

Catégorisation sémantico-discursive des évaluations exprimées dans la blogosphère

Matthieu Vernier¹ Laura Monceaux¹ Béatrice Daille¹ Estelle Dubreil¹

(1) LINA / CNRS UMR 6241, Université de Nantes

{Prenom.Nom}@univ-nantes.fr

Résumé. Les blogs constituent un support d’observations idéal pour des applications liées à la fouille d’opinion. Toutefois, ils imposent de nouvelles problématiques et de nouveaux défis au regard des méthodes traditionnelles du domaine. De ce fait, nous proposons une méthode automatique pour la détection et la catégorisation des évaluations localement exprimées dans un corpus de blogs multi-domaine. Celle-ci rend compte des spécificités du langage évaluatif décrites dans deux théories linguistiques. L’outil développé au sein de la plateforme UIMA vise d’une part à construire automatiquement une grammaire du langage évaluatif, et d’autre part à utiliser cette grammaire pour la détection et la catégorisation des passages évaluatifs d’un texte. La catégorisation traite en particulier l’aspect axiologique de l’évaluation, sa configuration d’énonciation et sa modalité dans le discours.

Abstract. Blogs are an ideal observation for applications related to the opinion mining task. However, they impose new problems and new challenges in this field. Therefore, we propose a method for automatic detection and classification of appraisal locally expressed in a multi-domain blogs corpus. It reflects the specific aspects of appraisal language described in two linguistic theories. The tool developed within the UIMA platform aims both to automatically build a grammar of the appraisal language, and the other part to use this grammar for the detection and categorization of evaluative segments in a text. Categorization especially deals with axiological aspect of an evaluative segments, enunciative configuration and its attitude in discourse.

Mots-clés : fouille d’opinion, langage évaluatif, catégorisation des évaluations.

Keywords: opinion mining, appraisal language, appraisal classification.

1 Introduction

Ces dernières années, les blogs ont conquis leur place à côté des médias traditionnels et deviennent une source d’informations incontournable. Les blogueurs l’utilisent majoritairement à des fins d’auto-représentation, et la plupart se forment autour des affects ou des idées propres à leur auteur. Le blog est souvent construit autour d’une stratégie argumentative où l’auteur cherche à convaincre, plus ou moins intensément, ses lecteurs d’adopter son point de vue ou sa façon d’évaluer le monde. De par la force et la rapidité des échanges sur le Web, un blog peut devenir célèbre au sein de la communauté à laquelle il appartient en très peu de temps. Certains blogs sont d’ailleurs si influents que les informations qui en sont issues semblent se propager aux autres blogs pour créer des phénomènes de *buzz* ou de sujets émergents dont les évolutions peuvent être observées au fil des jours ou des semaines.

Dans le cadre de la blogosphère, les recherches en fouille d'opinion suscitent un intérêt majeur, notamment afin de caractériser la façon dont l'un de ces sujets émergents est évalué : est-il évalué globalement positivement ou négativement ? Sur quels aspects est-il positivement ou négativement évalué ? Selon les sujets émergents, les domaines d'applications sont particulièrement nombreux : la sociologie, la politique, le marketing, la création de réseaux sociaux autour d'opinions communes, etc. La quantité des blogs, leur qualité grandissante et leur réactivité vis à vis de l'actualité mondiale suggèrent la possibilité d'analyser en temps réel l'évolution des avis portés sur un même sujet cible. L'enjeu pour la recherche en traitement de la langue et un des verrous technologiques de cette problématique résident principalement sur la façon de détecter et catégoriser automatiquement l'expression d'une opinion dans la langue. D'un point de vue terminologique, nous préférons adopter le terme d'évaluation et de langage évaluatif pour faire référence aux différentes modalités linguistiques possibles pour exprimer l'opinion, l'appréciation, le jugement, l'émotion, l'accord, le désaccord porté sur un sujet. Nous justifions cet emploi en adoptant le point de vue des théories linguistiques anglo-saxonnes (Martin & White, 2005) et françaises (Charaudeau, 1992; Galatanu, 2000).

Les billets publiés sur les blogs par les auteurs, et les commentaires qui y sont associés, sont particulièrement hétérogènes à la fois dans la diversité des thèmes abordés et dans leur forme discursive. Un billet peut parler d'un seul concept ou, dans la majorité des cas, évoquer plusieurs concepts bien distincts. L'auteur peut être présent ou absent dans l'énonciation (utilisation du pronom « je ») ; sa stratégie argumentative peut chercher à créer une atmosphère volontairement subjective ou, au contraire, créer une fausse impression d'objectivité. Notre objectif est d'aboutir à une méthode pour la fouille d'opinion qui tienne compte de la diversité possible des thématiques et des formes discursives des textes. Nous cherchons ainsi à détecter les évaluations exprimées localement dans les blogs et à catégoriser leur modalité, leur configuration d'énonciation et leur axiologie (positive ou négative).

Dans cet article, nous présentons l'évolution des problématiques du domaine afin de préciser en quoi les méthodes existantes en fouille d'opinion ne sont pas précisément adaptées à nos objectifs (section 2). Pour caractériser du mieux possible le phénomène complexe de l'évaluation, nous nous appuyons sur deux modèles linguistiques pour proposer un schéma d'annotation (section 3). A partir d'un corpus annoté manuellement, notre approche consiste à apprendre automatiquement une grammaire de l'évaluation et à utiliser celle-ci a posteriori pour détecter et catégoriser la modalité, la configuration d'énonciation et l'axiologie des évaluations issues des blogs (section 4). Nous expérimentons et évaluons notre méthode sur un corpus test à travers deux exemples d'applications (section 5).

2 Évolution des problématiques en fouille d'opinion

L'évolution des travaux en fouille d'opinion depuis une dizaine d'années semble guidée par deux courants : comment adapter des méthodes pour traiter du monodomaine à du multidomaine sans repasser par une phase d'entraînement coûteuse ? comment adapter des méthodes qui analysent un texte dans sa globalité vers des méthodes qui analysent séparément différents passages d'un texte ? Ces deux axes sont également deux de nos préoccupations induites par notre support d'observations : les blogs.

2.1 Du mono-thématique au multi-thématique

Quelques travaux en fouille d'opinion prenant les blogs comme support d'étude (Mishne & Glance, 2006; Mullen & Malouf, 2006) se sont limités à analyser les blogs d'une même thématique : le cinéma ou la politique notamment. Par opposition, notre objectif est de prendre en compte le plus large éventail possible de domaines discutés sur les blogs. Ce point soulève une problématique qui prend sa source dans les travaux pionniers de la fouille d'opinion. De nombreux travaux proposent des méthodes par apprentissage supervisé pour classer automatiquement des textes à partir d'un corpus homogène thématiquement : des critiques de films, de livres, de téléphones, d'appareil photos ou de voyages (Pang *et al.*, 2002; Dave *et al.*, 2003; Maurel *et al.*, 2008). Pour chaque domaine d'application, ces méthodes bénéficient de l'existence de ressources exploitables : les sites de critiques en ligne où chaque texte est associé à une note attribuée en amont par les utilisateurs du site. Ces méthodes permettent d'extraire statistiquement des unités textuelles (mots, n-grammes) associées à une polarité positive ou négative selon leur fréquence d'apparition dans les textes notés positivement ou négativement. Souvent ces indices textuels ne font sens que dans le seul domaine étudié, voire uniquement dans le corpus d'étude. Ainsi, (Aue & Gamon, 2005) montre que le terme *le plastique* est négatif lorsqu'on évalue une *cuisine* et qu'il n'a aucune polarité selon les classifieurs entraînés sur d'autres domaines.

Des travaux plus récents (Aue & Gamon, 2005; Blitzer *et al.*, 2007) proposent des méthodes pour comparer les unités textuelles extraites statistiquement sur plusieurs corpus mono-thématiques et conserver uniquement celles qui font sens dans plusieurs domaines. Ces unités textuelles sont ainsi considérées comme suffisamment génériques pour pouvoir s'adapter à des textes d'un domaine où il n'existe pas de corpus d'entraînement disponible. Toutefois, à notre connaissance très peu d'expériences ont été menées à grande échelle sur des corpus réellement multi-thématiques. Le besoin d'envisager d'autres méthodes en supplément des classifieurs supervisés classiques s'exprime de plus en plus avec les nouveaux besoins applicatifs en fouille d'opinion.

2.2 De la catégorisation de texte à la catégorisation des évaluations

Un deuxième aspect qui distingue notre objectif des problématiques habituelles en fouille d'opinion réside dans le constat suivant : un billet ou un commentaire publié sur un blog parle, dans la majorité des cas, de plusieurs concepts bien distincts. Chaque concept peut être évalué différemment. Il n'est donc pas pertinent de chercher à catégoriser positivement ou négativement un blog dans sa globalité. D'autres travaux font ce même constat et envisagent différentes stratégies. (Nigam & Hurst, 2006) commence par extraire les groupes de phrases d'un corpus qui traitent d'une même thématique spécifiée en amont pour ainsi se replacer dans les conditions d'une classification mono-thématique. Selon le même principe, mais avec un niveau de granularité plus fin, (Hu & Liu, 2004) catégorise les phrases une par une à partir des adjectifs présents dans la phrase. Citons également (Whitelaw *et al.*, 2005) qui s'intéresse à caractériser les évaluations à un niveau intra-phrastique (*very good, not terribly funny*) à partir de la théorie de l'évaluation anglo-saxonne (Martin & White, 2005). Nous nous inscrivons davantage dans ce dernier axe de recherche en nous questionnant sur la définition d'une évaluation, quelles sont ses différentes formes et ses différents marqueurs linguistiques qui permettraient d'aboutir à une méthode robuste pour la détection et la catégorisation des évaluations.

3 Langage évaluatif

3.1 Théories linguistiques

L'évaluation est définie par (Lavelle, 1950) comme *l'acte de rupture de l'indifférence par laquelle nous mettons toutes les choses sur le même plan et considérons toutes les actions comme équivalentes*. Tout acte de langage révélant une rupture d'indifférence relève donc du phénomène évaluatif. Ces actes mettent en jeu des mécanismes sémantiques, pragmatiques ou énonciatifs complexes faisant l'objet de nombreuses études (Kerbrat-Orecchioni, 1997; Anscombe & Ducrot, 1983). (Charaudeau, 1992) montre qu'il existe cinq modalités permettant à un locuteur d'exprimer une évaluation (**l'opinion, l'accord ou le désaccord, l'acceptation ou le refus, le jugement et l'appréciation**). Chacune de ces modalités révèle une attitude particulière du locuteur : sa croyance plus ou moins certaine par rapport à l'évaluation qu'il exprime, le champ d'expérience dans lequel il se positionne (éthique, moral, intellectuel, esthétique, etc), sa position par rapport à son énoncé (présence ou absence du « je »). Selon Charaudeau, il existe des marqueurs lexicaux et des structures linguistiques spécifiques à ces modalités (Tab. 1).

Marqueurs	Modalité
être sceptique, douter, croire, penser, être convaincu	Opinion (conviction faible à forte)
effectivement, d'accord, non, être faux	Accord ou Désaccord
courageux, intéressant, lâche, mentir	Jugement (éthique, moral, intellectuel)
adorer, haïr, joli, triste	Appréciation (affect, hédonique)

TAB. 1 – Exemples de marqueurs lexicaux pour les modalités de l'évaluation

La théorie de (Galatanu, 2000) sur l'évaluation complète le modèle de Charaudeau en hiérarchisant les modalités sur une échelle de subjectivité. Lorsqu'un locuteur organise son énoncé, il peut choisir d'objectiver ou de subjectiver son discours en activant certaines modalités ou par la configuration énonciative des modalités qu'il active. Dans les exemples Tab. 2, la valeur mise en jeu « mentir » (modalité de jugement) intervient dans une stratégie argumentative différente. Le locuteur dissimule sa présence (configuration implicite) et éventuellement sur-modalise l'évaluation en marquant qu'il s'agit d'une opinion ou d'une appréciation. Ces exemples ont un impact différent sur l'effet d'objectivité ou de subjectivité de l'énoncé.

Exemple	Sur-modalité	Modalité
<i>Je doute qu'il <u>mente</u></i>	Opinion faible explicite	Jugement implicite
<i>Il est évident qu'il <u>ment</u></i>	Opinion forte implicite	Jugement implicite
<i>Oui, c'est un <u>menteur</u></i>	Accord	Jugement implicite
<i>Il <u>ment</u></i>		Jugement implicite
<i>Je n'aime pas qu'il <u>mente</u></i>	Appréciation explicite	Jugement implicite

TAB. 2 – Exemple de discours évaluatif différent pour la même valeur *mentir*

3.2 Langage évaluatif sur les blogs

À des fins d'apprentissage et de test, nous annotons les évaluations exprimées dans un corpus de blogs à partir d'un schéma d'annotations en adéquation avec les théories linguistiques présentées. Pour chaque évaluation annotée, nous précisons sa modalité, sa configuration d'énonciation, son axiologie et le concept évalué. Nous renvoyons à (Dubreil *et al.*, 2008) pour une description plus précise de la méthodologie d'annotation. Le **corpus d'entraînement** annoté est composé de 200 billets de blogs associés aux commentaires postés par les lecteurs sur ce

billet. Ils sont extraits automatiquement de la plateforme de blogs OverBlog et appartiennent volontairement à des thématiques les plus variées possibles (actualité, artiste, famille, gastronomie, internet, santé, science, voyage, etc), en excluant les billets qui ne contiennent pas ou très peu de texte. 4945 passages évaluatifs ont ainsi été annotés manuellement. Par ailleurs, nous constituons semi-manuellement trois **ressources lexico-sémantiques** :

- un **lexique de l'évaluation** (1115 entrées), développé par Sinequa (Stern, 2008), contenant les termes évaluatifs fréquents dans le corpus, associées à leur catégorie grammaticale, leur modalité, leur énonciation et leur axiologie. ex : *machiste, chapeau bas, douter*,
- un **lexique de l'intensité** (21 entrées) ex : *particulièrement, très*,
- un **lexique de la négation** (15 entrées) ex : *pas, aucun*,

Le corpus d'entraînement et les ressources lexicales sont à la base de notre méthode pour construire automatiquement une grammaire du langage évaluatif.

4 Méthode symbolique pour catégoriser les évaluations

Notre approche consiste à apprendre automatiquement les structures du langage qui sont spécifiquement utilisées pour évaluer et utiliser la généricité des structures apprises pour détecter et catégoriser les évaluations dans de nouveaux textes. Nous partons de l'idée de Chauradeau qu'il existe des structures lexicales, grammaticales et sémantiques pour chaque modalité d'évaluation. Par exemple, *j'en doute* et *nous en sommes persuadés* sont des évaluations de modalité d'opinion généralisables car : *douter* et *être persuadé* ont la même classe sémantique (*verbes marqueurs d'opinion*), *je* et *nous* ont la même fonction grammaticale et impliquent *explicitement* le locuteur.

4.1 Apprentissage automatique des structures évaluatives

Plateforme pour le TAL Nous tirons profit de la plateforme pour le TAL *UIMA*¹ pour construire une chaîne de traitements linguistiques. Cette plateforme permet de traiter des textes non structurés et d'y ajouter des annotations dans un format normalisé assurant la réutilisabilité des composants et l'échange des annotations entre composants.

Pré-traitements L'apprentissage des structures évaluatives (Fig. 1) est réalisé à partir du corpus d'entraînement annoté manuellement, présenté dans la section précédente. Les composants de pré-traitements annotent des informations morpho-syntaxiques et sémantiques à partir de ressources extérieures (TreeTagger et lexiques). Ces annotations portent sur des mots ou des suites de mots, nous utilisons le terme abstrait *symbole* pour nommer ce niveau de granularité. Dès lors, le flux de données qui transite entre composants est constitué d'une suite de symboles, chacun associé à des traits :

- **lexico-grammaticaux** : forme, lemme (lem), catégorie grammaticale (pos),
- **sémantiques** : type (*évaluation* (eval), *intensité* (int), *négation* (neg), *autre* (mot)), modalité (*appréciation* (app), *opinion* (op), *accord-désaccord* (acc)), configuration d'énonciation (*exclamative* (excl), *explicite* (exp), *implicite* (imp)), axiologie (*positif*, *négatif*, *ambigu*).

¹Unstructured Information Management Architecture : développée par IBM et Apache. <http://incubator.apache.org/uima/>.

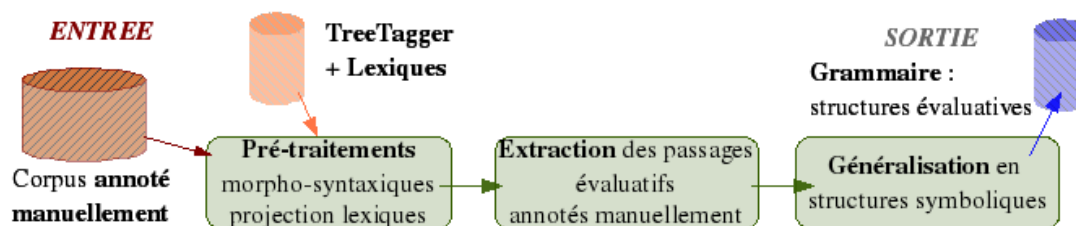
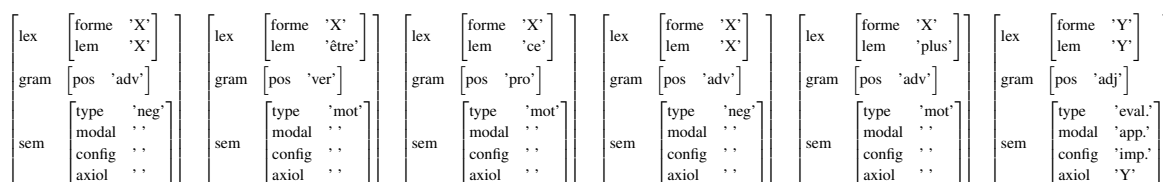


FIG. 1 – Chaîne de traitements pour l'apprentissage de structures évaluatives

Chaînes symboliques et généralisation À partir des 4945 passages évaluatifs annotés manuellement dans le corpus, et des chaînes symboliques correspondantes, il s'agit de gagner en généralité en spécifiant automatiquement les valeurs de trait qui peuvent être substituées par une autre valeur², ceci afin de repérer plus d'évaluations. La figure 2 représente la chaîne de symboles génériques construite à partir de l'exemple d'évaluation *n'est-ce pas plus original*.

FIG. 2 – Structure générique extraite à partir de l'exemple *n'est-ce pas plus original*.

Nos règles de généralisation sont les suivantes :

- Généralisation de la valeur des traits **axiol**, **forme** et de **lemme** (Y sur la fig. 2) pour tous les symboles de **type évaluation** et de **modalité appréciation**,
- Généralisation de la valeur du trait **lex** (X sur la fig. 2) pour certains symboles (adverbe, pronom ...) ainsi que le trait **lem** pour les symboles de type adverbe
- Ajout de l'opérateur standard * sur les symboles de **type intensité** et généralisation de la valeur des traits **forme** et de **lemme** de ces symboles,
- Ajout de l'opérateur standard + (une ou plusieurs fois) pour les symboles de **config explicite** et de **pos pronom** et généralisation de la valeur des traits **forme** et de **lemme** de ces symboles.

Ce processus de création de structure générique permet de distinguer des tournures évaluatives très proches mais dont la signification peut être radicalement différente : *n'est-ce pas plus original* est positif, *n'est pas plus original* est négatif et *ne semble pas plus original* est négatif mais avec une opinion de conviction moyenne. Le processus de généralisation permet également de regrouper certaines évaluations ayant la même structure évaluative (*c'est génial*, *c'est super*) d'où une réduction des 4945 structures annotées manuellement à 2830 structures apprises dans notre grammaire.

Grammaire du langage évaluatif Nous stockons chaque structure évaluative générique ainsi extraite dans une ressource au format XML. Pour chacune d'elle, le dernier composant de la chaîne ajoute des méta-données qui serviront lors de la catégorisation : le nombre d'occurrences de la structure dans le corpus d'entraînement, le nombre d'occurrences par modalité, le nombre d'occurrences par configuration d'énonciation et la tournure de la structure. Nous entendons par tournure, le fait que certaines structures peuvent être :

²Traits ayant pour valeur '?' dans Fig.2

- **directes** : *pour notre plus grand plaisir, pour notre plus grand malheur,*
- **inversives** : *loin d'être génial, loin d'être mauvais,*
- **figées positives** : *faire taire la critique, marcher nickel,*
- **figées négatives** : *équipe de bras cassé, c'est là où le bât blesse.*

4.2 Catégorisation sémantico-discursive des évaluations

Il s'agit ensuite d'utiliser la grammaire du langage évaluatif générée pour détecter et catégoriser les évaluations dans de nouveaux textes.

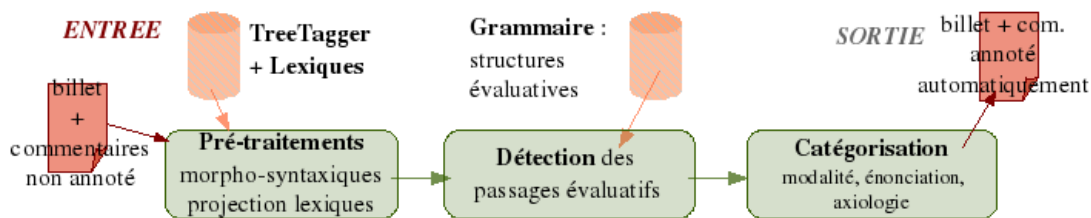


FIG. 3 – Chaîne de traitements pour la détection et la catégorisation de structures évaluatives

Détection La tâche de détection consiste à annoter les segments évaluatifs sans les catégoriser. Nous considérons que ces segments sont d'un niveau intra-phrastique. Les phrases d'un document sont donc traitées une par une. La stratégie du composant de détection consiste à :

- extraire les symboles de la phrase (et leurs traits) pour constituer une chaîne symbolique,
- tester l'unification des sous-chaînes symboliques avec les structures apprises, en commençant par les sous-chaînes les plus longues possibles et par le début de la phrase,
- créer une annotation lorsqu'une chaîne s'unifie avec une structure (S_i) de la grammaire.

Nous calculons un coefficient de confiance pour cette tâche à l'aide des méta-données de S_i contenues dans la grammaire.

$$\alpha_{detection}(S_i) = \frac{Eval(S_i)}{Eval(S_i) + NonEval(S_i)}$$

$Eval(S_i)$: nombre d'occurrences de S_i annotées dans le corpus d'entraînement

$NonEval(S_i)$: nombre d'occurrences de S_i non annotées dans le corpus d'entraînement

Catégorisation La catégorisation consiste à déterminer la modalité, la configuration énonciative et l'axiologie de l'évaluation détectée préalablement. Les méta-données de S_i stockées dans la grammaire permet de calculer la meilleure probabilité pour la modalité et l'énonciation :

$$\alpha_{catModalite}(S_i, M_j) = \frac{Eval(S_i, M_j)}{Eval(S_i, M_j) + NonEval(S_i, M_j)}$$

$$\alpha_{catConfigEnon}(S_i, appreciation, CE_j) = \frac{Eval(S_i, appreciation, CE_j)}{Eval(S_i, appreciation, CE_j) + NonEval(S_i, appreciation, CE_j)}$$

$Eval(S_i, M_j)$: nombre d'occurrences de S_i de modalité M_j lors de d'entraînement

$NonEval(S_i, M_j)$: nombre d'occurrences de S_i d'une autre modalité lors de l'entraînement

$Eval(S_i, appreciation, CE_j)$: nombre d'occurrences de S_i de type appréciation ayant une configuration d'énonciation CE_j lors de d'entraînement

La *tournure* de S_i est donnée dans la grammaire, le composant détermine alors l’axiologie de l’instance de S_i rencontrée :

- structure **directe** : la polarité axiologique est identique à celle du symbole axiologisé,
- structure **inverse** : la polarité axiologique est inverse à celle du symbole axiologisé,
- structure **figée** : la polarité axiologique est indiquée dans les méta-données de la structure.

Par exemple, *Pas la plus belle* est inversive, *belle* est positif donc l’évaluation est négative.

5 Expériences et résultats

Corpus Test Afin d’évaluer la détection et la catégorisation des structures évaluatives, un corpus test a été élaboré et annoté en se focalisant essentiellement sur les modalités d’opinion, d’accord-désaccord et d’appréciation³. Les billets de ce corpus test ont été extraits à partir de l’utilisation des mots clés suivants : Sarah Palin (25 billets) et Sushi (25 billets). Comme pour le corpus d’entraînement, le corpus a été annoté manuellement par un linguiste avec les mêmes contraintes : on dénote 955 instances d’évaluation. Dans la suite, nous allons évaluer chaque composant de notre outil de détection et catégorisation des évaluations.

Détection des évaluations Dans un premier temps, nous évaluons l’étape de détection des évaluations sur le corpus test, sachant que nous n’avons aucune connaissance lexicale en amont sur ce corpus. Nous considérons une évaluation comme correctement détectée si l’évaluation est correctement délimitée (même délimitation manuelle) ou si l’une des deux bornes est erronée à un ou deux mots d’écart par rapport à l’annotation manuelle (l’accord inter-annotateur sur les bornes n’est pas représentatif). Les structures symboliques apprises semblent être des indicateurs assez précis (88.4 % de précision) pour détecter les évaluations. Cependant de manière prévisible, le rappel (50,1 %) chute fortement. Cela peut s’expliquer par la grande variété orthographique présente dans les blogs pour accentuer les évaluations (ex : *j’adôoore*), ou par l’absence de connaissances sémantiques sur des adjectifs évaluatifs non présents dans le lexique de l’évaluation (ex : *télégénique*, *puritaine*) ou encore par la non connaissance de certaines structures figées particulièrement présentes dans les textes de Sarah Palin (ex : *pittbull aux lèvres rouges*, *fibres écolo*).

<i>Tâche</i>	Précision	Rappel
DETECTION	88,4% (478/541)	50,1% (478/955)
CATEGORISATION MODALITE	-	-
Opinion	88,0% (44/50)	100,0% (44/44)
Appreciation	100,0% (393/393)	99,0% (393/397)
Accord/Désaccord	100,0% (35/35)	94,6% (35/37)
CATEGORISATION CONFIG. ENONCIATIVE	-	-
Explicite	89,6% (60/67)	96,8% (60/62)
Implicite	97,6% (290/297)	97,9% (290/296)
Exclamative	100,0% (29/29)	82,9% (29/35)
CATEGORISATION AXIOLOGIQUE	-	-
Favorable	93,3% (279/299)	97,6% (279/286)
Défavorable	96,5% (83/86)	77,6% (83/107)
Ambiguë	8 occurrences	

TAB. 3 – Mesure des résultats obtenus par l’outil d’analyse des blogs sur le corpus test.

Catégorisation de la modalité Dans un deuxième temps, nous évaluons l’outil de catégorisation à partir des évaluations correctement détectées. Nous observons (voir fig. 3) qu’il y a très

³Nous regroupons les jugements et les appréciations sous cette modalité de par leur forte proximité sémantique.

peu d'ambiguïté entre les différentes modalités annotées. Les structures et les entités lexicales qui composent ces différents types d'évaluation sont assez bien distincts. De ce fait, les résultats obtenus sont presque maximaux et viennent corroborer les définitions théoriques données par Charaudeau.

Catégorisation de la configuration d'énonciation La tâche de catégorisation de la configuration d'énonciation des appréciations fournit de bons résultats (voir Tab.3). On ne s'intéresse ici qu'aux appréciations correctement détectées. Comme pour les modalités, les structures symboliques sont particulièrement différentes entre les appréciations implicites et explicites. En effet, l'absence de pronoms ou de verbes d'appréciation implique souvent le fait que l'appréciation soit implicite. L'ambiguïté se situe plutôt sur les appréciations exclamatives puisque souvent une phrase exclamative ne contient pas qu'une seule appréciation d'où la difficulté de rattacher l'exclamation à une seule ou toutes les appréciations.

Catégorisation axiologique La tâche de catégorisation axiologique des appréciations correctement détectées fournit également des résultats encourageants (voir fig.3). Suite à la phase d'entraînement, nous avons constaté qu'il y avait en effet peu d'ambiguïtés entre une tournure inversive et une tournure directe des structures symboliques. Les résultats montrent toutefois que le rappel des appréciations défavorables est plus faible, ce qui est dû essentiellement à des structures inversives non apprises (ex : *ce n'est pas chose facile*) et à la présence de cas d'évaluations ironiques (ex : *Toujours aussi passionnant*).

6 Conclusion et perspectives

L'outil de catégorisation des évaluations est très satisfaisant. Le plus difficile reste la détection de celles-ci, nécessitant notamment une amélioration automatique de la couverture lexicale. Pour réaliser cet objectif, nous envisageons une méthode non supervisée pour apprendre de nouveaux termes et de nouvelles expressions figées utilisées pour évaluer. L'association fréquente d'un terme ou d'une expression avec certaines structures grammaticales apprises (ex : *c'est un véritable* + NOM) est un indice permettant d'induire leur rôle évaluatif. En s'inspirant de Hatzivassiloglou et McKeown (1997), il est également possible de déterminer automatiquement la polarité axiologique des termes ainsi extraits. Un deuxième axe de perspectives consiste à rechercher le concept sur lequel porte l'évaluation en cherchant les groupes nominaux ou les noms propres les plus pertinents dans le co-texte selon un algorithme proche de la résolution d'anaphore. La méthode symbolique pour la fouille d'opinion présentée dans cet article permet de détecter et catégoriser l'axiologie positive ou négative et le rôle discursif des évaluations exprimées localement dans les blogs avec une bonne précision. La grammaire du langage évaluatif construite automatiquement contient environ 2800 règles et permet de faire la distinction entre des structures évaluatives assez proches pouvant induire en erreur les approches *sac-de-mots* classiques en fouille d'opinion. Les règles de grammaire et les ressources lexicales développées peuvent ainsi être appliquées sur des corpus dont la thématique n'est pas fixée en amont sans perdre en précision. Ces travaux s'inscrivent dans le projet BLOGOSCOPIE, soutenu par le programme Technologies Logicielles 2006 de l'ANR et réalisé en collaboration avec Syllabs, Sinequa et Over-Blog.

Références

- ANSCOMBRE J. & DUCROT O. (1983). *L'argumentation dans la langue*. Pierre Mardag.
- AUE A. & GAMON M. (2005). Customizing sentiment classifiers to new domains : A case study. In *Recent Advances in Natural Language Processing (RANLP)*.
- BLITZER J., DREDZE M. & PEREIRA F. (2007). Biographies, Bollywood, boom-boxes and blenders : Domain adaptation for sentiment classification. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL 2007)*, p. 440–447.
- CHARAUDEAU P. (1992). *Grammaire du sens et de l'expression*. Hachette Education, COMMUNICATION, PARA UNIVERSITAIRE.
- DAVE K., LAWRENCE S. & PENNOCK D. M. (2003). Mining the peanut gallery : Opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th World Wide Web Conference (WWW 2003)*, p. 519–528.
- DUBREIL E., VERNIER M., MONCEAUX L. & DAILLE B. (2008). Annotating opinion - evaluation of blogs. In *Workshop of LREC 2008 Conference, Sentiment Analysis : Metaphor, Ontology and Terminology (EMOT-08)*, p. 124.
- GALATANU O. (2000). Signification, sens, formation. In *Education et Formation, Biennales de l'éducation, (sous la direction de Jean-Marie Barbier, d'Olga Galatanu)*, Paris : PUF.
- HU M. & LIU B. (2004). Mining opinion features in customer reviews. In *Proceedings of the 19th National Conference on Artificial Intelligence*, p. 755–760 : AAAI Press / The MIT Press.
- KERBRAT-ORECCHIONI C. (1997). *L'énonciation, de la subjectivité dans le langage*. Colin (édition 2002).
- LAVELLE L. (1950). *Traité des valeurs*, volume tome 1. PUF.
- MARTIN J. & WHITE P. (2005). *The Language of Evaluation, Appraisal in English*. Palgrave Macmillan.
- MAUREL S., CURTONI P. & DINI L. (2008). L'analyse des sentiments dans les forums. In *actes de l'atelier FODOP'08 (fouille de données d'opinions) (INFORSID'08)*, p. 111–117.
- MISHNE G. & GLANCE N. (2006). Predicting movie sales from blogger sentiment. In *AAAI Symposium on Computational Approaches to Analysing Weblogs (AAAI-CAAW)*, p. 155–158.
- MULLEN T. & MALOUF R. (2006). A preliminary investigation into sentiment analysis of informal political discourse. In *AAAI Symposium on Computational Approaches to Analysing Weblogs (AAAI-CAAW)*, p. 159–162.
- NIGAM K. & HURST M. (2006). Towards a robust metric of polarity. In J. G. SHANAHAN, Y. QU & J. WIEBE, Eds., *Computing Attitude and Affect in Text : Theories and Applications*, number 20 in the Information Retrieval Series : Springer.
- PANG B., LEE L. & VAITHYANATHAN S. (2002). Thumbs up ? sentiment classification using machine learning techniques. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2002)*, p. 79–86.
- STERN R. (2008). Constitution d'un lexique de sentiment. In *Mémoire de Master de Recherche Linguistique-Informatique* : Université Paris 7.
- WHITELAW C., GARG N. & ARGAMON S. (2005). Using appraisal groups for sentiment analysis. In *Proceedings of the ACM SIGIR Conference on Information and Knowledge Management (CIKM)*, p. 625–631 : ACM.