

Catégorisation automatique d'adjectifs d'opinion à partir d'une ressource linguistique générique

Baptiste Chardon^{1,2}

(1) Synapse Développement, 33 rue Maynard, 31000 Toulouse

(2) IRIT, UMR5505, 31000 Toulouse

chardon@irit.fr, baptiste.chardon@synapse-fr.com

Résumé Cet article décrit un processus d'annotation manuelle de textes d'opinion, basé sur un schéma fin d'annotation indépendant de la langue et du corpus. Ensuite, à partir d'une partie de ce schéma, une méthode de construction automatique d'un lexique d'opinion à partir d'un analyseur syntaxique et d'une ressource linguistique est décrite. Cette méthode consiste à construire un arbre de décision basé sur les classes de concepts de la ressource utilisée. Dans un premier temps, nous avons étudié la couverture du lexique d'opinion obtenu par comparaison avec l'annotation manuelle effectuée sur un premier corpus de critiques de restaurants. La généralité de ce lexique a été mesurée en le comparant avec un second lexique, généré à partir d'un corpus de commentaires de films. Dans un second temps, nous avons évalué l'utilisabilité du lexique au travers d'une tâche extrinsèque, la reconnaissance de la polarité de commentaires d'internautes.

Abstract This paper introduces a manual annotation process of opinion texts, based on a fine-grained annotation scheme, independent of language and corpus. Then, from a part of this scheme, a method to build automatically an opinion lexicon from a syntactic analyser and a linguistic resource is described. This method consists in building a decision tree from the classes of the resource. The coverage of the lexicon has been determined by comparing it to the gold annotation of a restaurants review corpus. Its genericity was determined by comparing it to another lexicon generated from a different domain corpus (movie reviews). Eventually, the usefulness of the lexicon has been measured with an extrinsic task, the recognition of the polarity of reviews.

Mots-clés : Analyse d'opinion, Extension de lexique, Annotation d'opinions

Keywords: Opinion mining, Lexicon extension, Opinion annotation

1 Introduction

Afin de répondre au besoin grandissant d'automatisation du processus de traitement des données d'opinion, la recherche en extraction d'opinion a connu un essor important ces dernières années. Plusieurs projets de représentation et d'annotations de l'opinion dans les textes ont été réalisés : citons notamment le projet d'annotations MPQA (Wiebe et al., 2005), les travaux liés à la théorie Appraisal (Martin et al., 2005), et les travaux basés sur l'analyse du discours (Asher et al., 2009). En parallèle de

ces approches plus linguistiques, des méthodes computationnelles de constitution de ressources linguistiques ont été proposées, suivant deux approches principales : approches dites à base de corpus (Turney et al., 2002) et à base de dictionnaires (Kamps et al., 2004). Dans ce contexte, nos travaux répondent à plusieurs objectifs :

1. Étudier finement les expressions d'opinion en français. Pour cela, nous proposons en *section 2* un schéma générique d'annotation des opinions d'un texte prenant en compte le maximum d'éléments linguistiques jugés pertinents pour l'évaluation et la compréhension de ces opinions. Ce modèle a été utilisé et évalué dans le cadre d'une annotation de critiques francophones de restaurants.
2. Afin d'automatiser le processus d'extraction d'opinions, il est nécessaire d'élaborer une ressource linguistique décrivant les termes porteurs d'opinions. Nous proposons en *section 3* un algorithme permettant d'étendre une liste de mots germes à partir d'une ressource linguistique générique et d'un analyseur syntaxique.
3. Enfin, pour évaluer la qualité du lexique généré, nous proposons trois étapes de test. Tout d'abord, nous avons évalué la couverture du lexique en le comparant à celui produit lors de l'annotation manuelle. Ensuite, nous avons testé sa généralité en comparant les lexiques produits à partir de deux listes de germes différentes, issues de deux corpus de domaines distincts (critiques de films et critiques de restaurants). Nous avons enfin testé l'utilisabilité de ce lexique en déterminant automatiquement la polarité de critiques de restaurants. Les résultats de ces évaluations sont détaillés en *section 4*.

2 Modèle d'annotation pour l'étude fine des expressions d'opinion

Une opinion peut être positive, négative ou neutre. Cette distinction s'avère néanmoins insuffisante. En effet, à cela s'ajoutent des nuances d'intensité, par exemple entre des adjectifs tels que "*excellent*" et "*correct*". Une opinion ne se limite pas non plus aux seules expressions d'opinion : elle est nécessairement émise par une personne ou un groupe de personnes, qui peut éventuellement être différent du locuteur. De même, elle porte sur un sujet, plus ou moins précis. D'autres éléments externes peuvent également venir modifier la sémantique de l'opinion. Ces éléments, qui jouent le rôle d'opérateur, peuvent apporter des modifications au niveau de la polarité et/ou de l'intensité, du degré de certitude de l'émetteur, ou du degré de véracité d'une opinion rapportée. Enfin, une opinion s'inscrit dans un contexte discursif, et est par conséquent reliée à d'éventuelles autres opinions au travers des relations du discours. Notre objectif pour établir notre modèle d'annotation était de prendre en compte la totalité de ces éléments.

2.1 Projets d'annotation d'opinions existants

Parmi les travaux existants, le plus connu est sans doute le projet d'annotation MPQA (Wiebe et al., 2005), qui se base sur la notion d'état privé de (Quirk et al., 1985). Il prévoit une annotation de la polarité des opinions, de leur intensité sur quatre niveaux (échelle allant de faible à extrême), une distinction entre opinions explicites et implicites (i.e. supportées par des termes précis du texte, ou sous-entendues), et une annotation de l'émetteur de l'opinion.

D'un autre côté, (Whitelaw et al., 2005) ont utilisé la théorie Appraisal (Martin et al., 2005) afin de proposer une classification des expressions d'opinion via un large panel d'attributs. Néanmoins, ces travaux se limitent principalement à la classification fine d'adjectifs d'opinion et d'opérateurs d'intensité.

(Asher et al., 2009) ont quant à eux proposé un travail très descriptif d'annotation de l'opinion. Les opinions ont été classifiées suivant quatre catégories : sentiment, jugement, conseil et reportage (i.e. verbes introducteurs de la parole) ; chacune de ces catégories étant liées à des informations complémentaires de polarité, d'intensité, d'engagement de l'émetteur. Ces opinions étaient ensuite reliées à leur émetteur et à leur sujet, et enfin reliées entre elles via un sous-ensemble des relations discursives de la SDRT (Théorie des Représentations Discursives Segmentées, Asher et Lascarides, 2003).

Nos travaux se situent dans le cadre de la SDRT et dans la continuité de (Asher et al., 2009), avec une reprise des catégories et éléments précédemment évoqués. Ceux-ci ont été retravaillés et complétés, afin de prendre en compte notamment les opérateurs de modalité, d'intensité, et de négation, ainsi qu'un ensemble plus large de relations de discours. L'annotation manuelle répond à plusieurs objectifs. Tout d'abord, elle a permis de valider la pertinence des catégories et relations sélectionnées, dans le sens où un annotateur humain doit pouvoir annoter chaque phrase sans trop de doute. Elle a également permis de tester la validité de nos hypothèses concernant les éléments intervenant dans le calcul de l'opinion au niveau local, avec les opérateurs de négation, d'intensité et de modalité. De même, elle permet d'étudier au niveau global l'importance des phénomènes discursifs, via notamment les relations de la SDRT sélectionnées. L'annotation des éléments a enfin permis de mesurer et valider les résultats des applications d'extraction d'opinions en français.

2.2 Modèle proposé

Voici le modèle que nous avons utilisé pour l'annotation du document. Cette annotation est centrée sur les opinions véhiculées par le document et prend en compte le maximum d'éléments linguistiques jugés pertinents pour l'évaluation et la compréhension de ces opinions. Le modèle proposé est indépendant de la langue, même s'il s'appuie sur des exemples en français. De même, il est indépendant du domaine.

Un document est un triplet : $D = \langle O, U, R \rangle$, où O désigne l'ensemble des expressions d'opinion élémentaires du document, U l'ensemble des segments de discours du document, et R l'ensemble des relations entre les éléments de U . Plus précisément, O est un ensemble de quadruplets $\langle \text{Sujet}, \text{Émetteur}, \text{Expressions d'opinion}, \text{Opérateurs} \rangle$, tels que :

- *Sujet* désigne ici le sujet de l'opinion. *Émetteur* désigne l'émetteur de l'opinion.

- *Expressions d'opinion* désigne un quadruplet : $\langle \text{txt}, \text{cat}, \text{val}, \text{hyp} \rangle$, tels que *txt* désigne les bornes de l'annotation (lexicalisation), *cat* la catégorie sémantique de l'expression, *val* une mesure de la polarité et de l'intensité de l'expression (valeur entière entre -3 pour "très négatif" et +3 pour "très positif", 0 désignant une opinion mitigée ou ambiguë), enfin, *hyp* est un booléen, vrai si l'opinion est émise dans le cadre d'une hypothèse (Si ce restaurant est bon, nous y retournerons.). La catégorie sémantique peut être :

- **Explicite** : désigne une opinion explicitement supportée par des termes précis du texte. L'expression peut être un avis (*Le restaurant est bon.*) ou une recommandation (*Je recommande fortement les ramen.*).
- **Implicite** désigne les opinions sous-entendues, les faits étant interprétés comme opinion dans le contexte. (*De la terrasse, on voit toute la ville.*)

- *Opérateurs* désigne les différents opérateurs qui peuvent influencer sur l'opinion, nous distinguons :

- Les opérateurs d'engagement : verbes introducteurs de l'opinion (ex : Paul affirme que ce restaurant est bon.).
- Les opérateurs de négation : cet opérateur peut avoir une portée locale (ex : Ce restaurant n'est pas bon.), ou globale (ex : Il est impossible, même en tenant compte des points évoqués ci-dessus, de qualifier ce restaurant de bon.).
- Les opérateurs de modalité : (ex : Il est possible que ce restaurant soit bon.),
- Les opérateurs d'intensité (ex : Ce restaurant est très bon.).

U est l'ensemble des segments de discours contenant une expression d'opinion, ou étant relié à un élément de U par une relation $r \in R$ (cf. Asher et al., 2009, pour la définition d'un segment de discours).

R est un sous-ensemble des relations discursives de la SDRT, à savoir les relations $\{R_{continuation} R_{contraste} R_{attribution} R_{condition} R_{résultat} R_{support}\}$ (cf. Asher et al, 2009, pour la définition de ces relations).

2.3 Corpus utilisés et accords inter annotateurs

Le corpus principal annoté est constitué de commentaires sur des restaurants ou bars déposés en ligne sur le site <http://www.qype.fr/>. Ces commentaires sont libres, de temps à autre hors sujet, et contiennent parfois de multiples fautes d'orthographe. Afin de ne pas influencer l'annotateur, le corpus présenté ne contenait que le texte du commentaire, et aucune des informations annexes disponibles en ligne.

L'annotation a été effectuée par deux annotateurs familiers du domaine, sous l'environnement Glozz (<http://www.glozz.org/>). Le corpus était constitué de 40 documents (commentaires) a priori plutôt positifs et de 20 documents plutôt négatifs. Chaque commentaire est constitué en moyenne d'environ 3 phrases / 70 mots. Un manuel d'annotation a été fourni aux annotateurs afin de les guider dans la tâche. Après une première annotation de 10 documents, un retour a été effectué sur le manuel afin de préciser et corriger certains points ambigus. L'annotation des éléments discursifs a été laissée en suspens pour le moment.

À titre d'exemple, un de nos annotateurs propose l'annotation suivante de la phrase "*La viande est de très bonne qualité, le service également avec des serveurs anglophones.*" :

La <Sujet id=1 type="part of">viande</Sujet> est de <Opérateur id=2 type="intensite" valeur="+">très</Opérateur> <ExpressionOpinion id=3 categorie="explicite" valence="+2">bonne qualité</ExpressionOpinion>, le <Sujet id=4 type="part of">service</Sujet> <ExpressionOpinion id=5 categorie="explicite" valence="+2">également</ExpressionOpinion> avec des serveurs anglophones.

Les accords inter annotateurs présentés ici prennent en compte, pour le moment, la lexicalisation des expressions d'opinion (correspondance des bornes de l'annotation), leur polarité et leur catégorie. Le calcul d'accord s'est fait de manière souple : les deux annotateurs sont considérés comme en accord sur une annotation s'il y a chevauchement entre deux annotations. Les valeurs ont été calculées de la même manière que (Wiebe et al., 2005) l'ont effectué pour MPQA, l'accord $\text{agr}(A||B)$ indiquant la proportion d'annotations relevées par l'annotateur A également relevées par l'annotateur B.¹

¹ $\text{agr}(A||B) = |\{\text{annotations}(A) \text{ en accord avec annotations}(B)\}| / |\{\text{annotations}(A)\}|$.

	Agr(A B)	Agr(B A)	moyenne
Expressions d'opinion (OW)	86%	75%	80%
OW : explicites + polarité	77%	69%	73%
OW : implicites	23%	20%	21%

Figure 1 : Accords inter annotateurs obtenus

Les résultats observés pour les annotations d'expressions implicites montrent un désaccord entre les annotateurs. En effet ce qui est un simple fait pour un annotateur peut être interprété différemment pour un autre. Par exemple, pour le syntagme "*ambiance rock'n'roll*", *rock'n'roll* a été considéré soit comme expression d'opinion implicite positive (équivalente à "*bonne ambiance*"), soit comme simple fait (simple évocation du type de musique diffusée dans le bar). Ces annotations nous informent en fait beaucoup plus sur ce que perçoit l'annotateur, en fonction de ses centres d'intérêts, que sur ce que souhaitait exprimer le locuteur. À la suite de cette annotation, une version "gold standard" de l'annotation a été produite reprenant les annotations sur lesquelles les annotateurs sont tombés d'accord. Pour la suite de nos travaux, nous nous sommes donc concentrés sur les expressions d'opinion explicites, qui semblent effectivement repérables avec précision par un être humain.

3 Extension de lexique

L'objectif de la mise en place de ce modèle d'annotation est l'extraction d'opinions pour le calcul de l'opinion générale à l'échelle d'une phrase et d'un document dans des corpus issus de domaines variés. Pour mettre en œuvre ce processus, il nous faut constituer une ressource lexicale. Pour ce faire, nous proposons, dans un premier temps, de constituer une ressource pour les adjectifs à partir d'une classification générique des mots du français.

3.1 Principales méthodes d'élaboration de ressources lexicales

Pour étendre un lexique, deux principales catégories de méthodes sont utilisées dans la littérature. Les premières sont des méthodes à base de corpus qui n'utilisent que les informations présentes dans le corpus, annoté ou non, pour classer les mots. De ce fait, ces méthodes requièrent généralement un corpus assez important. (Hatzivassiloglou et McKeown, 1997) ont utilisé les conjonctions pour étendre leur lexique, partant de la supposition que deux adjectifs en conjonction (par exemple, reliés par un "et") ont une polarité similaire, et que deux mots en contraste conjonction (par exemple, reliés par un "mais") ont une polarité opposée. Turney et Littman ont proposé en 2002 et 2003 deux scores calculant la polarité d'un terme : le SO-PMI (Semantic Oriented Pointwise Mutual Information) basé sur les cooccurrences du mot avec un lexique germe polarisé, et la LSA (Latent Semantic Analysis), calculant la proximité des contextes d'utilisation de deux mots. Pour le français, Harb et al. (2008) ont utilisé les corrélations entre adjectifs et mots germes pour constituer une liste d'adjectifs d'opinion polarisés.

Les secondes méthodes, dites à base de dictionnaire, consistent à augmenter un lexique via des liens sémantiques et/ou syntaxiques, retrouvés dans des ressources linguistiques externes. (Kamps et al., 2004) ont utilisé la relation de synonymie de WordNet pour calculer une distance entre adjectifs. La distance relative avec les mots germes *good* et *bad* a permis la classification d'adjectifs entre positif et négatif. (Kim et Hovy, 2005) ont quant à eux directement étendu leurs lexiques à partir des synonymes et antonymes de ces mots germes. (Esuli et Sebastiani, 2006) ont amélioré ces méthodes en effectuant au préalable une désambiguïsation syntaxique. Les phrases de description des *synsets* de WordNet ont

également permis de classer ces derniers entre positif, négatif, et neutre, soit en testant la présence de mots germes dans les description (Andreevskaia et Bergler, 2006), soit en comparant les descriptions entre elles (Esuli et Sebastiani, 2006). Enfin, (Chesley et al., 2006) ont classé les adjectifs suivant le nombre d'occurrences de mots germes dans leur page de description du Wiktionary.

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

3.2 Ressource utilisée : Taxinomie des concepts Synapse Développement

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

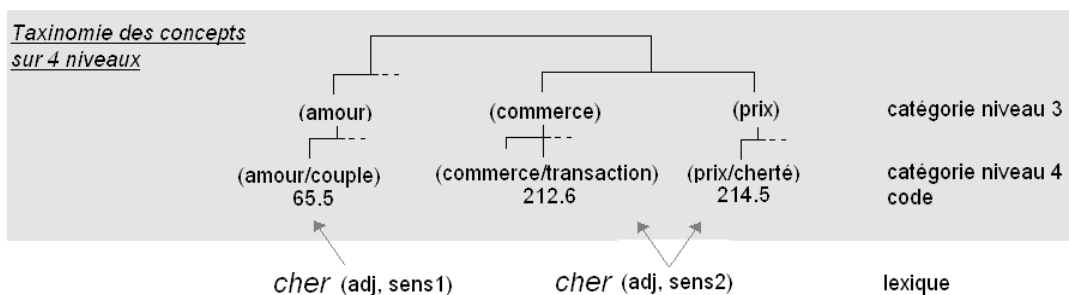


Figure 2 : Exemple de représentation des 2 sens de l'adjectif "cher".

L'adjectif "cher" est ici modélisé comme ayant 2 sens. Le sens 1 est défini par un lien vers un concept (65.5/amour/couple). Cette taxinomie se rapproche des *synsets* de WordNet, à la différence qu'ici le nombre de catégories conceptuelles est borné (3387 catégories au niveau 4, 256 au niveau 3), et que les mots sont définis par sens et non par classes de regroupements.

Nous avons donc étiqueté tous les corpus avec l'analyseur syntaxique de Synapse Développement (Laurent et al., 2009). À chaque mot est associé son lemme, sa catégorie syntaxique et le(s) numéro(s) de sens reconnu(s) lors de la désambiguïsation sémantique locale. Par extension, un ensemble de catégories conceptuelles de niveau 4 est donc associé à chaque mot. Nous avons observé que les catégories où l'on trouve des adjectifs d'opinion rassemblent des mots de sens proches et de même polarité. Elles regroupent cependant des mots d'intensité différentes. L'intensité des adjectifs d'opinion n'a donc pas été prise en compte lors de l'extension du lexique présentée ci-après.

3.3 Algorithme d'extension du lexique

3.3.1 Constitution du lexique étendu

À partir d'un petit nombre de documents annotés manuellement, on extrait une liste lemmatisée d'adjectifs germes d'opinion et leur polarité (positif/négatif). Cette étape est critique, au sens où un ensemble de mots germes trop restreint ou déséquilibré nuira aux résultats.

Initialisation : On crée ensuite un graphe sans arêtes dont les sommets sont les concepts de niveau 4 de la taxinomie. Pour chacun des mots du lexique germe extrait, on ajoute un sommet au graphe, auquel on associe la polarité p du mot. Si à ce mot sont associés un ou plusieurs concepts, on relie ce nouveau

sommet aux sommets associés à ces concepts. On obtient ainsi un graphe bipartite reliant les nœuds de type "taxinomie" via les nœuds de type "mot", que l'on peut nettoyer en supprimant les nœuds esseulés.

Simplification : L'objectif est de construire un arbre de décision à partir des catégories de taxinomie des adjectifs. Plus formellement, on souhaite affecter à chaque nœud "taxinomie" une polarité et une hauteur dans l'arbre telles que :

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

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

hauteur(t) est donc ici l'inverse de la priorité avec laquelle la catégorie est prise en compte : plus celle-ci est élevée, moins la catégorie est considérée comme discriminante pour déterminer la polarité.

Le graphe obtenu à l'étape d'initialisation se rapproche d'un graphe de CSP (Constraint Satisfaction Problem), où les nœuds "mots" seraient les contraintes à satisfaire. On applique donc un algorithme de résolution assez proche de la cohérence d'arcs. À chaque étape *h*, on affecte à tous les nœuds "taxinomie" cohérents, c'est-à-dire reliés uniquement à des mots de même polarité *p*, la hauteur *h* et la polarité *p*. On élimine ensuite tous les nœuds affectés, c'est-à-dire les nœuds "taxinomie" dont la valeur est connue, et les nœuds "mots" reliés à un nœud "taxinomie" affecté. Dans l'éventualité où aucun nœud ne serait affecté, on affecte un nœud quelconque à une polarité arbitraire (dans l'idéal minimisant le nombre de nœuds "mots" mal affectés) et à la hauteur *h*.

Construction de l'arbre de décision / classifieur du lexique : On obtient ainsi une liste de nœuds correspondant aux catégories de la taxinomie, chacun possédant une hauteur *h* et une polarité *p*. En associant à chaque nœud une branche de hauteur *h* menant sur une décision d'affectation de la valeur *p*, on obtient un arbre de décision permettant de classer les mots de la catégorie. Cet arbre de décision permet de générer un lexique visualisable en projetant les décisions sur tous les termes répertoriés dans la catégorie. En pratique, l'arbre seul suffit pour déterminer la polarité des adjectifs considérés.

La figure 3 présente un exemple de résolution de graphe possédant une ambiguïté sur la catégorie 12.5 (reliée à un mot positif et un mot négatif) : on ne peut résoudre ce conflit et affecter une polarité à la catégorie qu'à l'étape 2, aussi cette catégorie se retrouve-t-elle dans une branche de l'arbre de hauteur 2.

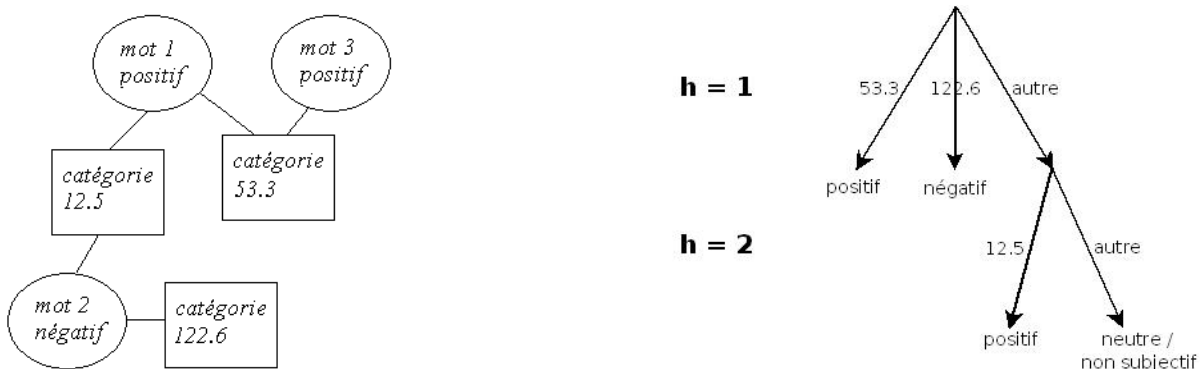


Figure 3 : Exemple de graphe simple et d'arbre de décision associé. Les étiquettes des branches de l'arbre s'interprètent comme : "si le terme rencontré appartient à la catégorie ...".

3.3.2 Renforcement du lexique

Afin de permettre une extension plus importante du lexique dans le cas où peu de mots germes seraient utilisés (c'est-à-dire si peu de textes annotés sont disponibles), nous avons rajouté une phase optionnelle

dite de renforcement du lexique. Deux catégories de la taxinomie sont jugées "proches" si elles comportent au moins un mot désambiguïsé en commun. Par exemple, sur la figure 2, prix/cherté est proche de commerce/transaction. À l'aide de cette relation de proximité, on peut répercuter la polarité des catégories positives ou négatives aux catégories proches, si cela n'engendre pas de conflit.

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

Nous nous sommes limité à une seule itération de cette méthode. Plus d'itérations induisaient une propagation de polarité à un nombre très élevé de catégories, ce qui se serait traduit par une chute en terme de précision non compensée par un gain substantiel en rappel.

4 Évaluations

Les évaluations se sont déroulées en trois temps. Tout d'abord, la couverture du lexique généré, ensuite, l'évaluation de sa généralité et enfin, son utilisabilité pour une tâche extrinsèque ont été successivement traitées.

4.1 Couverture du lexique

Le corpus de critiques de restaurants et bars manuellement annoté suivant le schéma présenté en section 2 a été utilisé pour cette évaluation. Ce corpus a été séparé en deux groupes : un corpus d'entraînement constitué de 5 commentaires positifs et 5 négatifs, choisis préalablement pour leurs polarités tranchées, et un corpus de test, constitué des 50 commentaires restants. La liste de mots germes était constituées de 15 mots positifs, et de 11 négatifs. En comparant les adjectifs reconnus avec ceux annotés manuellement, nous obtenons les résultats suivants :

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

Figure 4 : Couverture du lexique par rapport au lexique gold

Avant renforcement, 62% des mots reconnus n'existaient pas dans le lexique germe. Les résultats sans renforcement du lexique montrent une couverture relativement faible. En effet, les catégories étant assez fines, certains traits sémantiques approchants sont divisés en plusieurs catégories ; par exemple, les catégories "beauté" et "beauté humaine" sont distinctes. Par conséquent, des mots proches peuvent ne pas être reconnus comme tel. La phase de renforcement du lexique permet de corriger ce défaut : les deux catégories précédentes possèdent au moins un terme en commun ("*joli*"), et le caractère positif des adjectifs de l'une peut donc se propager à l'autre.

La précision est assez bonne dans le premier cas. Les faux positifs relevés avant renforcement du lexique sont principalement dus à deux phénomènes. Certains adjectifs sont ambigus hors contexte. Par exemple, "digne" utilisé seul est positif ("les tapas sont dignes"), mais utilisé dans l'expression "digne de X", sa polarité dépend de celle de X ("Le poulet est digne d'un mauvais fast-food."). D'autre part, certains adjectifs peuvent être employés, suivant le contexte, comme opérateurs d'intensité ou comme expressions d'opinion ("un super resto" vs. "un resto super bon", dans le second exemple, super est un opérateur intensifiant l'expression positive "bon").

4.2 Généricité du lexique

Le lexique ainsi obtenu a été généré à partir de mots germes issus d'un corpus spécialisé, et étendu via une ressource générique, indépendante de tout domaine. On peut donc se demander dans quelle mesure ce lexique est pertinent sur de nouveaux domaines spécialisés. Pour cela, dix commentaires issus d'un domaine différent ont été annotés afin de générer un second jeu de mots germes. Il s'agit de commentaires d'internautes déposés sur le site <http://www.allocine.fr/>, de taille similaire à ceux de Qype. Nous avons généré un second lexique à partir de ces mots germes, et comparé ce nouveau lexique à celui généré à la section précédente.

Les deux lexiques se sont avérés assez différents. En effet, le lexique des films comportait beaucoup plus de termes d'émotions fortes positives ("*poignant*") tandis que l'autre comportait des termes du registre culinaire ("*délicieux*") et des adjectifs négatifs de prix ("*cher*", "*élevé*"). Statistiquement, ceci s'est traduit par une chute de moitié du rappel lors de l'annotation d'un texte du corpus de restaurants avec ce lexique, la plupart des termes du lexique germe menant sur des catégories quasi absentes du corpus traité. Ceci montre que le lexique généré est plutôt spécifique au corpus, les catégories identifiées comme positives ou négatives dépendant des mots germes identifiés.

4.3 Utilisabilité du lexique

Les commentaires du corpus de restaurants sont associés à des étoiles (1 à 5). On peut par conséquent inférer la polarité que souhaitait exprimer le rédacteur à partir de celles-ci : positive (4 ou 5 étoiles), négative (1 ou 2 étoiles), mitigée (3 étoiles). Nous avons tenté de retrouver cette polarité pour les documents de test, en comptant simplement les adjectifs d'opinion : une majorité de positifs (resp. négatifs) signifiant un commentaire positif (resp. négatif), et une absence de majorité un commentaire neutre. Nous avons testé la difficulté de la tâche en calculant les scores obtenus avec la liste de mots germes seule, ainsi qu'avec la liste complète des adjectifs annotés pour obtenir un score maximal atteignable de 83%.

	Proportion de polarité reconnue
Mots germes	45%
Liste étendue sans renforcement	65%
Liste complète des adjectifs annotés	83%

Figure 5 : Proportion de polarité de documents reconnus

Les résultats montrent une amélioration de 20% de bonne classification avec l'extension du lexique. À noter que le résultat maximal atteignable en considérant uniquement les adjectifs est relativement loin de 100%. Ceci est principalement dû à deux facteurs. Tout d'abord, certains internautes se servent du commentaire uniquement pour compléter l'opinion véhiculée par les étoiles. De plus, notre algorithme de classification ne prend en compte ni les négations, ni le fait que le sujet de certains adjectifs d'opinion est hors sujet par rapport au restaurant jugé.

5 Conclusion et perspectives

Afin de mettre en place une extraction automatique d'opinions, nous avons proposé un schéma d'annotation fin des expressions d'opinion, et une méthode pour élaborer une ressource lexicale d'adjectifs d'opinion à partir de la taxinomie de concepts. Le modèle d'annotation proposé, indépendant de la langue et du domaine, inclut tous les éléments qui nous ont parus intervenir dans le calcul de

l'opinion, aussi bien au niveau local (émetteur, sujet, opérateurs) qu'au niveau du document (relations discursives). L'algorithme d'extension de lexique proposé repose sur une taxinomie fine des mots du français, une désambiguïisation sémantique et une lemmatisation effectuées par un analyseur syntaxique, et un jeu de mots germes. Évalué en trois étapes, il présente une couverture assez faible, améliorable dans une certaine mesure via une itération d'extension par proximité, au détriment de la précision. Les résultats appellent cependant à être confirmés sur une plus grande quantité de données annotées.

Afin d'améliorer cet algorithme, l'utilisation d'un dictionnaire de synonymes et antonymes pondérés est envisagée. Celui-ci pourrait venir en surcouche pour étendre la liste des adjectifs générés, ou en prétraitement pour étendre la liste des mots germes. Il pourrait également permettre d'introduire de classifier automatiquement les adjectifs par intensité. Après les adjectifs, d'autres catégories syntaxiques, comme les noms et les verbes, seront traitées. Les liens entre catégories syntaxiques (par exemple déverbaux) seront également envisagés suivant la même approche.

Références

- ANDREEVSKAIA A., BERGLER D., (2006). Mining WordNet for fuzzy sentiment: Sentiment tag extraction from WordNet glosses. Actes de *EACL 2006*.
- ASHER N., BENAMARA F., MATHIEU Y.Y. (2009). Appraisal of Opinion Expressions in Discourse. *Linguisticæ Investigationes*, 32:2.
- ASHER N., LASCARIDES A. (2003). *Logics of Conversation*. Cambridge University Press.
- CHESLEY P., VINCENT B., XU L., SRIHARI R.K. (2006). Using verbs and adjectives to automatically classify blog sentiment. *Training* 580.
- ESULI A., SEBASTIANI F. (2006). Determining term subjectivity and term orientation for opinion mining. Actes de *EACL 2006*, 193-200.
- ESULI A., SEBASTIANI F. (2006). SentiWordNet: A publicly available lexical resource for opinion mining. Actes de *LREC 2006*.
- HARB A., DRAY G., PLANTIE M., PONCELET P., ROCHE M., TROUSSET F. (2008). Apprenons les bons adjectifs!. Actes de *Atelier FODOP 2008*.
- HATZIVASSILOGLOU V., MCKEOWN K.R. (1997). Predicting the semantic orientation of adjectives. Actes de *EACL 1997*.
- KAMPS J., MARX M., MOKKEN R.J., DE RIJKE M. (2004). Using WordNet to measure semantic orientation of adjectives. Actes de *LREC 2004*, 1115-1118.
- KIM S.-M., HOVY E. (2005). Automatic Detection of Opinion Bearing Words and Sentences. Proceedings of *IJCNLP 2005*.
- LAURENT D., NEGRE S., SEGUELA P. (2009). L'analyseur syntaxique Cordial dans Passage. *TALN 2009*.
- MARTIN J.R., WHITE P.R.R. (2005). *The Language of Evaluation: Appraisal in English*. Basingstoke : Palgrave Macmillan.
- QUIRK R., GREENBAUM S., LEECH G., SVARTVIK J. (1985). *A Comprehensive Grammar of the English Language*. Longman.
- TURNER P.D., LITTMAN M.L. (2002). Unsupervised learning of semantic orientation from a hundred-billion-word corpus. Arxiv preprint.
- TURNER P.D., LITTMAN M.L. (2003). Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems* 21, 315-346.
- WIEBE J., WILSON T., CARDIE C. (2005). Annotating Expressions of Opinions and Emotions in Language. *Language Resources and Evaluation* 39, 165-210.
- WHITELAW C., GARG N., ARGAMON S. (2005). Using appraisal groups for sentiment analysis. *Proc. of the 14th ACM international conference on Information and knowledge management*, 625-631.