orientation sémantique propre (nombres, sigles ou abréviations).

7.3.3 Classification supervisée

Nous complétons nos expériences en construction de lexique affectif par deux méthodes d'apprentissage automatique. Nous cherchons à retrouver des mots partageant les mêmes voisins que des marqueurs d'opinion de référence, à l'aide d'un modèle de vecteurs de contextes premièrement, puis d'un étiquetage de séquence.

Extraction par similarité de contexte

Il est démontré que la représentation des mots en vecteurs de contextes permet de les regrouper efficacement en classes sémantiquement cohérentes [Bengio et al., 2003]. Récemment, la diffusion d'outils comme WORD2VEC [Mikolov et al., 2013] sur lesquels des propriétés supplémentaires de ces vecteurs ont été montrées, telle que la composition, ont popularisé cette méthode. Par cette expérience, nous souhaitons voir dans quelle mesure ce regroupement s'applique aux classes de mots subjectifs, positifs ou négatifs.

Nous employons pour cela l'implémentation Java de WORD2VEC disponible dans la librairie DEEPLEARNING4J[12]. Nous fixons empiriquement les paramètres à une fréquence minimum de 5 pour les mots considérés, une fenêtre de 3 mots autour des marqueurs candidats et une taille de 100 éléments pour les vecteurs construits. Ces vecteurs de contextes sont calculés à partir des phrases segmentées en mots lemmatisés par notre module de prétraitement propriétaire.

Après avoir créé un modèle de vecteurs pour chaque corpus, nous les interrogeons à trois reprises. Afin de recueillir des marqueurs sans tenir compte de leur polarité, nous utilisons comme mots de référence l'ensemble des entrées du lexique affectif. Nous limitons ensuite ces références aux subjectivèmes positifs, puis enfin aux négatifs, dans le but d'extraire dans chaque cas des éléments de la polarité correspondante.

Corpus	Subjectivèmes de référence	p@5	p@10	p@50	p@100
Banques	Tous	20,00	10,00	18,00	12,00
	Positifs	0,00	0,00	10,00	8,00
	Négatifs	0,00	0,00	2,00	4,00
Burger	Tous	0,00	0,00	24,00	19,00
	Positifs	0,00	20,00	18,00	13,00
	Négatifs	20,00	10,00	8,00	8,00
Chaussures	Tous	0,00	0,00	8,00	8,00
	Positifs	40,00	40,00	22,00	20,00
	Négatifs	0,00	0,00	4,00	4,00
Parcs	Tous	20,00	20,00	22,00	17,00
	Positifs	0,00	0,00	18,00	14,00
	Négatifs	0,00	20,00	20,00	13,00

TABLEAU 7.8 – Extraction de subjectivèmes par similarité de vecteurs de contextes

Pour cette expérience, nous n'avons pas limité le champ de recherche à certaines étiquettes grammaticales, ce qui est en partie la cause des basses précisions que nous pouvons observer dans le tableau 7.8. Les mots présentés comme similaires en contexte par l'algorithme sont en effet rarement des indicateurs de l'opinion, et lorsque cela est le cas, l'inférence de polarité n'est pas systématiquement juste. Nous attribuons ce phénomène à la petite taille de nos corpus, où l'application de tels calculs gourmands en données n'est manifestement pas adaptée. Les résultats de l'extraction d'éléments positifs dans le corpus Chaussures, satisfaisants en comparaison des autres expériences, peut s'expliquer par une grande majorité d'avis positifs dans ce corpus ; l'extraction bénéficie donc dans ce cas d'un biais de fréquence.

[12] deeplearning4j.org/word2vec

Extraction par étiquetage de séquence

La dernière méthode méthode que nous souhaitons tester est un étiquetage de séquence des subjectivèmes à l'aide d'un modèle CRF. Tout comme l'expérience précédente, cette classification est peu supervisée dans la mesure où les exemples positifs ne sont pas annotés manuellement mais indiqués par projection du lexique affectif existant. De cette manière, les phrases fournies pour l'apprentissage du modèle sont en tout point similaires à celles de la méthode CRF+S illustrées sur la figure 6.3, en section 6.3.3. Afin de disposer d'un corpus d'entraînement de taille raisonnable, mais également d'un corpus d'évaluation suffisamment grand pour faire apparaître des éléments candidats, nous scindons chaque corpus en deux parties de taille égale.

Corpus	#Éléments extraits	Précision (Subjectivité)	Précision (Positifs)	Précision (Négatifs)
Banques	25	72,00	52,94	88,88
Burger	44	86,36	77,41	92,30
Chaussures	42	69,04	63,88	83,33
Parcs	15	80,00	81,81	50,00

TABLEAU 7.9 – Extraction de subjectivèmes par étiquetage de séquence

L'évaluation de cet étiquetage, dont les résultats sont présentés dans le tableau 7.9, ne tient pas compte de l'ordre des éléments trouvés dans la mesure où ceux-ci sont extraits de façon indépendante d'une part, et sont peu nombreux d'autre part. La quantité modeste de ces marqueurs ne peut toutefois pas être tenue comme un échec de cette expérience, étant donné la précision importante de l'extraction. Par ailleurs, un avantage majeur de cette méthode, partagé avec celle des patrons de détection, est la possibilité d'identifier des subjectivèmes polylexicaux ou non triviaux, tel que « *ne répond pas* » (corpus Banques), « *victime de son succès* » (corpus Burgers) , « *taillent bien* » (corpus Chaussures) ou « *surpris à rêver* » (corpus Parcs).

7.4 Synthèse

Dans ce chapitre, nous nous sommes appuyés sur des contributions existantes en extraction de marqueurs d'opinion afin de définir une ou plusieurs méthodes adaptées aux contraintes qui délimitent le cadre de notre recherche. Nous notons deux familles de contributions :

- L'identification de marqueurs d'opinion au sein de **corpus homogènes à ceux analysés**, reposant sur des méthodes symboliques ou probabilistes ;
- L'utilisation de **ressources linguistiques externes**, allant d'un dictionnaire de synonymes à un moteur de traduction, pour étendre ou adapter un lexique existant.

Nos expériences préliminaires et les conclusions de la littérature nous ont orientés vers les méthodes exploitant le contexte des éléments recherchés, c'est-à-dire leur extraction depuis un corpus. Nous avons alors réalisé tour à tour l'extraction de marqueurs en cooccurrence avec des termes cibles, en conjonction avec des marqueurs connus, présents dans des patrons lexico-syntaxiques caractéristiques de l'opinion, et présentant un vecteur de contexte ou des séquences de mots voisins statistiquement similaires à des marqueurs connus. Nous retenons de ces expériences plusieurs points très positifs :

- Une extraction relativement naïve des **adjectifs proches de termes** d'un corpus permet d'acquérir un premier lexique affectif pertinent et spécifique du domaine ;
- Un lexique tel que celui-ci peut être enrichi efficacement au moyen de **patrons de détection** ou d'un **étiquetage de séquence** ;
- Par ailleurs, ces deux méthodes permettent d'extraire des **marqueurs implicites et multi-mots**, ce qui est essentiel pour la complétion de notre ressource.

8

Extraction de patrons

Dans le chapitre précédent, nous avons cherché à constituer par diverses méthodes un lexique affectif. Nous avons constaté qu'en dehors des adjectifs, les subjectivèmes les plus informatifs sont les expressions polylexicales. Celles-ci offrent suffisamment de contexte pour être d'explicites marqueurs d'opinion et sont bien souvent spécifiques au domaine du corpus duquel elles sont extraites. Nous avons toutefois observé que l'extraction automatique de ces marqueurs, bien que précise, n'enrichit pas le lexique de nombreuses entrées. De plus, celles-ci sont relativement peu couvrantes du fait que les n-grammes, au delà d'une certaine valeur de n, tendent vers une condition d'hapax, c'est-à-dire des éléments apparaissant une seule fois dans un corpus.

Une manière d'augmenter la couverture des expressions extraites est de les généraliser, en considérant une abstraction de certains des mots qui les composent, autrement dit les extraire en tant que patrons lexico-syntaxiques. Ainsi, l'identification automatique de nouveaux patrons de détection, qui est le sujet de ce chapitre, n'est pas seulement un moyen d'étendre les types de liens entre un terme et un subjectivème connus, mais prolonge notre travail de construction de lexique affectif.

La frontière entre marqueur et patron, qui jusqu'ici se dessinait distinctement, nous apparaît dès lors comme plus diffuse, et ne semble tenir qu'au choix d'un type de ressource. Un patron est en effet considéré comme une entrée du lexique s'il est totalement lexicalisé et n'est ajouté en tant que chemin de l'automate de reconnaissance que dans le cas où un de ses éléments est d'un niveau plus abstrait, telle qu'une étiquette grammaticale. Néanmoins, ce choix n'est pas sans conséquences, car du point de vue des performances de l'application en matière de temps de calcul, il est largement préférable d'accumuler des expressions dans le lexique affectif, quand bien même il s'agit d'hapax, tandis que l'ajout de chemins dans l'automate est synonyme d'une meilleure couverture des opinions dans un corpus.

Dans ce chapitre, nous cherchons à définir une méthode d'extraction de patrons présentant un compromis satisfaisant entre la lexicalisation et la généralisation. Nous montrons tout d'abord comment une identification automatique de nouveaux patrons lexico-syntaxiques peut être réalisée en combinant deux idées issues de la littérature [Riloff and Wiebe, 2003, Murray and Carenini, 2011]. Nous évaluons ensuite cette méthode pour l'extension d'une ressource existante sur des corpus d'avis en français et en anglais, et pour la création d'un nouvel ensemble de patrons à partir d'un corpus d'avis en chinois.

8.1 Extraction de patrons lexico-syntaxiques

L'extraction de patrons lexico-syntaxiques induit deux difficultés principales. Il s'agit d'une part de reconnaître des empans de texte linguistiquement cohérents et pertinents pour la détection d'opinion, et d'autre part de sélectionner pour chaque mot d'un patron le niveau d'abstraction le plus adapté.

8.1.1 Extraction par amorçage

Nous revenons ici sur le fonctionnement de la méthode d'extraction de patrons proposée par [Riloff and Wiebe, 2003], que nous avons d'ores et déjà mentionnée à plusieurs reprises dans ce travail. Cette méthode repose essentiellement sur le rapport de fréquence de structures préalablement définies au sein de contextes pertinents (soit dans notre cas, contenant une opinion) ou non. Dans le travail de [Riloff and Wiebe, 2003], ces structures sont formées par l'association d'une à deux étiquettes grammaticales et de l'emplacement du sujet cible d'opinion ; dans la version que nous en avons proposé au chapitre précédent, celles-ci sont les patrons de détection à trois niveaux d'abstraction contenant jusqu'à une dizaine d'éléments. Ce fonctionnement limite donc par définition l'acquisition de nouveaux patrons, en la conditionnant par des structures définies manuellement et sur la base d'une connaissance a priori des expressions d'opinion.

Pour notre méthode d'identification de patrons, nous proposons d'étendre le champ des extractions potentielles en ne supposant pas de telles structures syntaxiques. Nous considérons tous les segments de texte reliant un terme et un subjectivème connu, et nous leur attribuons un score de pertinence dont les facteurs sont leur spécificité à l'expression d'une opinion et leur niveau d'abstraction.

8.1.2 Patrons à instanciation variable

La seconde idée sur laquelle nous nous appuyons pour définir notre méthode d'extraction de patrons est la génération de patrons à partir de segments de texte de longueur définie.

Principe

Afin d'identifier des marqueurs d'opinion multi-mots dans des articles de blog, sujets à des approximations orthographiques substantielles, [Murray and Carenini, 2011] proposent d'appliquer les scores de pertinence de [Riloff and Wiebe, 2003] non pas sur des patrons définis, mais sur des n-grammes lexico-syntaxiques générés automatiquement à partir de segments de texte, appelés « n-grammes à instanciation variable ». Comme illustré sur la figure 8.1, un patron est généré pour chaque combinaison d'éléments sous leur forme lexicale ou syntaxique.

Segment extrait	*Really great idea*		
Patrons générés	*really*	*great*	*idea*
	really	*great*	NN
	really	JJ	*idea*
	RB	*great*	*idea*
	really	JJ	NN
	RB	*great*	NN
	RB	JJ	*idea*
	RB	JJ	NN

FIGURE 8.1 – Exemple donné dans [Murray and Carenini, 2011] de « patrons à instanciation variable » générés automatiquement

Dans le travail de [Murray and Carenini, 2011], les segments considérés sont limités aux trigrammes, ce qui ne peut être appliqué aux types de patrons que nous recherchons. En effet nous souhaitons que les patrons extraits contiennent *a minima* un terme et un marqueur d'opinion connu, ce qui limite drastiquement les possibilités de segments extraits.

Application

Nous procédons à plusieurs adaptations de cette méthode afin de l'appliquer aux segments de texte longs reliant un terme et un subjectivème. Premièrement, les unités lexicales des segments extraits dans notre cas sont éventuellement complexes. Il nous paraît en effet incohérent de différencier sur un plan lexical et syntaxique l'utilisation de chacun des constituants de polylexicaux, tels que les locutions prépositives (« *à partir de* », « *en cas de* ») ou adverbiales (« *d'une façon générale* », « *en fait* »). Deuxièmement, nous exploitons le niveau d'abstraction supplémentaire décrit au chapitre 4, en indiquant l'emplacement des termes et subjectivèmes bien entendu, mais également les négations au sein des patrons. Ainsi, en sus de leur forme lexicale, les unités concernées, telles que « *aucun* » ou « *pas du tout* », ne sont pas représentées par leur catégorie grammaticale mais par une étiquette « négation », que nous jugeons plus informative.

Combinatoire

Enfin, précisons que cette méthode d'extraction peut être fortement ralentie par la combinatoire des éléments lexicaux et syntaxiques des patrons. Pour chaque n-gramme extrait, l'algorithme génère en effet 2^n patrons, que nous devons conserver afin de calculer leur score de pertinence. Dans la mesure où cette extraction est réalisée de manière parallèle à la détection d'opinion, cette lenteur n'est pas rédhibitoire, cependant dans le cadre de nos expériences nous fixons une nombre maximum de 11 unités lexicales pour un segment compris dans les limites d'une phrase, ce qui peut correspondre parfois à son entièreté.

8.1.3 Extraction itérative de patrons

Le processus que nous proposons se déroule en deux étapes, soit une identification de fenêtres de mots caractéristiques de l'opinion par amorçage sur les patrons de détection d'opinion connus, puis une sélection des patrons candidats apparaissant dans ces fenêtres. Ces opérations sont en suite réitérées en considérant les nouveaux patrons comme amorces.

Amorçage

La première étape de cette méthode est en tout point similaire à celle que nous avons présentées en section 7.3.2 pour l'extraction de subjectivèmes polylexicaux. Des fenêtres de mots, jusqu'à 3 éléments de part et d'autre de chaque opinion reconnue à l'aide des patrons dont nous disposons, sont associées à un score (formule 7.2) représentant leur spécificité à la présence d'une opinion.

Sélection de patrons candidats

Les segments de texte de 3 à 11 unités lexicales compris dans une fenêtre retenue sont extraits. À partir de ces segments, nous générons les patrons à instanciation variable correspondants, c'est-à-dire toutes les combinaisons de n-grammes où un élément peut être instancié par sa forme lexicale ou sa forme abstraite (son étiquette grammaticale ou son rôle dans l'opinion). Les patrons sont ensuite ordonnés par un score (S_3) dont les facteurs dépendent de leur fréquence au sein de fenêtres caractéristiques de l'opinion, de leur capacité à extraire de nouvelles expressions et de leur niveau d'abstraction.

$$S_3(p) = \frac{\#p_{op}}{\#p} \times log_2(\#p_{op}) \times E_p \times \frac{p_{pos}}{|p|} \tag{8.1}$$

où :

- p est un patron candidat
- $\#p_{op}$ est le nombre d'occurrences de p contenues dans des fenêtres validées
- $\#p$ est le nombre total d'occurrences de p
- E_p est le nombre d'expressions uniques que couvre p
- p_{pos} est le nombre d'étiquettes grammaticales de p
- $|p|$ est la longueur du patron p en unités lexicales.

Itérations

Nous sélectionnons les patrons ayant obtenu le meilleur score S_3, en fixant empiriquement une limite de 10 patrons retenus, afin de les ajouter aux patrons de référence. Il est possible de réitérer ce cycle d'extraction jusqu'à convergence, c'est-à-dire qu'aucune nouvelle fenêtre pertinente n'est retenue, ou que les patrons candidats ne permettent pas d'extraire de nouvelles expressions. Dans les expériences suivantes, cette convergence est considérée atteinte lorsque l'évaluation d'un nouveau cycle est identique au précédent.

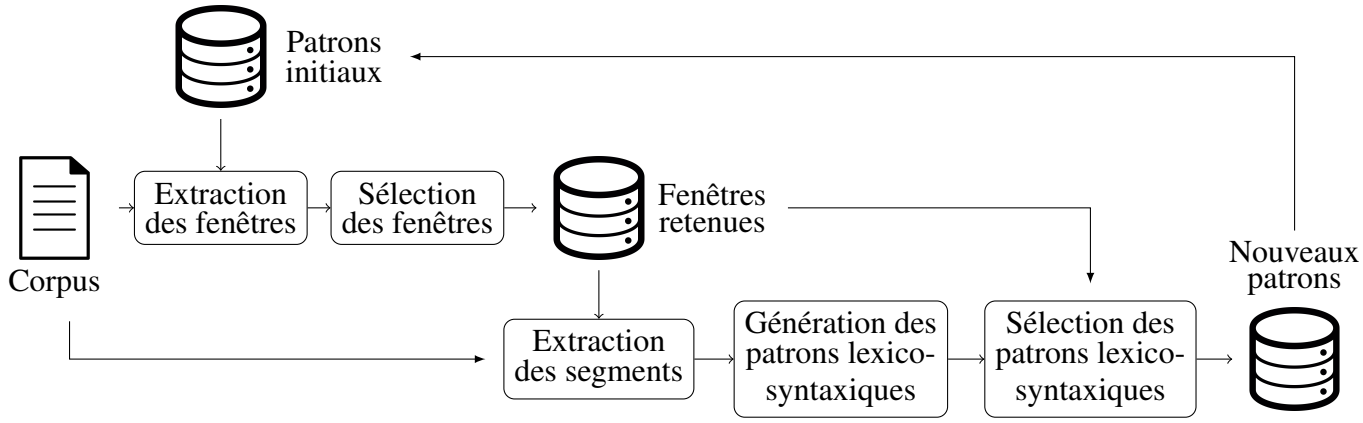


FIGURE 8.2 – Schéma de l'extraction de patrons de détection

8.2 Expériences

Nous évaluons notre méthode itérative d'extraction de patrons en l'appliquant à deux situations. Dans la première, la ressource de référence (« Patrons initiaux » sur la figure 8.2) est une liste de patrons relativement fournie et dont les éléments ont été corrigés ou optimisés au cours du temps. Dans la seconde, nous expérimentons cette méthode pour la création d'une nouvelle ressource, soit une génération d'une liste de patrons qui ne nécessite, pour les besoins de l'algorithme, qu'un unique patron défini manuellement.

8.2.1 Acquisition de patrons à partir d'une ressource existante

Nous souhaitons observer dans quelle mesure l'extraction automatique de patrons lexico-syntaxiques peut enrichir les règles de détection d'opinion en français et en anglais que nous avons manuellement construites et régulièrement mises à jour.

Cadre expérimental

Afin de comparer les performances obtenues avec les résultats précédents, nous menons ces expériences sur les corpus annotés en cibles d'opinion issus de l'atelier SEMEVAL 2016, sur lesquels nous nous sommes également appuyés aux chapitres 5 et 6. L'extraction de patrons est appliquée au corpus d'entraînement (335 documents dans le corpus en français, 350 dans le corpus en anglais), puis nous comparons sur le corpus d'évaluation les résultats de la détection d'opinion reposant sur les patrons initiaux dans un premier temps, et la ressource enrichie dans un second temps.

Résultats

S'il est vrai que la liste de patrons dans les deux langues a bel et bien été enrichie (plusieurs itérations sont effectivement nécessaires avant d'atteindre une convergence), nous constatons cependant que la performance globale des systèmes a peu changé, comme il est indiqué sur les tableaux 8.1 et 8.2 (-0,37pp F1 en français, +0,13pp F1 en anglais). Les résultats pour chacune des langues illustrent en réalité deux phénomènes différents. Dans le cas du français, de nouveaux patrons identifiés permettent une amélioration légère de la couverture des opinions (+0,14pp en rappel), au prix d'un plus grand nombre d'expressions détectées qui ne sont pas des opinions (-1,66pp en précision).

Ressource	Précision	Rappel	F1
Patrons initiaux	**44,13**	26,15	**32,84**
Patrons extraits (9 itérations jusqu'à convergence)	42,47	**26,29**	32,47

TABLEAU 8.1 – Extraction de patrons de détection en français

Ressource	Précision	Rappel	F1
Patrons initiaux	54,33	**34,50**	42,20
Patrons extraits (8 itérations jusqu'à convergence)	**60,11**	32,67	**42,33**

TABLEAU 8.2 – Extraction de patrons de détection en anglais

Les patrons extraits depuis le corpus en anglais sont plus précis – car davantage lexicalisés – lorsqu'ils correspondent à une expression d'opinion (+5,78pp en précision), mais certains chevauchent des segments de texte couverts par les patrons initiaux et entravent leur activation, ce qui induit une perte de couverture (-1,83pp en rappel).

8.2.2 Construction d'une nouvelle ressource

Les résultats précédents tendent à montrer que cette identification automatique de nouveaux patrons de détection peine à améliorer une ressource déjà établie. Nous conduisons à présent une seconde extraction afin d'évaluer la capacité de cette méthode à créer une liste de patrons dans une langue où aucune ressource n'existe, dans notre cas le chinois.

Corpus

Cette expérience nécessite deux corpus annotés en cibles d'opinion en chinois. Nous utilisons le corpus d'avis clients sur des appareils électroniques issu de l'atelier SEMEVAL ABSA 2016 afin d'évaluer la méthode d'identification de patrons, et le corpus MioChnCorp[1] [Yiou et al., 2015], constitué d'avis clients sur des hôtels, pour comparer cette détection d'opinion *ex nihilo* avec l'adaptation d'un modèle d'extraction probabiliste existant. En effet il ne s'agit pas tant de voir ici le gain qu'apporte la ressource créée par rapport à un seul patron de détection, mais aussi l'intérêt de cette méthode face à une approche plus commune, dans le cas où nous disposerions d'un premier modèle entraîné sur des données d'un domaine différent.

Lexique affectif

Les subjectivèmes chinois que nous employons sont issus d'une traduction d'un sous-ensemble de subjectivèmes en français par un locuteur natif. Par ailleurs, nous enrichissons ce lexique à l'aide de la méthode d'extraction par patrons présentée au chapitre 7. Ces marqueurs sont, comme en français ou en anglais, généralement des adjectifs. Cependant la langue chinoise distingue deux types d'adjectifs : les qualificatifs « simples », qui ont le même statut que nos adjectifs attributs ou épithètes, et les verbes statifs qui correspondent davantage à l'usage d'un participe passé pour qualifier un sujet. De ce fait les verbes statifs sont rarement neutres et forment par conséquent une classe de mots *de facto* subjectivèmes.

Marqueur extrait	Traduction	Occurrences
方便	*pratique*	15
快	*rapide*	8
实用	*fonctionnel*	7
出色	*incroyable*	6

TABLEAU 8.3 – Exemple de marqueurs d'opinion extraits du corpus en chinois

Patron d'amorçage

Nous définissons un seul patron d'amorçage, terme subjectivème , soit l'association la plus simple formant une opinion. Ce patron ne permet d'extraire qu'un nombre restreint d'expressions d'opinion, dont nous présentons quelques exemples dans le tableau 8.4, mais avec une précision satisfaisante, comme indiqué dans le tableau 8.5.

Expression extraite	Traduction	Expression extraite	Traduction
出片效果色彩深沉	*le filtre de film est sombre*	操作简单方便	*le fonctionnement est simple*
反差高	*le contraste est fort*	焦准确	*le focus est précis*
性能良好	*la performance est bonne*	降噪功能明显	*la réduction de bruit est significative*

TABLEAU 8.4 – Exemples d'opinions extraites du corpus par le patron d'amorçage terme subjectivème

[1] pan.baidu.com/s/1dDo9s8h

Résultats

Le traçage des résultats de chaque itération, illustré sur la figure 8.3, valide notre approche : la précision initiale est non seulement conservée mais est même améliorée (+2,76pp), tandis que le rappel croît de façon continue jusqu'à une stabilisation (+40,69pp). Le « saut » que nous observons sur la courbe à l'itération 5 est dû à l'identification d'un patron couvrant de nombreuses expressions, [*terme*] [ADV] [*subjectivème*], telle que 成像质量很好 (*la qualité d'image est très bonne*).

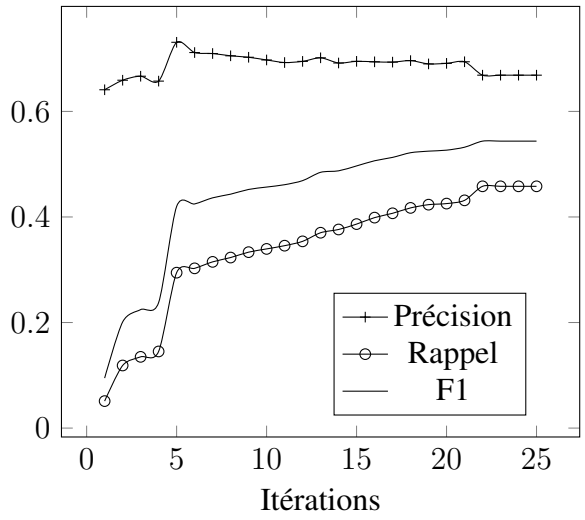


FIGURE 8.3 – Graphique de l'évaluation pas-à-pas de l'extraction itérative de patrons

Au total la ressource produite compte 72 patrons, dont nous présentons quelques exemples dans le tableau 8.6. La comparaison des deux systèmes évalués, indiquée dans le tableau 8.5, montre que notre méthode de construction de ressource par amorçage est globalement plus efficace pour la détection d'opinion dans ce corpus d'un nouveau domaine. Le modèle d'étiquetage de séquence entraîné sur les avis du domaine de l'hôtellerie, certes plus précis (+3,62pp en précision), est en effet significativement moins couvrant (-21,12pp en rappel).

Ressource	Précision	Rappel	F1
Patrons initiaux	64,10	5,11	9,46
Patrons extraits	66,86	**45,80**	**54,36**
CRF	**70,48**	24,68	36,56

TABLEAU 8.5 – Évaluation du modèle CRF (entraîné sur un domaine différent) et de l'extraction de patrons par amorce pour la détection d'opinion parmi des avis sur des appareils électroniques en chinois.

Ces résultats prometteurs laissent entrevoir des opportunités intéressantes en matière d'échange entre les approches symbolique et probabiliste. En effet s'il était possible d'appuyer un modèle de classification sur des premières annotations automatiquement créées à l'aide des patrons, cela pourrait en accélérer la mise en place. Nous procédons à la validation d'une telle supposition dans le chapitre suivant.

Patron	Expression	Traduction
subjectivème *terme*	多功能设置	*les paramètres sont multifonctions*
terme 也 ADV *subjectivème*	菜单界面也很友好	*l'interface est aussi très agréable*
subjectivème 的 *terme*	出色的机身设计	*excellent design du produit*
terme CS 不 *subjectivème*	对焦系统虽然不够	*le système de focus n'est pas suffisant*
terme 不 VERB ADV *subjectivème*	整机重量不是很重	*le poids n'est globalement pas un problème*

TABLEAU 8.6 – Exemples de patrons extraits et d'expressions correspondantes

8.3 Synthèse

Dans ce chapitre, nous avons présenté une méthode d'extraction de patrons de détection en nous inspirant de deux concepts existants.

- L'**extraction de marqueurs polylexicaux** au sein de fenêtres caractéristiques d'expressions subjectives, telle que proposée par [Riloff and Wiebe, 2003] ;
- Les **patrons à instanciation variable**, définis par [Murray and Carenini, 2011], à travers lesquels nous généralisons des marqueurs extraits en patrons.

À l'aide de cette méthode nous avons cherché à compléter des bases de patrons de détection existantes, puis à en créer une nouvelle pour une langue encore non traitée.

- Nous avons conclu que l'**enrichissement des ressources établies** ne permet pas d'améliorer significativement la détection d'opinion, voire en gêne le fonctionnement ;
- En revanche nous obtenons par cette méthode des résultats très positifs pour la **création d'une ressource**. Les patrons ainsi construits se révèlent notamment plus performants pour la détection d'opinion qu'un modèle d'étiquetage entraîné sur un domaine différent.

9

Annotation de corpus assistée

Il ressort des expériences précédentes que la constitution d'une ressource symbolique semble plus adaptée à l'initialisation d'un système de fouille d'opinion qu'à son développement sur le long cours. Afin d'améliorer continuellement ce système, il est donc nécessaire d'emprunter une approche différente. Sur la base de notre étude des systèmes hybrides décrits au chapitre 6, nous proposons de faire reposer ce système sur une méthode probabiliste supervisée.

Dans ce chapitre, nous expérimentons plusieurs moyens de réduire l'effort d'annotation intrinsèque à la mise en place d'une telle méthode. Nous proposons tout d'abord une solution à l'inconvénient que représente le « démarrage à froid » de la création de matériel annoté en nous appuyant sur une ressource symbolique, puis nous combinons cette amorce à une annotation itérative. Nous étudions enfin dans quelle mesure l'annotation peut être présentée de manière plus efficace et éventuellement ludifiée dans notre cadre de travail afin de pallier la fastidiosité de la tâche.

9.1 Annotation pour la fouille d'opinion

Nous argumentons ici brièvement pour la création de matériel annoté, puis nous abordons la notion existante d'apprentissage actif, dont l'objectif est de réduire l'effort d'annotation et sur laquelle nous nous appuyons pour les expériences présentées par la suite.

9.1.1 Motivations

Nous avons vu lors de la comparaison des méthodes au chapitre 6, qu'un modèle d'étiquetage de séquence offre globalement une meilleure détection de l'opinion à granularité fine qu'un système reposant sur un ensemble de règles. En ce sens, il paraît inévitable de créer un matériel d'entraînement, c'est-à-dire des corpus annotés. Par ailleurs l'intérêt de ce type de ressource ne se limite pas à une question de performance. Comme nous l'avons évoqué au chapitre 5, la maintenance d'exemples de phrases annotées est plus aisée que celle d'une liste de patrons car ces exemples sont explicites et ne constituent pas, à l'inverse des règles symboliques, un risque de dérive de détection. Cette transparence des éléments de la ressource nécessaire au modèle probabiliste est également synonyme d'un accès plus ouvert à la contribution, tandis que l'ajout manuel de patron est un travail d'expert.

Malgré ces avantages, que nous avons pu développer dans les chapitres précédents, l'annotation de corpus en cibles d'opinion reste une tâche fastidieuse, et de manière plus concrète, trop chronophage pour l'appliquer dans notre cadre de travail. Par conséquent, ce que nous recherchons s'apparente davantage à une aide à l'annotation, ou une « annotation assistée par ordinateur », pour reprendre une expression usitée dans le domaine du dessin et de la conception technique.

9.1.2 Travaux existants

La notion existante se rapprochant le plus de celle d'une « annotation assistée par ordinateur » est l'apprentissage actif. Cette appellation recouvre un groupe de méthodes visant à réduire l'ensemble d'entraînement d'un modèle probabiliste, en sélectionnant les exemples les plus informatifs pour l'annotation, comme illustré sur la figure 9.1.

La sélection peut se dérouler selon plusieurs stratégies différentes, comme l'échantillonnage incertain (*uncertainty sampling* [Lewis and Gale, 1994]) qui consiste à présenter à l'annotateur les exemples pour lesquels le modèle est le moins confiant, ou encore la requête par votes (*query-by-committee* [Seung et al., 1992]), où plusieurs modèles apprennent en concurrence à partir de traits de classification différents. Les exemples sur lesquels le « comité » est le plus en désaccord sont proposés à l'annotation. Il existe d'autres stratégies que nous ne détaillons pas ici, cependant il est admis que l'échantillonnage incertain, que nous avons utilisé, est la plus efficace [Settles, 2010]. En ce qui concerne l'apprentissage actif pour l'étiquetage de séquence, [Tomanek and Hahn, 2009] proposent de réduire davantage l'effort manuel nécessaire en ne présentant à l'annotateur que les segments de séquence incertains. La mesure de cet effort manuel est par

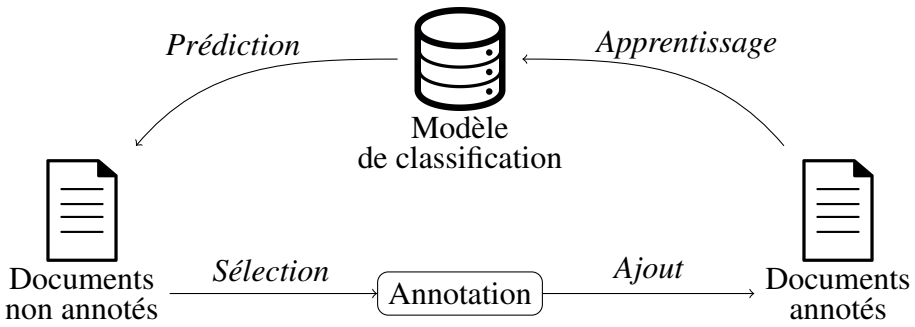


FIGURE 9.1 – Schéma global du principe d'apprentissage actif

ailleurs un sujet d'étude en tant que tel, car il n'est pas évident d'évaluer le coût d'annotation d'un exemple proposé par la méthode d'apprentissage actif. [Settles et al., 2008] montrent ainsi que ce coût peut être extrêmement variable selon l'exemple ou l'annotateur, et proposent une méthode d'estimation de l'effort nécessaire afin de réduire le coût total d'une annotation. [Tomanek, 2010] compare les temps totaux d'une tâche d'annotation suivant une stratégie d'apprentissage actif et une sélection aléatoire et constate que les éléments proposés aléatoirement sont globalement moins coûteux en temps d'annotation, mais que la sélection active des exemples reste bénéfique de par la réduction du nombre d'instances nécessaires pour obtenir les mêmes résultats. Enfin [Joshi et al., 2014] proposent une mesure de « complexité d'annotation de sentiment » (« *sentiment annotation complexity* », ou *SAC*) reposant sur le temps qu'un annotateur passe à regarder chacun des mots d'une phrase, enregistré à l'aide d'un appareil d'oculométrie.

9.1.3 Phases de l'annotation

L'amélioration du modèle supervisé n'est pas linéaire lors de l'ajout de contenu annoté, mais suit une progression pseudo-logarithmique. Dans ce travail nous distinguons par conséquent deux phases dans cette progression : une phase d'*initialisation*, où le modèle est relativement instable et peut être fortement influencé par de nouvelles annotations, et une phase d'*amélioration continue*, où le modèle est viable pour la détection d'opinion pour mais peut toujours bénéficier de contributions.

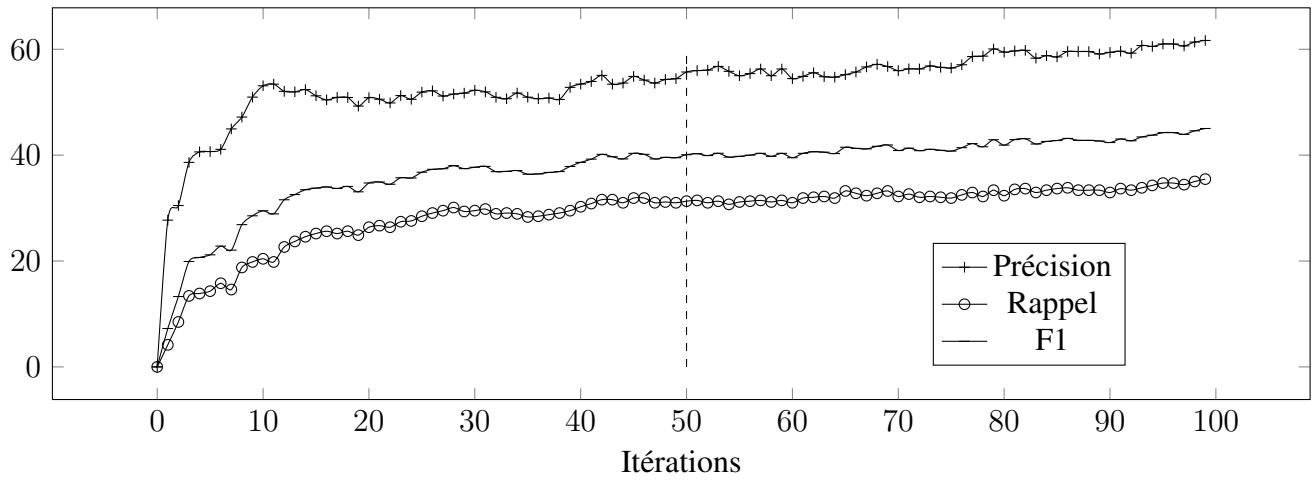


FIGURE 9.2 – Évaluation itérative du modèle CRF de détection d'opinion par ajouts d'exemples aléatoires

Nous identifions ces phases sur les données d'expérimentation, soit le corpus d'avis clients en français sur le domaine de la restauration annoté provenant de l'atelier SEMEVAL ABSA 2016, afin d'estimer le nombre de phrases nécessaires à la mise en place d'un modèle minimal. Un modèle d'étiquetage CRF est ainsi itérativement alimenté de 10 phrases annotées choisies aléatoirement dans le corpus d'entraînement. Nous constatons, comme illustré sur la figure 9.2, que l'apport des nouvelles annotations s'estompe lorsqu'environ 500 phrases sont fournies.

9.2 Amorçage à l'aide de patrons

En addition à un apprentissage actif, nous proposons, pour réduire l'effort d'étiquetage manuel pendant la phase d'initialisation, d'amorcer l'annotation à l'aide des patrons de détection. Si le rôle de l'apprentissage actif est en effet de rechercher les éléments déterminants parmi l'ensemble des premières annotations nécessaires, celui-ci repose sur une première sélection aléatoire, ce que nous supposons sous-optimal.

9.2.1 Opinions initiales

La notion d'amorce pose intrinsèquement la question du choix des exemples initiaux. Il n'est en effet pas envisageable de fournir à un premier modèle l'ensemble des phrases du corpus d'entraînement telles qu'analysées par les patrons, au risque de l'entraîner sur des exemples erronés, et plus simplement parce que cela ne laisserait aucune nouvelle annotation possible. Afin de limiter ces phrases à un sous-ensemble pertinent, nous expérimentons deux types de filtres, l'un sur les patrons, l'autre sur les cibles d'opinion extraites.

Sélection de patrons

La première méthode que nous proposons d'appliquer pour définir un ensemble initial de phrases d'entraînement est un filtre des patrons selon leur précision de détection. [Riloff et al., 2003] montrent ainsi que cette sélection réduit la dérive sémantique de l'amorce. Nous comptabilisons pour cela les occurrences d'expressions d'opinion correctes et incorrectes retrouvées pour chaque patron dans le corpus d'entraînement, comme indiqué pour quelques exemples dans le tableau 9.1. Un seuil de précision de 0,50 est empiriquement fixé afin de sélectionner des patrons pertinents et suffisamment couvrants.

Patron	#Opinions correctes	#Occurrences	Précision
concept VERB ADV *subjectivème* - ADJ	31	38	0,81
concept *negation* *subjectivème* - ADJ	4	7	0,57
subjectivème - NOUN ADP *concept*	5	9	0,55
concept VERB *subjectivème* - ADJ	100	185	0,54
concept *subjectivème* - VERB	47	95	0,49
concept *subjectivème* - ADJ	153	406	0,37
concept ADV *subjectivème* - ADJ	45	142	0,31
subjectivème - ADJ *concept*	87	287	0,30

TABLEAU 9.1 – Précision des patrons d'amorçage sur le corpus d'entraînement

Sélection de termes

Une seconde approche consiste à filtrer non pas la méthode de détection mais les éléments détectés. Il est en effet raisonnable de penser, à la manière de [Hu and Liu, 2004], que les sujets les plus fréquents parmi des avis clients sont dans un grand nombre de cas des cibles d'opinion. Nous conservons ainsi les termes dont le nombre d'occurrences est supérieur ou égal à 10, ce qui représente 47 éléments dans le corpus d'entraînement, tels que « *service* », « *restaurant* » ou « *accueil* ».

9.2.2 Amorçage

Nous comparons les filtres décrits – ainsi que l'absence et la combinaison de filtres – pour la méthode symbolique sur le corpus d'entraînement (et non le corpus d'évaluation puisque cela entraînerait un biais), puis nous comparons ces filtres sur leur capacité à sélectionner des exemples annotés pour un modèle CRF sur le corpus d'évaluation. L'ensemble de ces résultats sont indiqués dans le tableau 9.2. Nous pouvons y constater premièrement que l'apport d'un sous-ensemble de phrases d'exemple pour le modèle probabiliste est peu prévisible, puisque les résultats de chaque méthode sur le corpus d'entraînement ne correspondent pas à ceux du modèle CRF observés sur le corpus d'évaluation. À titre d'exemple, la combinaison des filtres (« P+C »), présentant la précision la plus élevée et le rappel le plus faible dans le premier cas, produit un modèle dont le rappel est important et la précision est la plus basse.

	Évaluation sur le corpus d'entraînement			Amorce sur le corpus d'évaluation			
Méthode	Précision	Rappel	F1	# Phrases	Précision	Rappel	F1
Tous patrons	46,83	**29,28**	**36,03**	1 705	50,23	15,95	**24,21**
Patrons précis (P)	68,18	8,49	15,11	183	**64,44**	4,32	8,10
Cible fréquente (C)	75,24	21,86	33,87	88	13,62	**36,51**	19,84
P+C	**82,31**	6,06	11,29	25	12,76	32,49	18,32

TABLEAU 9.2 – Comparaison des méthodes pour la constitution d'une amorce

Nous observons ensuite dans quelle mesure cette amorce influe les itérations d'apprentissage actif, que nous simulons à partir des annotations disponibles du corpus d'entraînement. Par souci de lisibilité, seule la courbe de progression du modèle dont l'amorce a produit les meilleurs résultats, c'est-à-dire la combinaison des filtres (« P+C », en rouge) est tracée sur la figure 9.3 aux côtés des courbes de référence, soit l'ajout d'annotations par échantillonnage incertain à partir de 10 annotations aléatoires (« Sans amorce », en bleu), et l'ajout totalement aléatoire présenté en section 9.1.3 (« Aléatoire », en noir). Si les performances des trois modèles sont similaires à la dernière itération, nous pouvons constater que cet « état stable » est atteint dès l'itération 15 pour le modèle amorcé, contre 30 itérations pour l'annotation sans amorce.

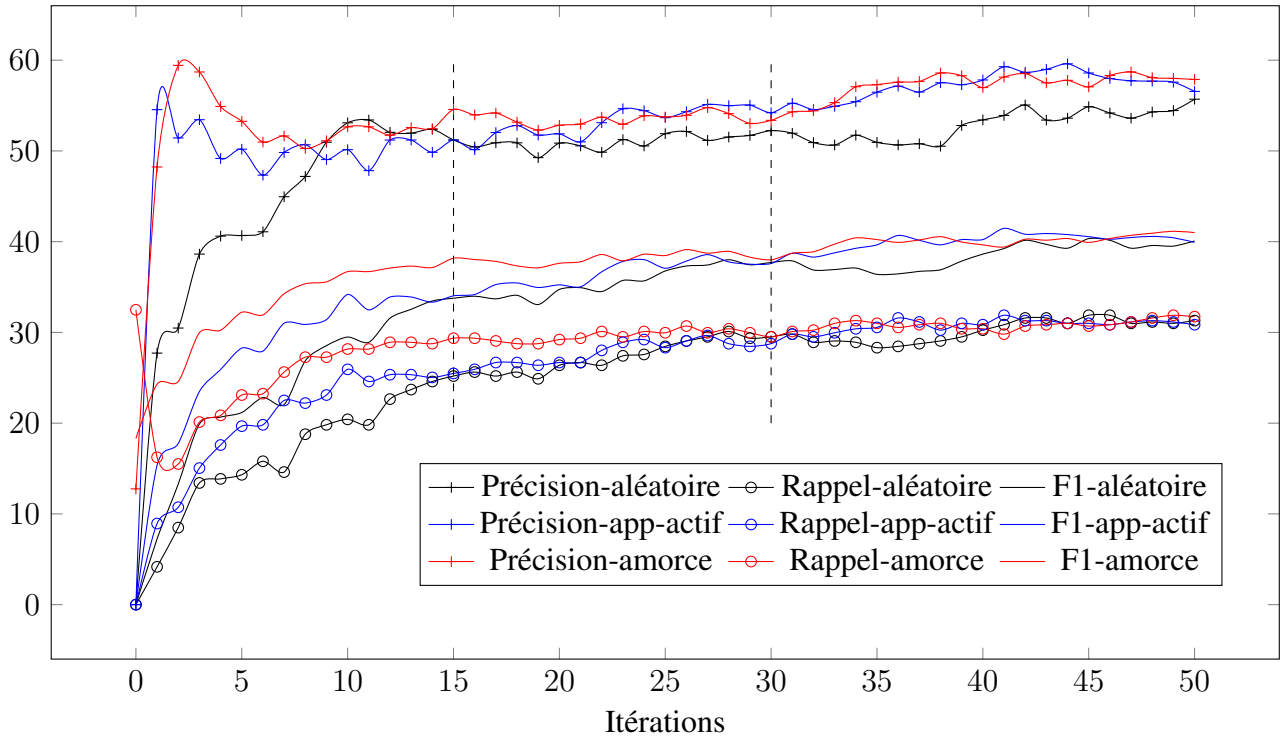


FIGURE 9.3 – Progression des performances des modèles sur le corpus d'évaluation

9.3 Rendre l'annotation accessible

Nous nous interrogeons maintenant sur des solutions d'aide lors de la phase d'*amélioration continue*, afin de compléter l'assistance apportée à l'annotateur au cours de son travail. Or nous constatons, par expérience vécue et observée, qu'un facteur important de la difficulté perçue de la tâche d'annotation est la manière dont elle est présentée (ce qui concerne par ailleurs aussi la phase d'initialisation). Après avoir présenté les particularités de la maintenance de ressources annotées dans notre cadre de travail, nous notons les interfaces existantes puis nous proposons un système répondant à nos besoins. Enfin nous étudions la question de la ludification de la tâche d'annotation.

9.3.1 Contraintes d'annotation

Annotation en comité restreint

Premièrement, la majeure partie des documents que nous souhaitons annoter ne peut être divulguée publiquement, ce qui interdit le recours au *crowdsourcing*, c'est-à-dire le fait de produire une ressource de qualité en demandant à une large population un moindre travail. Le rapport est donc inverse dans notre cas : un comité restreint doit effectuer l'intégralité de l'annotation. De surcroît, cela rend complexe la validation des annotations dans la mesure où ce comité compte une minorité d'experts linguistes.

Propriété des données

Deuxièmement, les entreprises clientes restent propriétaires des données recueillies et peuvent donc en exiger la suppression, qu'il s'agisse de documents annotés ou non. Si nous ne pouvons empêcher cela, il est néanmoins possible de limiter les pertes de ces données enrichies en multipliant les sources d'où nous sélectionnons des documents à étiqueter, ce qui signifie multiplier les travaux d'annotation.

Annotation et outils existants

Enfin, il est impératif que l'outil d'annotation que nous utilisons puisse s'intégrer dans notre suite logicielle sans y être nocif. Cette adaptation n'est pas seulement motivée par le respect de confidentialité mais également par le besoin de travailler avec les annotations automatiquement créées que nous exportons, comme nous avons pu le faire dans la première partie de ce chapitre.

9.3.2 Interface d'annotation

Travaux existants

Du fait du besoin omniprésent de création de ressources annotées, de multiples systèmes d'annotation proposant une interface utilisateur graphique ont été développés. Ainsi, MMAX2 [Müller and Strube, 2006] facilite l'étiquetage de coréférences dans un texte par une visualisation en chemins. KNOWTATOR [Ogren, 2006] a été conçu pour assister l'annotation dans le domaine médical, et est intégré au logiciel de gestion de connaissance Protégé[1]. CORPUSTOOL [O'Donnell, 2008] se veut être un outil simple et général pour l'annotation de textes et supporte plusieurs niveaux d'analyse. WEBANNOTATOR [Tannier, 2012] est une extension du navigateur Firefox[2] qui permet de labelliser divers types de contenus sur des pages web sans en dénaturer la structure en HTML. Le logiciel TEAMWARE [Bontcheva et al., 2013] de la suite GATE est un outil très complet pour l'annotation collaborative. Enfin l'outil d'annotation BRAT [Stenetorp et al., 2012] (développé par la suite sous le nom WEBANNO [Yimam et al., 2013]) représente le système existant le plus adapté à nos besoins, de par une interface simple, une configuration souple et la possibilité de structurer une tâche d'annotation en un projet collaboratif à petite échelle.

[1] protege.stanford.edu

[2] www.mozilla.org/fr/firefox/new/

Limites des outils

Les inconvénients que nous identifions à l'utilisation d'outils tels que BRAT concernent davantage la présentation des données que les fonctionnalités de l'outil. Nous listons ici les principaux.

- Sélection au mot : afin de faciliter le travail d'annotation, nous cherchons principalement à réduire l'effort cognitif nécessaire. Cela signifie centrer ce travail sur un unique problème, auquel l'utilisateur peut répondre de façon instantanée [Krug, 2005]. Le fait de présenter un texte brut dont la sélection est réalisée au niveau caractère interroge l'utilisateur sur plusieurs points, qui ne sont pas tous en rapport avec l'annotation visée. Il lui est ainsi demandé de résoudre par exemple la tokenisation des phrases et l'appartenance des tokens à d'éventuelles locutions polylexicales, ce qui, à notre sens, ne concerne pas l'annotation d'opinion qui suppose un prétraitement d'ores et déjà réalisé. Enfin de manière plus pragmatique, la sélection de caractères d'un empan de texte, à la souris par exemple, est une action lente comparativement au choix immédiat d'une étiquette pour cet empan [Weinschenk, 2011].
- Rejet d'un document : un second inconvénient que nous avons observé lors de l'utilisation de ces outils est l'impossibilité de rejeter un document, mais seulement de passer au suivant sans ajouter d'annotation. Or la valeur d'un document non annoté est bien différente de celle d'un document qui ne nécessite pas d'annotation. Ce système oblige l'utilisateur à recourir soit à la création d'une annotation d'un type « rejet », qu'il faut ensuite traiter, soit à l'identification et la suppression manuelle de documents de la pile à annoter.
- Progression de la tâche : nous avons également constaté une frustration des annotateurs due au manque de vision sur l'avancement de leur travail. Il n'est effectivement pas trivial d'évaluer mentalement l'apport de chaque annotation ajoutée au modèle, et pour les mêmes raisons de simplification que celles citées précédemment, cela ne devrait pas être demandé à l'utilisateur.
- Format d'annotation : enfin le choix d'un format peu standard pour les annotations créées à partir de l'outil rend complexe son interaction avec d'autres éléments de la chaîne de maintenance des ressources, notamment les systèmes d'apprentissage consommant ces annotations.

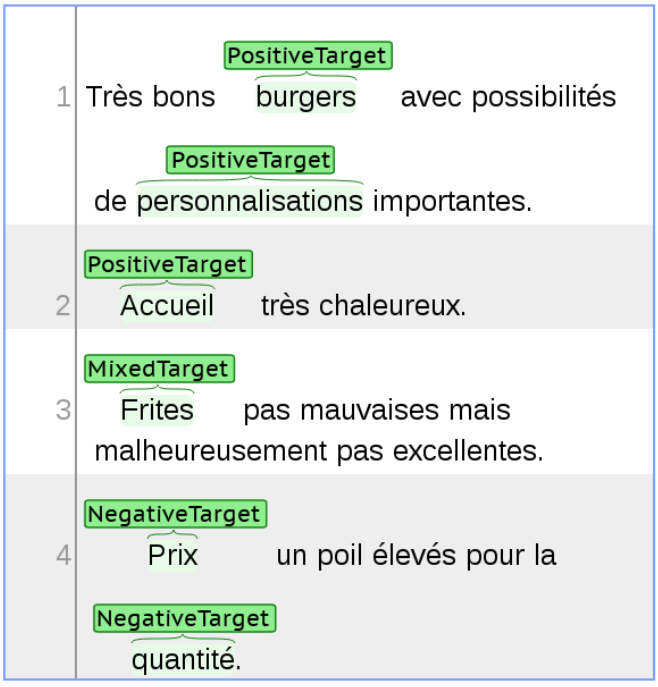


FIGURE 9.4 – Capture d'écran de Brat

ID	Étiquette	Début	Fin	Texte
T0	PositiveTarget	10	17	burgers
T1	PositiveTarget	70	77	Accueil
T2	PositiveTarget	39	56	personna- lisations
T3	MixedTarget	95	101	Frites
T4	NegativeTarget	154	158	Prix
T5	NegativeTarget	182	190	quantité

FIGURE 9.5 – Format d'annotation de Brat

Proposition

D'après notre étude des interfaces d'annotation de textes existantes, il apparaît que le domaine de la création de ressources annotées pourrait grandement bénéficier d'un rapprochement avec celui du design d'interfaces utilisateur.

Le prototype fonctionnel que nous avons conçu est une proposition dans ce sens. Il s'agit d'un logiciel doté d'une interface graphique épurée proposant deux types d'annotation, la catégorisation de documents et l'étiquetage de séquences de mots. Dans les deux cas, l'avancement de la tâche est visible, soit par une simple barre de progression faisant état du nombre de documents traités et restants, soit par l'affichage des performances du modèle à chaque ajout – comme illustré sur la figure 9.6 – si un corpus d'évaluation est fourni. Les documents problématiques, car non pertinents ou bruités, peuvent être passés et sont sauvegardés dans une base de données séparée. Enfin, les données annotées sont exportées au format attendu par les librairies d'apprentissage automatique que nous utilisons afin de proposer aisément une annotation par apprentissage actif d'une part, et de disposer d'une ressource directement disponible pour la construction d'un modèle d'autre part.

FIGURE 9.6 – Capture d'écran de l'interface d'annotation proposée (avec évaluation)

En ce qui concerne l'étiquetage de séquence, les mots sont présentés comme des blocs parmi lesquels il est possible de naviguer et fonctionnent comme des unités distinctes, ce qui en accélère la sélection et l'étiquetage comparativement à une sélection de caractères. Cette visualisation des mots permet en outre de révéler les documents dont le prétraitement est visiblement incorrect, ce qui offre la possibilité de ne pas les considérer et de les conserver ainsi à part pour faciliter l'analyse des erreurs.

À la différence d'autres outils d'annotation, cette interface n'est pas généraliste mais spécifiquement conçue pour la création de contenus annotés pour la fouille d'opinion. Cela se traduit par exemple par un ensemble prédéfini d'étiquettes, représentées par des couleurs pour réduire ici encore la charge cognitive lors de l'annotation. Cette spécificité d'utilisation, établie à partir de notre travail sur la modélisation de la fouille d'opinion, limite l'effort de configuration nécessaire pour notre équipe d'annotateurs et les rend ainsi plus autonomes.

9.3.3 Ludification de l'annotation

Rendre l'annotation accessible, ce n'est pas seulement construire un outil efficace, mais aussi changer la perception de la tâche, généralement considérée rébarbative. Certains travaux proposent ainsi de joindre l'utile à l'agréable en incitant les annotateurs à contribuer à une ressource par des mécaniques de jeu. Nous présentons ici quelques exemples de ces travaux et nous étudions les intérêts et limites de cette approche.

Travaux existants

PHRASES DETECTIVES[3] [Chamberlain et al., 2008] est une interface d'annotation de liens anaphoriques dont l'accroche narrative (ou *storytelling*) évoque un détective sur une piste d'indices. JEUXDEMOTS[4] [Lafourcade and Joubert, 2008] est un jeu d'association de mots récompensant par des points les joueurs qui indiquent de nombreux mots sémantiquement liés. C'est également l'objectif du jeu WORD SHERIFF [5] [Parasca et al., 2016], cependant dans celui-ci les utilisateurs jouent simultanément, ce qui crée une situation ludique différente. ZOMBILINGO [6] [Fort et al., 2014] est une interface d'annotation de rôles grammaticaux utilisant un *storytelling* du monde des zombies, souvent associé à la notion de jeu. BISAME[7] [Millour and Fort, 2016] demande le même type d'annotation sur des phrases en alsacien. Le joueur est ici seulement récompensé par des points (pas de *storytelling*), toutefois l'affichage d'un classement des participants contribue à la motivation des contributeurs. Enfin, ROBOCORP [8] [Dziedzic and Włodarczyk, 2016] utilise des mécaniques plus profondes pour l'annotation d'entités nommées en polonais. Le jeu propose ainsi un scénario, dans lequel un personnage – un petit robot – manque de vocabulaire et requiert par conséquent une participation du joueur pour être compris.

Intérêts et limites

Le sujet des travaux mentionnés précédemment est fortement lié à celui de la ludification des supports éducatifs et du travail collaboratif en entreprise, appelés dans les deux cas « jeux sérieux » ou « jeux avec un but ». Le principe général de ce processus est d'emprunter des éléments caractéristiques d'un jeu pour motiver la réussite d'un travail et le rendre moins rebutant. Les conséquences visées sont une meilleure productivité et une réduction du sentiment d'anxiété, voire un certain bien-être [Oprescu et al., 2014].

Ces « jeux sérieux » sont cependant régulièrement critiqués pour leur pauvreté en matière de mécaniques de jeu [Linehan et al., 2011], sous l'argument qu'un jeu (vidéo) n'est pas un logiciel auquel est ajoutée une « propriété de jeu », mais un programme conçu dès ses fondations autour de mécaniques propres à une expérience ludique, ce qui permet justement de susciter chez l'utilisateur un engagement fort [Bogost, 2011]. Ce phénomène est appelé « *chocolate covered broccoli* » (du « broccoli nappé de chocolat ») dans la littérature, symbolisant la tentative de rendre plus attrayant un élément indésirable à l'aide d'un élément plus apprécié, ce qui produit en définitive une association flatteuse ni pour l'un ni pour l'autre.

De ce débat entre création d'un nouveau programme et ludification (ou *gamification*) d'un outil existant émergent plusieurs questions au sujet de l'annotation. Celle du coût de développement tout d'abord, car le temps passé à l'élaboration d'un tel jeu doit être justifié par une productivité accrue lors de l'utilisation ; celle de l'engagement des utilisateurs également, dans la mesure où l'adhérence de l'ensemble de notre comité d'annotation à l'univers d'un jeu n'est pas assurée, tandis qu'une représentation plus simple d'une forme de « récompense », par exemple une attribution d'un score comptabilisant les annotations ajoutées, sera plus facilement adoptée ; enfin une vue sur les données réelles doit être conservée, afin de ne pas en décorréler l'annotation au point d'altérer le jugement de l'utilisateur.

Dans notre cas, ces éléments établissent un cadre qui est, à notre sens et à celui de notre équipe d'annotateurs, le plus adapté à la maintenance des ressources utilisées. Ce cadre théorique pourra par la suite être instancié au cours d'expérimentations autour de la ludification.

[3] www.phrasedetectives.org
[4] www.jeuxdemots.org
[5] comp3096.herokuapp.com
[6] zombilingo.org
[7] bisame.herokuapp.com
[8] citygames.zetsystem.com.pl/gwap/GWAP/

9.4 Synthèse

Dans ce chapitre, nous avons cherché à réduire l'effort réel et perçu que demande l'annotation de corpus pour la mise en place d'une méthode probabiliste supervisée, et en particulier l'étiquetage de séquence.

Dans un premier temps, nous avons proposé d'appliquer une stratégie d'apprentissage actif dont nous avons optimisé les premières itérations à l'aide de notre méthode symbolique de détection d'opinion.

- Un **ensemble de documents annotés automatiquement** au moyen des patrons les plus précis prévient le « démarrage à froid » de l'annotation ;
- Sur nos données d'étude, l'apprentissage actif réalisé à partir de cette amorce se révèle **plus efficace qu'en choisissant un ensemble initial aléatoirement**, dans le sens où des performances stables sont atteintes en moins d'itérations d'annotation.

Le second procédé par lequel cette tâche peut être rendue plus accessible est à notre sens l'amélioration de l'outil d'annotation, d'un point de vue ergonomique notamment. Les systèmes existants présentant certaines limites pour notre cadre de travail, nous avons proposé une nouvelle interface plus adaptée, actuellement utilisée par notre équipe d'annotateurs. Enfin nous avons étudié l'approche de ludification de l'annotation.

- L'interface proposée se veut la plus simple possible, en demandant uniquement à l'utilisateur une annotation d'opinion sur un contenu déjà prétraité. Cette annotation ne requiert pas de mouvements chronophages, mais uniquement des **interactions humain-machine instantanées**. Par ailleurs les documents non pertinents peuvent être ignorés ce qui représente là encore un gain de temps ;
- Nous retenons de la littérature sur le sujet de la **ludification** (ou *gamification*) des **arguments très contrastés** : promesse de productivité pour les uns, chimère exploitant l'image de média léger du jeu vidéo pour les autres, cette approche nous semble avant tout nécessiter une expérimentation en conditions réelles par des utilisateurs. À présent que nous avons défini un cadre de travail pour cela, nous y procéderons au cours de futurs travaux.

10

Conclusion

Les trois parties qui constituent ce manuscrit ne suivent pas l'ordre chronologique de nos travaux, mais peuvent être vues comme trois chemins que nous avons empruntés pour découvrir notre sujet, y apporter notre contribution, et au terme desquels nous apercevons de nouveaux sentiers à explorer.

10.1 Bilan

L'étude des prises de parole des consommateurs, et en particulier des éléments qui les composent, a été pour nous l'occasion d'aborder une réflexion sur la notion même d'opinion dans le discours. Le travail de définition de [Liu, 2012] sur ce sujet est un formidable point de départ, en ce qu'il établit un cadre d'étude scindant l'opinion en parties dont les rôles sont clairement identifiés. En cherchant à déterminer les éléments les plus utiles dans notre cas, nous avons été amené à questionner ce cadre et à l'adapter aux données que nous traitons et que nous présentons aux utilisateurs d'un outil de fouille d'opinion.

- Il apparaît clairement que l'analyse des avis clients ne doit pas présupposer des sujets abordés et de la façon dont ceux-ci sont qualifiés, mais doit permettre au contraire de **faire émerger ces éléments des corpus**. Cette « écoute libre » signifie d'une part une certaine **tolérance** vis-à-vis des structures attendues, en considérant par exemple comme sujets non pas seulement les groupes nominaux, mais les groupes verbaux voire adjectivaux, et comme indicateurs de l'opinion non pas seulement les adjectifs mais les noms, verbes et plus particulièrement les expressions complexes, qui sont bien souvent spécifiques à un domaine. D'autre part, il s'agit ensuite de **trier les informations pertinentes** parmi cette large collecte, qui par ailleurs contient de nombreux éléments bruités.

- Afin de réaliser cette sélection, nous proposons une **modélisation de l'opinion multi-classe**, dont chacune représente une entité, c'est-à-dire l'ensemble cohérent de termes exprimant la même notion. Nos expériences au moyen d'un étiquetage de séquence selon cette modélisation montrent non seulement que celle-ci permet d'obtenir une extraction **plus performante que dans le cas de l'annotation binaire** (opinion / non opinion), mais de surcroît que cette approche est une solution prometteuse au problème de l'**adaptation au domaine** [Lark et al., 2017].

Les éléments que nous cherchons sont extraits du discours au moyen de méthodes provenant de paradigmes opposés et dont les apports spécifiques sont relativement bien identifiés pour d'autres applications du traitement automatique des langues. À l'aube de notre parcours, le choix entre ces deux voies a donné lieu à de nombreux débats tant les avantages de l'un semblent compenser les inconvénients de l'autre. Ainsi, plutôt que de choisir entre la robustesse de l'approche symbolique et les performances de l'approche probabiliste, entre la rapidité de production de patrons et la facilité de maintenance des ressources annotées, ou entre la dérive des structures linguistiques définies a priori et l'imprévisibilité des modèles de classification, nous décidons de consacrer une partie importante de notre travail à en établir des comparaisons et à tenter de combiner leurs résultats.

- Nous cherchons essentiellement à concilier dans un même outil de fouille d'opinion un **fonctionnement par règles linguistiques** – que nous avons nommées « patrons de détection d'opinion » dans ce travail – faisant figure d'approche sécurisante d'un point de vue applicatif, car d'ores et déjà mis en place d'une part et déterministe d'autre part, et un **fonctionnement statistique par l'emploi d'un modèle markovien CRF** dont les performances pour l'extraction d'opinion à granularité fine ont été plusieurs fois prouvées, notamment lors des ateliers SEMEVAL sur ce sujet, mais qui représente néanmoins un pari risqué pour une application produisant des résultats directement visibles par les utilisateurs.

- Lors de nos expériences, nous avons pu confirmer une supériorité du modèle probabiliste pour l'identification des cibles d'opinion, cependant la différence entre les deux paradigmes s'estompe lorsque la polarité de ces cibles est prise en compte. La combinaison que nous proposons, soit l'**utilisation des résultats de la méthode symbolique comme traits de classification pour l'étiquetage de séquence**, se révèle plus performante que chacune des deux approches, c'est pourquoi cette solution sera intégrée dans la chaîne de traitement de notre outil de fouille d'opinion.

Muni de ces informations essentielles, nous avons pu nous concentrer sur la construction des différents types de ressources sur lesquelles les méthodes d'extraction s'appuient. Dans chaque cas, nous avons cherché un moyen – compatible avec le fonctionnement de l'outil de fouille d'opinion auquel nous contribuons – d'acquérir des appuis pertinents, couvrants, et dont les processus d'extraction ne sont pas spécifiques à une langue en particulier.

- La constitution d'un **lexique affectif**, ingrédient indispensable pour toutes les méthodes de détection d'opinion, a été notre première priorité. Nous retrouvons dans la littérature de nombreux travaux proposant soit leur **identification au sein de corpus**, par proximité avec des sujets abordés ou par similarité de contexte avec des mots marqueurs d'opinion connus, soit une **extension de lexique existant** – qui peut être instancié par quelques éléments définis manuellement – par recherche de liens sémantiques dans une base de connaissance. Cette dernière option se révèle cependant peu précise, quel que soit le type de lien sémantique employé, c'est pourquoi nous avons préféré travailler à cette construction à partir de corpus [Lark et al., 2015]. Les deux méthodes les plus prometteuses dans cette voie ont été une **extraction reposant sur des patrons symboliques** et un **étiquetage de séquence à partir de marqueurs d'opinion projetés** automatiquement. En sus de la pertinence globale des éléments retrouvés, ces deux méthodes sont particulièrement intéressantes pour leur capacité à détecter des **expressions complexes** dont la prise en compte est fondamentale [Lark et al., 2016].

- Nous avons ensuite généralisé l'extraction de marqueurs d'opinion multi-mots afin de **générer de nouveaux patrons**. La méthode que nous proposons repose sur un classement de segments de phrases dont la longueur et l'abstraction (c'est-à-dire le fait de considérer certains mots sous une forme plus générique que leur forme lexicale, comme l'étiquette grammaticale) sont très variables. Un **grand nombre de patrons candidats** sont donc générés, et leur différentiation ne tient qu'au **score de pertinence** que nous leur attribuons. Cela peut expliquer les résultats moyens que nous avons obtenus pour ce qui est de l'enrichissement d'une liste de patrons d'ores et déjà établie et maintes fois corrigée manuellement. En revanche, cette méthode se montre très efficace pour la **création d'une nouvelle ressource**. Si les opinions détectées par ces patrons générés ne sont pas exemptes d'erreur, il est important de noter que cette méthode est bien plus performante pour l'**initialisation d'un système de fouille d'opinion** dans un domaine que son alternative, soit l'utilisation d'un modèle d'étiquetage de séquence entraîné sur un domaine différent.

- La perspective de cette initialisation à l'aide de patrons nous a amené à réfléchir sur une seconde hybridation de notre système, cette fois dans l'objectif de pallier le problème du **choix des premiers exemples lors de l'annotation** (ou « démarrage à froid »). Nous avons établi que dans le cas d'un **apprentissage actif**, qui nous semble la meilleure option pour l'établissement d'un modèle probabiliste de détection d'opinion, il est plus efficace d'**amorcer l'annotation** par un ensemble de documents automatiquement traités au moyen de l'approche symbolique. Le travail sur l'optimisation des itérations d'annotation suivantes nous a conduit à étudier les problèmes d'**ergonomie d'interface et de ludification**. Nous avons ainsi proposé une **interface simple** et **adaptée à l'étiquetage d'opinion** pour les annotateurs de notre entreprise, contrastant avec les outils existants plus complexes.

10.2 Travaux futurs

Chacune des pistes que nous avons suivies a dévoilé dans une certaine mesure des chantiers sur lesquels nous souhaiterions travailler prochainement. Nous différencions parmi ceux-ci les évolutions naturelles de nos contributions et les expériences que nous n'avons pas jugées suffisamment concluantes pour être présentées dans ce travail, mais dont le principe nous apparaît encore aujourd'hui comme suffisamment prometteur pour être réitérées.

Dans cette deuxième catégorie, nous incluons nos tentatives d'instancier de manière plus concrète les couples terme—subjectivème que nous avons décrits au chapitre 2. Nous avons en effet expérimenté la construction de lexique affectif « conditionnel », où un marqueur d'opinion ne peut être activé qu'avec une liste fermée de termes cibles d'opinion. Cette restriction permet en principe d'améliorer la précision des opinions extraites en limitant les risques de dérive sémantique des marqueurs d'opinion, cependant nous n'avons pas réussi à concilier cet apport avec le principe d'« écoute libre » des avis clients.

L'acquisition automatique d'annotations au moyen d'un apprentissage par renforcement (*reinforcement learning*) nous a également semblé prometteuse, mais les résultats obtenus jusqu'ici n'ont pas été probants. Le principe de cette acquisition est relativement similaire à celui de notre méthode d'extraction de patrons, dans le sens où il s'agit de générer des permutations d'éléments puis de n'en conserver que les plus pertinents pour l'analyse. Dans le cas de cette expérience, ces permutations correspondent aux annotations de cibles d'opinion possibles à partir de phrases dont les emplacements des termes sont connus. Les phrases ainsi annotées – éventuellement choisies par échantillonnage incertain – sont ajoutées à un ensemble d'entraînement d'un modèle supervisé, puis nous observons à l'aide d'un ensemble d'évaluation si cet ajout améliore ou dégrade les performances de ce modèle.

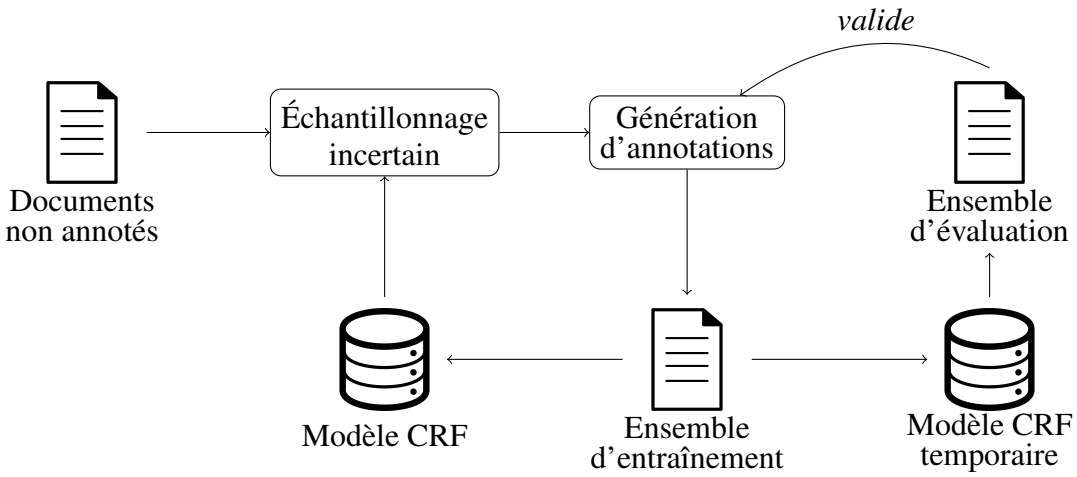


FIGURE 10.1 – Schéma de l'annotation automatique par renforcement

Ce processus est bien évidemment séduisant de par l'automatisation de l'annotation, toutefois les limites auxquelles nous nous sommes heurtés lors de nos expériences en ont atténué l'intérêt. D'une part, le critère de sélection présente un fort risque de sur-apprentissage pour le corpus d'évaluation utilisé, ce qui produit des exemples d'entraînement incohérents. En effet les « meilleurs combinaisons » générées, au sens de la mesure F1 du modèle ainsi alimenté, sont régulièrement des cibles d'opinion incorrectes pour un annotateur humain, ce qui est contre-intuitif avec l'idée de l'apprentissage supervisé. D'autre part, la combinatoire de cette génération ne nous autorise à ajouter que quelques phrases par itération à l'ensemble d'entraînement, ce qui ne permet pas de réaliser une mesure très fiable.

Les évolutions que nous avons envisagées pour nos contributions constituent des pistes de travail plus concrètes. La validation de la modélisation multi-classe de l'opinion, présentée au chapitre 3, fera l'objet d'une étude prochaine sur un plus grand nombre d'entités. En outre, cette vision modulaire de la fouille d'opinion correspond à une volonté de notre entreprise d'accueil de structurer davantage les briques logicielles impliquées dans l'analyse du texte autour de la notion d'entité.

Le second chantier auquel nous sommes appelés à participer à court terme est la multiplication des amorces pour la mise en place de modèles de détection d'opinion, non seulement pour différents domaines mais également pour différentes langues. Le multilinguisme est en effet devenu un sujet de préoccupation majeur depuis le début de nos travaux et nous avons à cœur d'y contribuer au moyen des méthodes d'acquisition de marqueurs d'opinion, développées au chapitre 7, et du système hybride, décrit chapitres 8 et 9. Ce déploiement ne consiste cependant pas uniquement en une application de méthodes établies, mais requiert de plus amples études sur la création de patrons et sur l'amorçage de l'étiquetage de séquence dans chaque langue.

Enfin, si l'interface d'annotation que nous avons proposée à notre équipe d'annotateurs est d'ores et déjà utilisée, celle-ci est perfectible et, comme nous l'avons évoqué au chapitre 9, pourrait bénéficier d'une ludification. Les travaux sur le sujet de l'ergonomie de telles interfaces auxquels nous nous sommes intéressés, à première vue éloignés de notre problématique, montrent à notre sens le besoin de futures contributions dans cette voie pour la création de contenus annotés, qui sont actuellement des ressources rares. Ces productions seraient d'autant plus précieuses dans le cadre d'une analyse à l'échelle du mot telle que la notre, qui requiert une annotation subtile et chronophage.

Table des matières

Liste des tableaux

Table des figures

Bibliographie

[Abacha and Zweigenbaum, 2011] Abacha, A. B. and Zweigenbaum, P. (2011). Medical entity recognition : A comparison of semantic and statistical methods. In Proceedings of BioNLP 2011 Workshop, pages 56–64, Portland, OR, USA. 73

[Abney, 1996] Abney, S. (1996). Statistical methods and linguistics. The balancing act : Combining symbolic and statistical approaches to language, pages 1–26. 72

[Alm et al., 2005] Alm, C. O., Roth, D., and Sproat, R. (2005). Emotions from text : Machine learning for text-based emotion prediction. In Proceedings of HLT-EMNLP 2005, pages 579–586, Vancouver, BC, Canada. 61

[Alshawi, 1994] Alshawi, H. (1994). Qualitative and quantitative models of speech translation. The Balancing Act : Combining symbolic and statistical approaches to language. 73

[Augustyn et al., 2006] Augustyn, M., Hamou, S. B., Bloquet, G., Goossens, V., Loiseau, M., and Rinck, F. (2006). Lexique des affects : constitution de ressources pédagogiques numériques. In Proceedings of CEDIL 2006, pages 407–414, Grenoble, France. 62

[Awasthi et al., 2006] Awasthi, P., Rao, D., and Ravindran, B. (2006). Part of speech tagging and chunking with HMM and CRF. In Proceedings of the NLPAI Machine Learning Competition, pages 69–40, Mumbai, India. 22

[Baccianella et al., 2010] Baccianella, S., Esuli, A., and Sebastiani, F. (2010). Sentiwordnet 3.0 : An enhanced lexical resource for sentiment analysis and opinion mining. In Proceedings of LREC 2010, pages 2200–2204, Valletta, Malta. 86, 88

[Bengio et al., 2003] Bengio, Y., Ducharme, R., Vincent, P., and Jauvin, C. (2003). A neural probabilistic language model. In Journal of machine learning research, volume 3, pages 1137–1155. 95

[Benveniste, 1958] Benveniste, E. (1958). De la subjectivité dans le langage. In Problèmes de linguistique générale, page 258–266. 23

[Blair-Goldensohn and Hannan, 2008] Blair-Goldensohn, S. and Hannan, K. (2008). Building a sentiment summarizer for local service reviews. In NLP in the Information Explosion Era Workshop, pages 339–348, Beijing, China. 40

[Blei et al., 2003] Blei, D. M., Ng, A. Y., Jordan, M. I., and Lafferty, J. (2003). Latent dirichlet allocation. In Journal of Machine Learning Research, volume 2, pages 993–1022. 87

[Blitzer et al., 2007] Blitzer, J., Dredze, M., Pereira, F., et al. (2007). Biographies, bollywood, boom-boxes and blenders : Domain adaptation for sentiment classification. In Proceedings of ACL 2007, volume 7, pages 440–447, Prague, Czech Republic. 20, 29

[Bogost, 2011] Bogost, I. (2011). Gamification is bullshit. The gameful world : Approaches, issues, applications, pages 65–80. 113

[Bontcheva et al., 2013] Bontcheva, K., Cunningham, H., Roberts, I., Roberts, A., Tablan, V., Aswani, N., and Gorrell, G. (2013). Gate teamware : A web-based, collaborative text annotation framework. In Language Resources and Evaluation, volume 47, pages 1007–1029. 110

[Boubel, 2012] Boubel, N. (2012). Construction automatique d'un lexique de modifieurs de polarité. In Proceedings of TALN 2012, volume 3, pages 123–136, Grenoble, France. 55

[Bourigault, 1994] Bourigault, D. (1994). Lexter : un Logiciel d'EXtraction de TERminologie : application à l'acquisition des connaissances à partir de textes. PhD thesis. 22

[Bradley and Lang, 1999] Bradley, M. M. and Lang, P. J. (1999). Affective norms for English words (ANEW) : Instruction manual and affective ratings. Technical report. 62

[Breck et al., 2007] Breck, E., Choi, Y., and Cardie, C. (2007). Identifying expressions of opinion in context. In Proceedings of IJCAI 2007, pages 2683–2688, Hyderabad, India. 64, 65

[Brun and Gala, 2012] Brun, C. and Gala, N. (2012). Propagation de polarités dans des familles de mots : impact de la morphologie dans la construction d'un lexique pour l'analyse d'opinions. In Proceedings of TALN 2012, pages 495–502, Grenoble, France. 89

[Brun et al., 2016] Brun, C., Perez, J., and Roux, C. (2016). XRCE at SemEval-2016 Task 5 : Feedbacked Ensemble Modelling on Syntactico-Semantic Knowledge for Aspect Based Sentiment Analysis. In Proceedings of the SemEval 2016 ABSA Workshop, pages 277–281, San Diego, CA, USA. 64

[Cambria et al., 2015] Cambria, E., Gastaldo, P., Bisio, F., and Zunino, R. (2015). An elm-based model for affective analogical reasoning. In Neurocomputing, volume 149, pages 443–455. 27

[Chamberlain et al., 2008] Chamberlain, J., Poesio, M., and Kruschwitz, U. (2008). Phrase detectives : A web-based collaborative annotation game. In Proceedings of I-Semantics 2008, pages 42–49, Karlsruhe, Germany. 113

[Chen and Skiena, 2014] Chen, Y. and Skiena, S. (2014). Building Sentiment Lexicons for All Major Languages. In Proceedings of ACL 2014, pages 383–389, Baltimore, ML, USA. 89

[Chen et al., 2014] Chen, Z., Mukherjee, A., and Liu, B. (2014). Aspect Extraction with Automated Prior Knowledge Learning. In Proceedings of ACL 2014, pages 347–358, Baltimore, ML, USA. 89

[Choi and Cardie, 2009] Choi, Y. and Cardie, C. (2009). Adapting a polarity lexicon using integer linear programming for domain-specific sentiment classification. In Proceedings of EMNLP 2009, pages 590–598, Singapore. 29, 87

[Choi et al., 2005] Choi, Y., Cardie, C., Riloff, E., and Patwardhan, S. (2005). Identifying sources of opinions with conditional random fields and extraction patterns. In Proceedings of HLT-EMNLP 2005, pages 355–362, Vancouver, BC, Canada. 64, 73

[Church and Hanks, 1990] Church, K. W. and Hanks, P. (1990). Word association norms, mutual information, and lexicography. In Computational linguistics, volume 16, pages 22–29. 22

[Daille, 1996] Daille, B. (1996). Study and implementation of combined techniques for automatic extraction of terminology. In The balancing act : Combining symbolic and statistical approaches to language, pages 49–66. 74

[Das and Bandyopadhyay, 2010] Das, A. and Bandyopadhyay, S. (2010). SentiWordNet for Indian languages. In Proceedings of the 8th Workshop on Asian Language Resources, pages 56–63, Beijing, China. 89

[de Albornoz et al., 2012] de Albornoz, J., Plaza, L., and Gervás, P. (2012). SentiSense : An easily scalable concept-based affective lexicon for sentiment analysis. In Proceedings of LREC 2012, pages 3562–3567, Istanbul, Turkey. 27

[Deng and Wiebe, 2015] Deng, L. and Wiebe, J. (2015). Joint prediction for entity/event-level sentiment analysis using probabilistic soft logic models. In Proceedings of EMNLP 2015, pages 179–189, Lisbon, Portugal. 73

[Dey and Haque, 2009] Dey, L. and Haque, S. M. (2009). Opinion mining from noisy text data. In Proceedings of AND 2008, volume 12, pages 205–226, Singapore. 89

[Ding et al., 2008] Ding, X., Liu, B., and Yu, P. S. (2008). A holistic lexicon-based approach to opinion mining. In Proceedings of WSDM 2008, pages 231–240, Palo Alto, CA, USA. 39, 86

[Dziedzic and Włodarczyk, 2016] Dziedzic, D. and Włodarczyk, W. (2016). Making NLP games with a purpose fun to play using Free to Play mechanics : RoboCorp case study. In Proceedings of EACL - Games4NLP Workshop, pages 187–197, Valencia, Spain. 113

[Ekman and Oster, 1979] Ekman, P. and Oster, H. (1979). Facial expressions of emotion. In Annual review of psychology, volume 30, pages 527–554. 61

[Esuli, 2008] Esuli, A. (2008). Automatic generation of lexical resources for opinion mining : models, algorithms and applications. PhD thesis. 88

[Fan et al., 2008] Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R., and Lin, C.-J. (2008). Liblinear : A library for large linear classification. In Journal of Machine Learning Research, volume 9, pages 1871–1874. 62

[Fellbaum, 1998] Fellbaum, C. (1998). WordNet : An Electronic Lexical Database. MIT Press. 22, 35, 88

[Feng et al., 2011] Feng, S., Bose, R., and Choi, Y. (2011). Learning general connotation of words using graph-based algorithms. In Proceedings of EMNLP 2011, pages 1092–1103, Edinburgh, Scotland. 87

[Fort et al., 2014] Fort, K., Guillaume, B., and Stern, V. (2014). Zombilingo : manger des têtes pour annoter en syntaxe de dépendances. In Proceedngs of TALN 2014, pages 15–16, Marseille, France. 113

[Garcia Fernandez and Ferret, 2012] Garcia Fernandez, A. and Ferret, O. (2012). Etude de différentes stratégies d'adaptation à un nouveau domaine en fouille d'opinion. In Proceedings of TALN 2012, pages 391–398, Grenoble, France. 29

[Généreux and Evans, 2006] Généreux, M. and Evans, R. (2006). Towards a validated model for affective classification of texts. In Proceedings of the Workshop on Sentiment and Subjectivity in Text, pages 55–62, Sydney, Australia. 27

[Godbole et al., 2007] Godbole, N., Srinivasaiah, M., and Skiena, S. (2007). Large-Scale Sentiment Analysis for News and Blogs. In Proceedings of ICWSM 2007, pages 219–222, Boulder, CO, USA. 89

[Grefenstette, 1993] Grefenstette, G. (1993). Automatic thesaurus generation from raw text using knowledge-poor techniques. In Making sense of Words. Ninth Annual Conference of the UW Centre for the New OED and text Research, Oxford, England. 17

[Hamdan et al., 2015] Hamdan, H., Bellot, P., and Bechet, F. (2015). Lsislif : CRF and Logistic Regression for Opinion Target Extraction and Sentiment Polarity Analysis. In Proceedings of the SemEval 2015 ABSA Workshop, pages 753–758, Denver, CO, USA. 64

[Hamilton et al., 2016] Hamilton, W. L., Clark, K., Leskovec, J., and Jurafsky, D. (2016). Inducing domain-specific sentiment lexicons from unlabeled corpora. arXiv preprint arXiv :1606.02820. 87

[Hamon et al., 2015] Hamon, T., Fraisse, A., Paroubek, P., Zweigenbaum, P., and Grouin, C. (2015). Analyse des émotions, sentiments et opinions exprimés dans les tweets : présentation et résultats de l'édition 2015 du défi fouille de texte (DEFT). In Proceedings of TALN 2015, Caen, France. 34

[Hatzivassiloglou, 1996] Hatzivassiloglou, V. (1996). Do we need linguistics when we have statistics ? a comparative analysis of the contributions of linguistic cues to a statistical word grouping system. In The balancing act : Combining symbolic and statistical approaches to language, pages 67–94. 73

[Hatzivassiloglou and McKeown, 1997] Hatzivassiloglou, V. and McKeown, K. R. (1997). Predicting the semantic orientation of adjectives. In Proceedings of ACL 1997, pages 174–181, Madrid, Spain. 55, 86, 94

[Hernandez et al., 2015] Hernandez, N., Jadi, G., Lark, J., and Monceaux, L. (2015). Exploitation de lexiques pour la catégorisation fine d'émotions, de sentiments et d'opinions. In Proceedings of DEFT 2015, pages 51–60, Caen, France. 61

[Hochreiter and Schmidhuber, 1997] Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. In Neural Computation, volume 9, pages 1735–1780, Cambridge, MA, USA. 63

[Hu and Liu, 2004] Hu, M. and Liu, B. (2004). Mining and summarizing customer reviews. In Proceedings of KDD 2004, pages 168–177, Seattle, WA, USA. 39, 42, 86, 88, 91, 92, 108

[Huang et al., 2014] Huang, M., Ye, B., Wang, Y., Chen, H., Cheng, J., and Zhu, X. (2014). New Word Detection for Sentiment Analysis. In Proceedings of ACL 2014, pages 531–541, Baltimore, ML, USA. 87

[Huang et al., 2015] Huang, Z., Xu, W., and Yu, K. (2015). Bidirectional LSTM-CRF models for sequence tagging. arXiv preprint arXiv :1508.01991. 63

[Jacquemin, 1994] Jacquemin, C. (1994). FASTR : A unification-based front-end to automatic indexing. In Intelligent Multimedia Information Retrieval Systems and Management, volume 1, pages 34–47. 19, 22

[Jakob and Gurevych, 2010] Jakob, N. and Gurevych, I. (2010). Extracting Opinion Targets in a Single- and Cross-Domain Setting with Conditional Random Fields. In Proceedings of EMNLP 2010, pages 1035–1045, Cambridge, MA, USA. 20, 29, 39, 64, 68

[Jijkoun et al., 2010] Jijkoun, V., de Rijke, M., and Weerkamp, W. (2010). Generating focused topic-specific sentiment lexicons. In Proceedings of ACL 2010, pages 585–594, Uppsala, Sweden. 87

[Jijkoun et al., 2008] Jijkoun, V., Khalid, M. A., Marx, M., and De Rijke, M. (2008). Named entity normalization in user generated content. In Proceedings of the second workshop on Analytics for noisy unstructured text data, pages 23–30, Singapore. 19

[Joshi et al., 2014] Joshi, A., Mishra, A., Senthamilselvan, N., and Bhattacharyya, P. (2014). Measuring Sentiment Annotation Complexity of Text. In Proceedings of ACL 2014, pages 36–41, Baltimore, ML, USA. 107

[Kanayama and Nasukawa, 2006] Kanayama, H. and Nasukawa, T. (2006). Fully automatic lexicon expansion for domain-oriented sentiment analysis. In Proceedings of EMNLP 2006, pages 355–363, Sydney, Australia. 20, 86

[Kapur and Clark, 1996] Kapur, S. and Clark, R. (1996). The automatic construction of a symbolic parser via statistical techniques. The Balancing Act : Combining Symbolic and Statistical Approaches to Language. 72

[Kerbrat-Orecchioni, 1980] Kerbrat-Orecchioni, C. (1980). L'Énonciation, De la subjectivité dans le langage. Armand Colin. 24

[Kim and Hovy, 2004] Kim, S.-M. and Hovy, E. (2004). Determining the sentiment of opinions. In Proceedings of COLING 2004, page 1367, Geneva, Switzerland. 35, 39, 88

[Klavans and Resnik, 1996] Klavans, J. L. and Resnik, P. (1996). The balancing act : Combining symbolic and statistical approaches to language. 72

[Krug, 2005] Krug, S. (2005). Don't Make Me Think : A Common Sense Approach to the Web (2nd Edition). New Riders Publishing. 111

[Ku et al., 2005] Ku, L.-W., Lee, L.-Y., Wu, T.-H., and Chen, H.-H. (2005). Major topic detection and its application to opinion summarization. In Proceedings of SIGIR 2005, pages 627–628, Salvador, Brazil. 40

[Kudo, 2005] Kudo, T. (2005). CRF++ : Yet another CRF toolkit. https ://taku910.github.io/crfpp/. 66

[Lafferty et al., 2001] Lafferty, J. D., McCallum, A., and Pereira, F. C. N. (2001). Conditional random fields : Probabilistic models for segmenting and labeling sequence data. In Proceedings of ICML 2001, pages 282–289, San Francisco, CA, USA. 44, 61, 63

[Lafourcade and Joubert, 2008] Lafourcade, M. and Joubert, A. (2008). Jeuxdemots : un prototype ludique pour l'émergence de relations entre termes. In Proceedings of JADT 2008, pages 657–666, Lyon, France. 113

[Lark et al., 2017] Lark, J., Morin, E., and Peña Saldarriaga, S. (2017). A comparative study of target-based and entity-based opinion extraction. In Proceedings of CICLING 2017, Budapest, Hungary. 43, 116

[Lark et al., 2015] Lark, J., Morin, E., and Peña Saldarriaga, S. (2015). CANÉPHORE : un corpus français pour la fouille d'opinion ciblée. In Proceedings of TALN 2015, pages 418–424, Caen, France. 74, 117

[Lark et al., 2016] Lark, J., Morin, E., and Saldarriaga, S. (2016). Extraction d'opinions ambigües dans des corpus d'avis clients. In Proceedings of TALN 2016, pages 451–458, Paris, France. 117

[Lavalley et al., 2010] Lavalley, R., Clavel, C., and Bellot, P. (2010). Extraction probabiliste de chaînes de mots relatives à une opinion. Traitement Automatique des Langues, 51 :101–130. 39

[Levenshtein, 1966] Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. In Soviet physics doklady, volume 10, pages 707–710. 22

[Lewis and Gale, 1994] Lewis, D. D. and Gale, W. A. (1994). A sequential algorithm for training text classifiers. In Proceedings of SIGIR 1994, pages 3–12, Dublin, Ireland. 44, 106

[L'Homme, 2004] L'Homme, M.-C. (2004). La terminologie : principes et techniques. Pum. 20

[Li et al., 2010] Li, F., Han, C., Huang, M., Zhu, X., Xia, Y.-J., Zhang, S., and Yu, H. (2010). Structure-aware review mining and summarization. In Proceedings of COLING 2010, pages 653–661, Beijing, China. 64

[Li et al., 2012] Li, F., Pan, S., Jin, O., Yang, Q., and Zhu, X. (2012). Cross-domain co-extraction of sentiment and topic lexicons. In Proceedings of ACL 2012, pages 410–419, Uppsala, Sweden. 64, 87

[Linehan et al., 2011] Linehan, C., Kirman, B., Lawson, S., and Chan, G. (2011). Practical, appropriate, empirically-validated guidelines for designing educational games. In Proceedings of SIGCHI 2011, pages 1979–1988, Vancouver, BC, Canada. 113

[Liu, 2012] Liu, B. (2012). Sentiment Analysis and Opinion Mining. Morgan & Claypool. 6, 20, 25, 26, 30, 41, 116

[Liu et al., 2005] Liu, B., Hu, M., and Cheng, J. (2005). Opinion observer : analyzing and comparing opinions on the web. In Proceedings of WWW 2005, pages 342–351, Chiba, Japan. 20, 21, 40, 129

[Liu et al., 2003] Liu, H., Lieberman, H., and Selker, T. (2003). A model of textual affect sensing using real-world knowledge. In Proceedings of IUI 2003, page 125, Miami, FL, USA. 27

[Liu et al., 2015] Liu, Q., Gao, Z., Liu, B., and Zhang, Y. (2015). Automated Rule Selection for Aspect Extraction in Opinion Mining. In Proceedings of IJCAI 2015, pages 1291–1297, Buenos Aires, Argentina. 39

[Llosa, 1996] Llosa, S. (1996). Contributions à l'étude de la satisfaction dans les services. PhD thesis. 40

[Maks and Vossen, 2012] Maks, I. and Vossen, P. (2012). Building a fine-grained subjectivity lexicon from a web corpus. In Proceedings of LREC 2012, number 3, pages 3070–3076, Istanbul, Turkey. 87

[Marchand, 2013] Marchand, M. (2013). Fouille d'opinion : ces mots qui changent de polarité selon le domaine. In Proceedings of CORIA 2013, pages 347–352, Neuchâtel, Switzerland. 20

[Marchand et al., 2014] Marchand, M., Besançon, R., Mesnard, O., and Vilnat, A. (2014). Influence des marqueurs multi-polaires dépendant du domaine pour la fouille d'opinion au niveau du texte. In Proceedings of TALN 2014, pages 1–12, Marseille, France. 29

[Martin and White, 2003] Martin, J. R. and White, P. R. (2003). The language of evaluation, volume 2. Springer. 24

[Meng et al., 2012] Meng, X., Wei, F., Liu, X., Zhou, M., Li, S., and Wang, H. (2012). Entity-centric topic-oriented opinion summarization in twitter. In Proceedings of SIGKDD 2012, pages 379–387. 40

[Mikolov et al., 2013] Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv :1301.3781, pages 1–12. 22, 95

[Millour and Fort, 2016] Millour, A. and Fort, K. (2016). Construction de ressources langagières annotées par myriadisation pour le traitement automatique des langues peu dotées : le cas de l'alsacien. Technical report. 113

[Moghaddam and Ester, 2013] Moghaddam, S. and Ester, M. (2013). The flda model for aspect-based opinion mining : Addressing the cold start problem. In Proceedings of WWW 2013, pages 909–918, New York, NY, USA. 70

[Mohammad et al., 2009] Mohammad, S., Dunne, C., and Dorr, B. (2009). Generating high-coverage semantic orientation lexicons from overtly marked words and a thesaurus. In Proceedings of EMNLP 2009, pages 599–608, Singapore. 89

[Morin and Jacquemin, 2004] Morin, E. and Jacquemin, C. (2004). Automatic acquisition and expansion of hypernym links. In Computers and the Humanities, page 363–396. 22

[Müller and Strube, 2006] Müller, C. and Strube, M. (2006). Multi-level annotation of linguistic data with mmax2. In Corpus technology and language pedagogy : New resources, new tools, new methods, volume 3, pages 197–214. 110

[Murray and Carenini, 2011] Murray, G. and Carenini, G. (2011). Subjectivity detection in spoken and written conversations. Natural Language Engineering. 98, 99, 104, 130

[Nazarenko et al., 2009] Nazarenko, A., Zargayouna, H., Hamon, O., and Van Puymbrouck, J. (2009). Évaluation des outils terminologiques : enjeux, difficultés et propositions. In Traitement automatique des Langues, volume 50, pages 257–281. 90

[Needleman and Wunsch, 1970] Needleman, S. B. and Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. In Journal of molecular biology, volume 48, pages 443–453. 22

[Ogren, 2006] Ogren, P. V. (2006). Knowtator : a protégé plug-in for annotated corpus construction. In Proceedings of NAACL-HLT 2006 (demonstrations), pages 273–275, New York City, NY, USA. 110

[Oprescu et al., 2014] Oprescu, F., Jones, C., and Katsikitis, M. (2014). I play at work—ten principles for transforming work processes through gamification. Frontiers in psychology, 5. 113

[Otmakhova and Shin, 2015] Otmakhova, Y. and Shin, H. (2015). Do We Really Need Lexical Information ? Towards a Top-down Approach to Sentiment Analysis of Product Reviews. In Proceedings of HLT-NAACL 2015, pages 1559–1568, Denver, CO, USA. 86

[O'Donnell, 2008] O'Donnell, M. (2008). The uam corpustool : Software for corpus annotation and exploration. In Proceedings of AESLA 2008, pages 3–5, Almeria, Spain. 110

[Pak and Paroubek, 2010] Pak, A. and Paroubek, P. (2010). Construction d'un lexique affectif pour le français à partir de twitter. In Proceedings of TALN 2010, Montréal, Canada. 36, 87

[Pang and Lee, 2008] Pang, B. and Lee, L. (2008). Opinion Mining and Sentiment Analysis. Now Publishers Inc. 6, 20

[Pang et al., 2002] Pang, B., Lee, L., and Vaithyanathan, S. (2002). Thumbs up ? : sentiment classification using machine learning techniques. In Proceedings of EMNLP 2002, pages 79–86, Philadelphia, PA, USA. 34, 35

[Parasca et al., 2016] Parasca, I.-E., Rauter, A. L., Roper, J., Rusinov, A., Bouchard, G., Riedel, S., and Stenetorp, P. (2016). Defining words with words : Beyond the distributional hypothesis. In Proceedings of the 1st Workshop on Evaluating Vector-Space Representations for NLP, pages 122–126, Berlin, Germany. 113

[Park et al., 2002] Park, Y., Byrd, R. J., and Boguraev, B. K. (2002). Automatic glossary extraction : Beyond terminology identification. In Proceedings of COLING 2002, pages 1–7, Taipei, Taiwan. 19

[Pazienza et al., 2005] Pazienza, M. T., Pennacchiotti, M., and Zanzotto, F. M. (2005). Terminology extraction : an analysis of linguistic and statistical approaches. In Knowledge mining, pages 255–279. 73

[Perez-Rosas et al., 2012] Perez-Rosas, V., Banea, C., and Mihalcea, R. (2012). Learning Sentiment Lexicons in Spanish. In Proceedings of LREC 2012, page 73, Istanbul, Turkey. 89

[Petz et al., 2013] Petz, G., Karpowicz, M., Fürschuß, H., Auinger, A., Stříteskỳ, V., and Holzinger, A. (2013). Opinion mining on the web 2.0–characteristics of user generated content and their impacts. In Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data, pages 35–46. 19

[Piao et al., 2005] Piao, S. S., Rayson, P., Archer, D., and McEnery, T. (2005). Comparing and combining a semantic tagger and a statistical tool for mwe extraction. In Computer Speech & Language, volume 19, pages 378–397. 74

[Pontiki et al., 2016] Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., AL-Smadi, M., Al-Ayyoub, M., Zhao, Y., Qin, B., De Clercq, O., Hoste, V., Apidianaki, M., Tannier, X., Loukachevitch, N., Kotelnikov, E., Bel, N., Jiménez-Zafra, S. M., and Eryigit., G. (2016). Semeval-2016 task 5 : Aspect based sentiment analysis. In Proceedings of SemEval 2016, San Diego, CA, USA. 34, 64

[Pontiki et al., 2015] Pontiki, M., Galanis, D., Papageorgiou, H., Manandhar, S., and Androutsopoulos, I. (2015). Semeval-2015 task 12 : Aspect based sentiment analysis. In Proceedings of SemEval 2015, pages 486–495, Denver, CO, USA. 26, 34, 64

[Pontiki et al., 2014] Pontiki, M., Galanis, D., Pavlopoulos, J., Papageorgiou, H., Androutsopoulos, I., and Manandhar, S. (2014). Semeval-2014 task 4 : Aspect based sentiment analysis. In Proceedings of SemEval 2014, Dublin, Ireland. 34, 64

[Popescu and Etzioni, 2005] Popescu, A.-M. and Etzioni, O. (2005). Extracting Product Features and Opinion from Reviews. In Proceedings of HLT-EMNLP 2005, pages 339–346, Vancouver, Canada. 50

[Price, 1994] Price, P. (1994). Combining linguistic with statistical methods in automatic speech understanding. The Balancing Act : Combining symbolic and statistical approaches to language, pages 76–83. 72

[Qiu et al., 2009] Qiu, G., Liu, B., Bu, J., and Chen, C. (2009). Expanding domain sentiment lexicon through double propagation. In Proceedings of IJCAI 2009, pages 1199–1204, Pasadena, CA, USA. 39, 87

[Qiu et al., 2011] Qiu, G., Liu, B., Bu, J., and Chen, C. (2011). Opinion word expansion and target extraction through double propagation. In Computational Linguistics, volume 37, pages 9–27. 50

[Rastier, 1995] Rastier, F. (1995). Le terme : entre ontologie et linguistique. In La banque des mots, pages 35–65. 17

[Read and Carroll, 2009] Read, J. and Carroll, J. (2009). Weakly supervised techniques for domain-independent sentiment classification. In Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion, pages 45–52, Hong Kong, HK. 29

[Reichheld, 2001] Reichheld, F. (2001). Loyalty Rules ! Harvard Business School Press. 37

[Riloff and Wiebe, 2003] Riloff, E. and Wiebe, J. (2003). Learning extraction patterns for subjective expressions. In Proceedings of EMNLP 2003, pages 105–112, Sapporo, Japan. 73, 98, 99, 104

[Riloff et al., 2003] Riloff, E., Wiebe, J., and Wilson, T. (2003). Learning Subjective Nouns using Extraction Pattern Bootstrapping. In Proceedings of CONLL 2003, pages 25–32, Edmonton, Canada. 22, 50, 73, 86, 87, 92, 108

[Rosé et al., 2003] Rosé, C. P., Roque, A., Bhembe, D., and Vanlehn, K. (2003). A hybrid text classification approach for analysis of student essays. In Proceedings of HLT-NAACL-EDUC 2003, pages 68–75, Edmonton, Canada. 73

[Rosé and Waibel, 1994] Rosé, C. P. and Waibel, A. (1994). Recovering from parser failures : A hybrid statistical/symbolic approach. arXiv preprint cmp-lg/9407025. 73

[Sager, 1990] Sager, J. C. (1990). A practical course in terminology processing. John Benjamins Publishing. 17

[Seddah et al., 2012] Seddah, D., Sagot, B., Candito, M., Mouilleron, V., and Combet, V. (2012). The french social media bank : a treebank of noisy user generated content. In Proceedings of COLING 2012, Mumbai, India. 19

[Settles, 2010] Settles, B. (2010). Active learning literature survey. Technical report. 106

[Settles et al., 2008] Settles, B., Craven, M., and Friedland, L. (2008). Active learning with real annotation costs. In Proceedings of the NIPS workshop on cost-sensitive learning, pages 1–10, Vancouver, Canada. 107

[Seung et al., 1992] Seung, H. S., Opper, M., and Sompolinsky, H. (1992). Query by committee. In Proceedings of COLT 1992, pages 287–294, Pittsburgh, PA, USA. 106

[Socher et al., 2010] Socher, R., Manning, C. D., and Ng, A. Y. (2010). Learning continuous phrase representations and syntactic parsing with recursive neural networks. In Proceedings of the NIPS-2010 Deep Learning and Unsupervised Feature Learning Workshop, pages 1–9, Whistler, BC, Canada. 22

[Spertus, 1997] Spertus, E. (1997). Smokey : Automatic recognition of hostile messages. In Proceedings of AAAI-IAAI 1997, pages 1058–1065, Providence, RI, USA. 34

[Staiano and Guerini, 2014] Staiano, J. and Guerini, M. (2014). DepecheMood : a Lexicon for Emotion Analysis from Crowd-Annotated News. arXiv preprint arXiv :1405.1605. 87

[Stenetorp et al., 2012] Stenetorp, P., Pyysalo, S., Topić, G., Ohta, T., Ananiadou, S., and Tsujii, J. (2012). Brat : A web-based tool for nlp-assisted text annotation. In Proceedings of EACL 2012, pages 102–107, Avignon, France. 110

[Taboada et al., 2011] Taboada, M., Brooke, J., Tofiloski, M., Voll, K., and Stede, M. (2011). Lexicon-based methods for sentiment analysis. In Computational Linguistics, volume 37, pages 267–307. 55

[Täckström and McDonald, 2011] Täckström, O. and McDonald, R. (2011). Discovering fine-grained sentiment with latent variable structured prediction models. In Proceedings of ECIR 2011, pages 368–374, Dublin, Ireland. 35

[Tannier, 2012] Tannier, X. (2012). Webannotator, an annotation tool for web pages. In Proceedings of LREC 2012, pages 316–319, Istanbul, Turkey. 110

[Thelen and Riloff, 2002] Thelen, M. and Riloff, E. (2002). A Bootstrapping Method for Learning Semantic Lexicons using Extraction Pattern Contexts. In Proceedings of EMNLP 2002, pages 214–221, Philadelphia, PA, USA. 50, 73, 87, 93

[Tomanek, 2010] Tomanek, K. (2010). Resource-aware annotation through active learning. PhD thesis. 107

[Tomanek and Hahn, 2009] Tomanek, K. and Hahn, U. (2009). Semi-supervised active learning for sequence labeling. In Proceedings of IJCNLP 2009, pages 1039–1047, Singapore. 106

[Turney, 2002] Turney, P. (2002). Thumbs up or thumbs down ? : semantic orientation applied to unsupervised classification of reviews. In Proceedings of ACL 2002, pages 417–424, Philadelphia, PA, USA. 34, 35, 87

[Vernier, 2011] Vernier, M. (2011). Analyse à granularité fine de la subjectivité. PhD thesis. 89

[Vernier et al., 2011] Vernier, M., Monceaux, L., and Daille, B. (2011). Identifier la cible d'un passage d'opinion dans un corpus multithématique. In Proceedings of TALN 2011, Montpellier, France. 28

[Vincze and Bestgen, 2011] Vincze, N. and Bestgen, Y. (2011). Identification de mots germes pour la construction d'un lexique de valence au moyen d'une procédure supervisée. In Proceedings of TALN 2011, pages 223–234, Montpellier, France. 86

[Wang and Cardie, 2014] Wang, L. and Cardie, C. (2014). A Piece of My Mind : A Sentiment Analysis Approach for Online Dispute Detection. In Proceedings of ACL 2014, pages 693–699, Baltimore, ML, USA. 35

[Weinschenk, 2011] Weinschenk, S. (2011). 100 things every designer needs to know about people. Pearson Education. 111

[Westerhout and Monachesi, 2008] Westerhout, E. and Monachesi, P. (2008). Creating Glossaries Using Pattern-Based and Machine Learning Techniques. In Proceddings of LREC 2008, Marrakech, Morocco. 74

[Wiebe and Mihalcea, 2006] Wiebe, J. and Mihalcea, R. (2006). Word sense and subjectivity. In Proceedings of ACL 2006, pages 1065–1072, Sidney, Australia. 39

[Wiebe et al., 2005] Wiebe, J., Wilson, T., and Cardie, C. (2005). Annotating expressions of opinions and emotions in language. Language Resources and Evaluation, Volume 39, pages 165–210. 86

[Wiebe et al., 1999] Wiebe, J. M., Bruce, R. F., and O'Hara, T. P. (1999). Development and Use of a Gold - Standard DataSet for Subjectivity Classifications. In Proceedings of ACL 1999, pages 246–253, College Park, ML, USA. 34

[Wiegand and Klakow, 2010] Wiegand, M. and Klakow, D. (2010). Bootstrapping supervised machine-learning polarity classifiers with rule-based classification. In Proceedings of WASSA 2010, pages 59–66, Lisbon, Portugal. 74

[Wilson et al., 2005] Wilson, T., Wiebe, J., and Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In Proceedings of HLT 2005, pages 347–354, Vancouver, Canada. 39, 55

[Wu et al., 2006] Wu, C.-H., Chuang, Z.-J., and Lin, Y.-C. (2006). Emotion recognition from text using semantic labels and separable mixture models. In Asian and Low-Resource Language Information Processing, volume 5, pages 165–183. 61

[Yang et al., 2007] Yang, C., Lin, K. H.-Y., and Chen, H.-H. (2007). Emotion classification using web blog corpora. In Proceedings of WI 2007, pages 275–278, Silicon Valley, CA, USA. 61

[Yang et al., 2014] Yang, M., Peng, B., Chen, Z., Zhu, D., and Chow, K. (2014). A Topic Model for Building Fine-grained Domain-specific Emotion Lexicon. In Proceedings of ACL 2014, pages 421–426, Baltimore, ML, USA. 27, 87

[Yi et al., 2003] Yi, J., Nasukawa, T., Bunescu, R., and Niblack, W. (2003). Sentiment analyzer : Extracting sentiments about a given topic using natural language processing techniques. In Proceedings of ICDM 2003, pages 427–434, Melbourne, Florida, USA. 50

[Yimam et al., 2013] Yimam, S. M., Gurevych, I., de Castilho, R. E., and Biemann, C. (2013). Webanno : A flexible, web-based and visually supported system for distributed annotations. In Proceedings of ACL 2013 (Demonstrations), pages 1–6, Sofia, Bulgaria. 110

[Yiou et al., 2015] Yiou, L., Hang, L., Wu, J., and Li, X. (2015). An Empirical Study on Sentiment Classification of Chinese Review using Word Embedding. In Proceedings of PACLIC 2015, Shanghai, China. 102

[Yu and Hatzivassiloglou, 2003] Yu, H. and Hatzivassiloglou, V. (2003). Towards answering opinion questions : Separating facts from opinions and identifying the polarity of opinion sentences. In Proceedings of EMNLP 2003, pages 129–136, Sapporo, Japan. 34

[Yu et al., 2013] Yu, H., Ho, C., Juan, Y., and Lin, C. (2013). Libshorttext : A library for short-text classification and analysis. Technical report. 62

[Zhiqiang and Wenting, 2014] Zhiqiang, T. and Wenting, W. (2014). DLIREC : Aspect Term Extraction and Term Polarity Classification System. In Proceedings of SemEval 2014 ABSA Workshop, pages 235–240, Dublin, Ireland. 64

Annexe : traits de classification

Template de traits utilisé pour les expériences employant la librairie CRF++. Les chiffres entre crochets correspondent ici respectivement à la position du mot par rapport au mot courant, et au type d'information utilisé parmi : 1) la forme textuelle du mot, 2) sa forme lemmatisée, 3) sa catégorie grammaticale, 4) son rôle du point du vue de l'opinion (terme, subjectivème, négation) et 5) le fait que le mot soit contenu dans une opinion reconnue par l'approche symbolique.

```
# Unigram
U111:%x[-1,1]
U112:%x[0,1]
U113:%x[1,1]

U221:%x[-1,2]
U222:%x[0,2]
U223:%x[1,2]

U331:%x[-1,3]
U332:%x[0,3]
U333:%x[1,3]

U441:%x[-1,4]
U442:%x[0,4]
U443:%x[1,4]

# Bigram
B111:%x[-1,1]
B112:%x[0,1]
B113:%x[1,1]

B221:%x[-1,2]
B222:%x[0,2]
B223:%x[1,2]

B331:%x[-1,3]
B332:%x[0,3]
B333:%x[1,3]

B431:%x[-1,4]
B432:%x[0,4]
B433:%x[1,4]

## Traits utilisés dans le cas
## de la méthode hybride
#U441:%x[-1,5]
#U442:%x[0,5]
#U443:%x[1,5]

#B431:%x[-1,5]
#B432:%x[0,5]
#B433:%x[1,5]
```

UNIVERSITE BRETAGNE LOIRE

UNIVERSITÉ DE NANTES

Thèse de Doctorat

Joseph LARK

Construction semi-automatique de ressources pour la fouille d'opinion

Semi-automatic acquisition of opinion mining resources

Résumé
Identifier les leviers de satisfaction des consommateurs est aujourd'hui capital dans un monde où la relation que tisse une entreprise avec ses clients est sa plus grande richesse. Le domaine de la fouille d'opinion, dans lequel s'inscrit cette thèse, propose des méthodes permettant de répondre à ce besoin. Celles-ci nécessitent cependant une mise à jour constante de ressources spécialisées qui sont la pierre angulaire des outils d'analyse d'opinion.
Ce travail vise à développer des stratégies d'acquisition et de structuration de ces ressources, qui prennent la forme de lexiques, de patrons morpho-syntaxiques ou de textes annotés. Chacune de ces formes présente des difficultés d'acquisition propres, auxquelles s'ajoute la complexité de mettre à jour ces ressources en fonction de la langue à traiter ou du domaine des corpus analysés, notion primordiale en fouille d'opinion.
Premièrement, nous menons une étude des éléments fondamentaux autour desquels l'opinion est construite dans le discours, conduisant à une nouvelle modélisation en étiquetage de séquence de l'opinion. Nous traitons ensuite la question de l'apport des différents types de ressources, dont il ressort que la meilleure stratégie est de les utiliser de concert. Enfin, nous proposons des méthodes d'acquisition pour chacune des ressources répondant non seulement aux besoins de la fouille d'opinion mais également aux contraintes du contexte industriel au sein duquel ces recherches sont menées.

Mots clés
Fouille d'opinion, traitement automatique des langues, ressources spécialisées.

Abstract
Identifying satisfaction triggers among customers is a crucial need in today's business world, as a strong customer relationship is now a most vital asset. The domain of opinion mining, in which this thesis falls into, offers several methods to answer this need. These methods, however, require a continuous update of specialized resources which are the cornerstone of many opinion mining tools.
The objective of this work is to develop acquisition and structuration strategies for these resources, which can be lexicons, morphosyntactic rules or annotated data. Each of these items presents its own extraction difficulties, on top of the general issue of their update in a language- or domain-specific setting. Indeed, language constraints are fundamental in opinion mining, so the proposed methods must take these into account.
First, we study the core elements from which opinion expressions are built in customer feedback. This study leads us to suggest a new modelisation of opinion mining as a sequence labeling task. We then compare the benefits of each type of resource through a benchmark of several opinion mining methods, and conclude that the best performing strategy is a hybrid approach. Finally, we present results for resource acquisition methods that answer not only the needs of opinion mining but also the constraints from the industrial setting in which this work has been conducted.

Key Words
Opinion mining, natural language processing, specialized resources.