

## Un système de détection d'opinions fondé sur l'analyse syntaxique profonde

Caroline Brun

Xerox Research Centre Europe, 6 chemin de Maupertuis, 38240 Meylan, France  
Caroline.Brun@xrce.xerox.com

**Résumé :** Dans cet article, nous présentons un système de détection d'opinions construit à partir des sorties d'un analyseur syntaxique robuste produisant des analyses profondes. L'objectif de ce système est l'extraction d'opinions associées à des produits (les concepts principaux) ainsi qu'aux concepts qui leurs sont associés (en anglais «features-based opinion extraction»). Suite à une étude d'un corpus cible, notre analyseur syntaxique est enrichi par l'ajout de polarité aux éléments pertinents du lexique et par le développement de règles génériques et spécialisées permettant l'extraction de relations sémantiques d'opinions, qui visent à alimenter un modèle de représentation des opinions. Une première évaluation montre des résultats très encourageants, mais de nombreuses perspectives restent à explorer.

**Abstract:** In this paper, we present an opinion detection system built on top of a deep robust syntactic parser. The goal of this system is to extract opinions associated to products but also to characteristics of these products, i.e. to perform feature-based opinion extraction. To carry out this task, and following the results of a target corpus study, the robust syntactic analyzer is enriched by the association of polarity to pertinent lexical elements and by the development of generic rules extracting semantic relations of opinions, in order to feed an opinion representation model. A first evaluation gave very encouraging results, but many perspectives remain to be explored.

**Mots-clés :** détection d'opinions, analyse de sentiments, analyse syntaxique robuste, extraction d'information

**Keywords:** opinion detection, sentiment analysis, robust parsing, information extraction

### 1 Introduction

La fouille d'opinions (parfois aussi qualifiée d'analyse de sentiments) fait l'objet d'un engouement tout particulier que ce soit dans les milieux académiques ou dans l'industrie. En effet, avec l'émergence de groupes de discussions, forums, blogs, sites compilant des avis consommateur, on trouve une masse très importante de documents contenant des informations exprimant des opinions, constituant une source énorme de données pour des applications de veille diverses (technologique, marketing, concurrentielle, sociétale). De nombreux travaux de recherche, à la croisée du TALN et de la fouille de données, se penchent sur le problème de la détection d'opinions. Dans cet article, nous présentons le système de détection d'opinions développé pour l'anglais dans le cadre du projet européen Scoop<sup>1</sup>, système basé sur l'utilisation d'un analyseur syntaxique robuste adapté à l'analyse des opinions. Après une brève revue des travaux du domaine, nous décrirons l'analyse de corpus que nous avons réalisée, sur un premier corpus cible constitué de revues sur des imprimantes, photocopieurs et scanners extraits du site de revues grand public "Epinion", et qui a conduit à la conception et au développement du système, lui-même décrit en détails dans la section suivante. Nous présenterons ensuite l'évaluation préliminaire de ce système et concluons sur le bilan et les perspectives envisagées.

---

<sup>1</sup> <http://www.scoopproject.eu/overview.html>

## 2 Travaux en fouille d'opinions

Outre les travaux concernant la constitution de ressources lexicales pour l'analyse d'opinions, évoqués dans la section 4.2.2, deux groupes de travaux autour de la fouille d'opinions se distinguent : ceux visant à classifier les textes selon une polarité globale (positive, négative et quelquefois neutre), généralement basés sur des approches supervisées (comme (Pang et al. 2002), ou (Charton et Acuna-Agost 2007)), et ceux visant à l'extraction précise d'opinions exprimées sur des aspects positifs ou négatifs. Ces derniers considèrent qu'un *concept principal* (par exemple, un produit) se compose de plusieurs *concepts associés* que l'on peut évaluer séparément (par exemple, qualité, vitesse d'impression, et résolution pour une imprimante). Notre système se situe dans cette catégorie. Le but ici est de détecter les concepts associés (caractéristiques d'un produit) ainsi que les opinions émises à propos de ces concepts. On distingue donc trois sous-tâches : l'extraction des concepts associés, la découverte des opinions (positive, négative et parfois neutre), et la production d'un résumé des informations associées à un concept donné. Pour extraire les concepts associés, les méthodes se basent généralement sur des critères de fréquence couplés à des heuristiques linguistiques (Yi et al. 2003), (Popescu and Etzioni 2005). Pour associer les opinions aux concepts, (Hu et Liu 2004) extraient les segments contenant un concept et additionnent les orientations du vocabulaire d'opinion présent dans le même segment. (Vernier et al. 2009) proposent une méthode symbolique de détection et de catégorisation des évaluations localement exprimées dans un corpus de blogs multi-domaine. Certains systèmes utilisent des dépendances syntaxiques pour relier source et cible de l'opinion comme (Kim and Hovy 2006) ou (Bloom et al. 2007). Notre système se situe dans cette famille, car le traitement syntaxique de phénomènes complexes (négation, comparaison, dépendances à longue distance) semble une étape nécessaire à l'extraction fine d'opinions. Une particularité de notre système est son architecture à deux niveaux : un premier niveau, général et valide sur tout type de corpus, et un deuxième niveau, adapté pour chaque sous-domaine d'application.

## 3 Analyse de corpus

Nous avons utilisé un corpus de revues "grand public", disponibles sur le site web "Epinion" (<http://www.epinions.com/>), un site généraliste compilant des millions de revues concernant produits, films, livres, etc. La première application visée par notre système concernant les revues d'utilisateurs d'imprimantes en anglais, nous avons extrait de ce site environ 3500 revues d'imprimantes. Ces revues, semi structurées, contiennent : le nom du produit, la note globale (de 0 à 5 étoiles), le titre de la revue, la date de création, les «Pros», «Cons» et «Bottom line», et la revue complète en texte libre, se terminant par la mention "Recommended" : "yes" ou bien "no". Une étude menée sur une centaine de ces revues (choisies au hasard), servant de corpus de développement au système, confirme deux éléments capitaux :

1. Un ensemble de phénomènes linguistiques complexes sont mis en jeu dans l'expression des opinions et montrent que la seule polarité (positive ou négative) associée à des mots clés est insuffisante pour la détection fine d'opinions :
  - La négation, lexicale ou syntaxique, qui inverse la polarité des opinions : (i) *I can't use it without problems.* (ii) *There is no way I can recommend this printer.*
  - La modalité, qui affecte le caractère avéré des opinions exprimées: (i) *Considering the high cost of the printer, the quality should be outstanding and there should be no problems.*
  - La comparaison, exprimant une opinion comparative entre deux éléments: (i) *I would be happier with a better price,* (ii) *Performance is better than many competing lasers.*
  - La coréférence, pour la détection du sujet de l'opinion : sur l'exemple suivant, l'auteur réfère à d'autres produits (mentions soulignées) qui ne sont pas le sujet de la revue (mentions en gras) : (i) ***Xerox DocuPrint P8ex Laser Printer**: When my previous printer (HP LaserJet 5) did not last as long as I would like it to have lasted, I had many options for a new printer. I had one functional HP remaining, one NEC and then I bought **this Xerox P8ex**. I have since purchased another Xerox laser all-in-one because I was quite satisfied with the performance of the **P8ex**.*
2. Concernant le vocabulaire d'opinion, il est nécessaire de tenir compte des problèmes :
  - D'ambiguïtés, car un mot d'un même domaine peut selon les contextes indiquer des opinions de polarités différentes, comme l'adjectif "fast" dans le domaine des imprimantes : (i) *It uses ink twice as **fast**.* [opinion négative], (ii) *It is also a **fast**, high quality printer.* [opinion positive]
  - De polarité dépendante du domaine : (i) *It walks like a **lemon** and quacks like a **lemon*** : Dans le domaine des produits, "lemon" est négatif, alors que c'est un mot objectif (neutre) en général.

Suivant cette étude, notre système est conçu selon une architecture à deux niveaux, un premier niveau contenant du vocabulaire de polarité invariante selon les domaines, et des règles génériques, puis un

deuxième niveau, contenant du vocabulaire de polarité liée au domaine et des règles également spécifiques au domaine.

## 4 Description du système

### 4.1 Objectifs et modèle adopté

Nous souhaitons développer un système d'extraction d'opinions sur des revues de produits. Notre objectif est de classer les revues comme positives ou négatives (opinion au niveau du document), mais aussi d'extraire finement les opinions émises sur les concepts principaux (les imprimantes) et associés au produit (vitesse, qualité d'impression etc., pour une imprimante). En effet, au delà d'une opinion globale, positive ou négative, il est très intéressant de détecter ce que les utilisateurs aiment ou au contraire n'aiment pas dans un produit donné. Pour atteindre cet objectif, nous adoptons la représentation formelle d'une opinion proposée par (Liu, B. 2010), sous forme d'un quintuplet de la forme  $(o_j, f_{jk}, so_{ijkl}, h_i, t_l)$  où :  $o_j$  est le concept principal cible de l'opinion (=l'objet),  $f_{jk}$  est un concept associé au concept principal  $o_j$ ,  $h_i$  est le détenteur de l'opinion,  $so_{ijkl}$  est la valeur de l'opinion (positive, négative ou neutre) émise par  $h_i$  sur le concept  $f_{jk}$ , et  $t_l$  est le moment où l'opinion est exprimée. Nous proposons d'instancier, à partir des résultats de l'analyse syntaxique profonde fournie par XIP, des relations sémantiques alimentant des quintuplets conformes à ce modèle formel.

### 4.2 Description des développements

Notre système d'extraction d'opinions utilise l'analyseur syntaxique XIP, (Ait-Mokhtar et al., 2002), qui produit, de façon robuste, une représentation syntaxique profonde (contrôle des infinitives, normalisation de la forme passive, etc.) et intègre un module de reconnaissance des entités nommées, (Brun & Ehrmann 2009), auquel nous avons intégré la reconnaissance des noms d'imprimantes, pour les besoins de ce projet. Nous cherchons à extraire, à partir du texte d'une revue, des relations sémantiques permettant d'instancier les quintuplets (cf. section 4.1) modélisant les opinions. Dans le cadre de notre application, nous simplifions le modèle en considérant que le moment où l'opinion est exprimée est la date de création de la revue et que le détenteur de l'opinion est son auteur. De plus, s'il n'est pas explicitement mentionné dans la phrase, par défaut, l'objet de l'opinion est le produit revu. La valeur de l'opinion est modélisée par un trait ("positif" ou "négatif") associé à la relation de sentiment. Enfin, nous conservons dans la relation le prédicat porteur de la polarité, car il peut être utile pour une phase de normalisation sémantique ultérieure. Nous cherchons donc à extraire des relations sémantiques de la forme : SENTIMENT[POLARITE](CONCEPT\_PRINCIPAL,CONCEPT\_ASSOCIÉ,PRÉDICAT\_PORTEUR), comme :

- (1) "This printer is slow" → SENTIMENT[NEGATIF](printer,\_, slow)
- (2) "The laser print quality is great." → SENTIMENT[POSITIF](Default\_Product, print quality, great)

Dans (1), le prédicat porteur du sentiment est "slow", le concept principal est "printer", l'opinion (négative) porte sur ce concept dans son entier (et non sur un concept associé). Dans (2), le prédicat porteur du sentiment est "great", le concept associé sur lequel porte l'opinion (positive) est "print quality" et comme il n'est pas explicitement mentionné, le concept principal de l'opinion est l'objet de la revue (Default\_product). Nous avons implanté un lexique de vocabulaire porteur d'opinion, et des règles d'extraction de relations sémantiques de sentiments, selon l'architecture à deux niveaux proposée dans la section 3.

#### 4.2.1 Extraction des concepts associés

Les objets cibles de notre application sont les produits commentés dans les revues. Le vocabulaire qui les désigne est : *printer, copier, scanner, machine, product*. Afin d'extraire les concepts associés à ces objets, nous utilisons une méthode proche de (Popescu and Etzioni 2005), qui cherchent des relations de méronymie (partie-tout). Nous utilisons XIP pour extraire automatiquement les noms les plus fréquents qui modifient un concept principal selon les deux relations syntaxiques suivantes : MOD-ANTEPOSE(CONCEPT\_PRINCIPAL,CONCEPT\_CANDIDAT-ASSOCIE) correspondant à "printer quality", pour lequel "quality" est le candidat, et MOD-PREP[OF](CONCEPT\_CANDIDAT-ASSOCIE,CONCEPT\_PRINCIPAL), correspondant à "the speed of the machine", pour lequel "speed" est le candidat. 736 candidats concepts associés sont ainsi extraits, avec leur fréquence. Nous considérons ensuite qu'un candidat est effectivement un concept associé s'il est au moins une fois en relation attributive avec les adjectifs "good" et "bad" dans

le corpus. Ces relations attributives sont extraites automatiquement du corpus avec XIP. Au final, nous obtenons une liste de 76 concepts associés, les plus fréquents étant : *quality, speed, photo, color, software, cartridge, price, resolution, ...* Une vérification manuelle permet de constater que ce sont bien des concepts associés désignant des parties matérielles (*cartridge, drum*), des caractéristiques techniques (*performance, resolution, speed*), ou des concepts voisins (*price, support, warranty*).

#### 4.2.2 Constitution du lexique de vocabulaire d'opinion

La constitution d'un vocabulaire subjectif encodant les polarités (positives ou négatives) est une nécessité pour toute application de fouille d'opinion. Ce vocabulaire est constitué d'adjectifs ("beautiful", "ugly"), de noms ("talent", "nuisance"), de verbes ("love", "hate") et d'adverbes, ("admirably", "annoyingly"). De nombreux travaux s'attaquent à ce problème. Par exemple, (Agarwal et Bhattacharyaa 2006) classent des adjectifs selon leur polarité en calculant leur degré d'association avec des adjectifs "déclencheurs" de polarité connue. (Vegnaduzzo 2004) utilise aussi des adjectifs déclencheurs et une méthode distributionnelle basée sur la syntaxe. (Esuli et Sebastiani 2006) développent SentiWordnet : ils analysent quantitativement les définitions associées aux synsets avec différents classifieurs statistiques, et associent des mesures de positivité, négativité et objectivité à chaque synset. Nous n'avons pas utilisé SentiWordnet dans le cadre de notre application, car l'ambiguïté de chaque entrée lexicale de Wordnet est préservée. Ne disposant pas non plus d'un corpus manuellement annoté (comme (Vernier et al. 2009)), nous avons à nouveau utilisé XIP pour extraire automatiquement sur l'ensemble du corpus, des relations syntaxiques, afin de sélectionner le vocabulaire potentiellement subjectif. Ce sont des relations attributives, pour extraire les noms et adjectifs attribués d'un concept principal ou associé ou encore du pronom «I», comme dans «the **size** of the printer is **huge**» ou «**I** am extremely **unhappy**», des relations de modifieurs antéposés d'un concept principal ou associé, comme dans «It prints **great photos**» et des relations Sujet-Verbe-Objet et Sujet-Verbe, pour extraire les verbes dont le sujet est un concept principal ou associé, et le complément d'objet le pronom "I", comme dans «**I** am **disappointed** with this **product**», ou des verbes dont le sujet est un concept principal ou associé, comme dans «This **printer stinks**!». Les résultats sont filtrés selon la catégorie du candidat et sa fréquence, puis la polarité ainsi que l'appartenance au lexique général ou au lexique de domaine sont affectées manuellement. WordNet est également utilisé pour étendre avec les synonymes et antonymes des mots validés. Nous obtenons 130 verbes dans le lexique général et 42 dans le lexique spécialisé, 465 adjectifs dans le lexique général et 230 dans le lexique spécialisé, et 145 noms dans le lexique général et 42 dans le lexique spécialisé. Nous avons ainsi constitué un lexique de polarité "générique", valable pour toute application, et un lexique lié au domaine des imprimantes. Cependant, la seule mention des polarités est insuffisante au développement des règles. Elles nécessitent des traits de sous-catégorisation sémantique sur les prédicats encodant la portée de l'opinion. Les traits sémantiques ajoutés indiquent si l'opinion porte sur le sujet, (1), le complément d'objet des verbes, (2), ou sur certains compléments prépositionnels, (3). Nous ajoutons aussi des traits sémantiques de domaine, car dans (4), c'est la combinaison [easy + to + verbe indiquant une fonctionnalité de l'imprimante] qui exprime une opinion positive :

- (1) *These **printers** never cease to **amaze** me with both quality and price.*
- (2) *I **appreciate** the **swiftness** of this machine.*
- (3) *We have had several **problems** with a **LaserJet 8000**.*
- (4) *This machine was very **easy** to **setup**.*

#### 4.2.3 Développement du système de règles

Une fois encodé le vocabulaire, nous avons développé manuellement un ensemble de règles, construites sur la sortie de l'analyse syntaxique profonde, et extrayant des relations sémantiques d'opinion. Les règles sont également divisées en deux sous-ensembles : des règles génériques et des règles propres au domaine. Les règles génériques testent, pour un schéma syntactico-sémantique donné, la présence de vocabulaire d'opinion dans les arguments, et en fonction des informations de sous-catégorisation sémantique associées, extraient la relation de sentiment correspondante. Par exemple, la règle XIP suivante :

If (SUBJ-N(#1[polarite!polarite:!,topic-subj],#2[concept-principal]) )  
SENTIMENT(#2,#1)

Indique que si on a détecté le sujet normalisé (SUBJ-N) d'un verbe (#1) exprimant une opinion (trait polarite : *positif* ou *négatif*) sur son sujet (trait de sous-catégorisation *topic-subj*) sur un concept principal (#2, trait *concept-principal*), alors on crée une relation de sentiment en utilisant la percolation

(!polarite :!) qui associe la valeur positive ou négative selon l'orientation du verbe (#1). Cette règle permet par exemple de traiter les exemples suivants :

- (1) «*These **printers** never cease to **amaze** me.*» → SENTIMENT[POSITIF](printer,\_,amaze)  
 (2) «*I was quite **disappointed** with this **machine**.*» → SENTIMENT[NEGATIF](machine,\_,disappoint)

Des règles très similaires sont également développées dans le cas où les opinions portent sur les concepts associés. Le système contient une soixantaine de règles génériques, couvrant la majorité des constructions repérées lors de notre étude de corpus.

Nous avons aussi mis l'accent sur le traitement de la négation, en développant des règles permettant de traiter les cas de négation en style télégraphique (“*Not quite as fast as HP says*”), d'interaction entre négation et quantification (“*I never had so many problems*”) ou de double négation (“*I can't say I don't appreciate this printer*”). Ces règles de négation se combinent ensuite avec les relations de sentiments déjà extraites à l'étape précédente pour inverser la polarité, pour les exemples du type “*I really don't like this feature*”, “*This is not a good photo printer*”, etc.

De plus, une couche de règles spécifiques au domaine des imprimantes a été développée, traitant les associations du type :

- (1) “*It is easy to set up.*” → SENTIMENT[POSITIF](it,\_,easy-to-set-up)  
 (2) “*It uses a lot of ink*” → SENTIMENT[NEGATIF](it,\_,use-ink)

Dans (1), c'est l'association [easy + to + verbe de fonctionnalité de l'imprimante] qui exprime une opinion positive. Dans (2), un verbe de consommation («consume», «use», «eat», ...) ayant pour complément d'objet un consommable de l'imprimante («ink», «paper», «cartridge», ...) indique une opinion négative. Une vingtaine de règles de ce type, basées sur le système de traits sémantiques décrit dans la section précédente, testent la présence de telles associations et produisent la relation de sentiment adaptée.

Enfin, une dizaine de règles prennent en compte la structure même des revues du site “Epinion” : elles permettent d'utiliser certains indices structurels (“*Cons*”, “*Pro*”, “*Recommended*”,...) pour le calcul des opinions (par exemple “*Cons : none*” indique une opinion positive).

## 5 Evaluation

Ne disposant pas d'un corpus de revues d'imprimantes annoté en termes de relations d'opinions, nous avons mis à profit la structure des revues du site “Epinion” pour faire une évaluation préliminaire «gros grain» des performances du système : nous comparons la recommandation finale de l'utilisateur avec le pourcentage global d'opinions positives ou négatives extraites par notre système, sur tout type de concept, principal ou associé. Si la revue est favorable et que le pourcentage d'opinions positives extraites par le système est supérieur à 50%, ou à l'inverse, si la revue est défavorable et que le pourcentage d'opinions négatives extraites est supérieur à 50%, nous considérerons que le système classe correctement la revue selon une opinion globale. Le tableau I montre les résultats obtenus sur le corpus de test (3337 revues). Ces résultats sont très encourageants car en accord avec les résultats de l'état de l'art obtenus pour des tâches similaires, cf. (Pang et al. 2002) ou (Paroubek et al. 2007). De plus, afin de valider la qualité et la portabilité de notre grammaire générale sur d'autres domaines, nous avons conduit une évaluation similaire sur des revues de films, extraites aussi du site “Epinion”. Seuls les concepts principaux et associés sont modifiés dans le lexique de domaine, et nous n'utilisons que la grammaire générale d'extraction d'opinions. Les résultats, également satisfaisants, sont aussi donnés sur le tableau suivant :

	Revue favorable		Revue défavorable		Total des revues	
	Imprimantes	Cinéma	Imprimantes	Cinéma	Imprimantes	Cinéma
nombre	2517	1343	820	420	3337	1763
Classées positive	2440	1264	218	143	2658	1407
Classées négative	77	79	602	277	679	356
précision	97%	94%	73.4%	66%	91%	87.5%

Tableau I : résultats de l'évaluation pour le corpus “imprimantes” et le corpus “cinéma”

Pour les deux évaluations, le système a plus de difficulté à classer correctement les revues défavorables. Il y a plusieurs explications à cela : tout d'abord, comme le vocabulaire d'opinion est en partie extrait à partir des corpus, et que la proportion de revues défavorables est plus faible, il est possible qu'il y ait un problème de couverture pour le vocabulaire négatif. D'autre part, les rédacteurs semblent avoir un discours différent selon qu'ils recommandent ou non un produit ou un film : une analyse rapide des résultats montre que, dans les cas défavorables, les rédacteurs ont tendance à utiliser la comparaison avec d'autres produits

ou films qu'ils ont au contraire appréciés. Notre système ne traitant que très partiellement la comparaison et n'intégrant pas encore de module de coréférence, nombre d'opinions positives exprimées sur d'autres produits ou films sont comptabilisées de façon erronée au crédit du produit/film cible de la revue.

## 6 Conclusion et perspectives

Dans cet article, nous présentons un système d'extraction automatique d'opinions sur des revues de produits en ligne, utilisant des relations syntaxiques profondes pour construire des relations sémantiques correspondant aux opinions exprimées dans les revues. Nous avons développé un lexique contenant les polarités associées au vocabulaire subjectif mais aussi une structuration en traits sémantiques, et un ensemble de règles manuelles extrayant les relations d'opinion. L'évaluation des performances du système sur la tâche de classification des revues est très encourageante. Nous devons à présent poursuivre les développements en cours afin de prendre en compte des phénomènes linguistiques peu couverts par notre système, à savoir la modalité et les comparatives, ayant jusqu'à présent mis l'accent sur le traitement de la négation. D'autre part, une étape essentielle pour atteindre notre objectif d'extraction d'opinions sur les concepts associées aux produits revus est l'intégration d'un module de coréférence. A la suite de cela, nous réaliserons une évaluation plus fine de nos résultats.

## Références

- AGARWAL A., BHATTACHARYA, P. (2006). Augmenting Wordnet with Polarity Information on Adjectives. *3rd International Wordnet Conference*, Jeju Island, Korea, South Jeju (Seogwipo).
- AIT-MOKTHAR, S. CHANOD J.P., ROUX C. (2002). Robustness beyond Shallowness: Incremental Dependency Parsing. *Special Issue of NLE Journal*.
- BLOOM, K. NAVENDU G., ARGAMON S. (2007). Extracting Appraisal Expressions. Actes de *HLT-NAACL*, Rochester, USA.
- BRUN C., EHRMANN. (2009). Adaptation of a Named Entity Recognition System for the ESTER 2 Evaluation Campaign. *IEEE NLP-KE 2009*, Dalian, China, Sep 24-27.
- CHARTON E., ACUNA-AGOST R. (2007). Quel modèle pour détecter une opinion? Trois propositions pour généraliser l'extraction d'une idée dans un corpus, Actes du *Défi Fouille de Textes, Grenoble*.
- ESULI A, SEBASTIANI F. (2006). *SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining*. Actes de 5th Conference on Language Resources and Evaluation (LREC'06), pp. 417-422.
- HU M., AND LIU B. (2004). Mining and summarizing customer reviews. Actes de *ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD-2004)*, Seattle, USA.
- KIM S-M, HOVY E. (2006). Identifying and analyzing judgment opinions. Actes de *HLT-NAACL*, New-York, USA.
- LIU, B. (2010). Sentiment Analysis and Subjectivity, *Handbook of Natural Language Processing, 2eme ed.*
- PANG, B. LEE, L., VAITHYANATHAN, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. Actes de *EMNLP (Conférence on Empirical Methods in Natural Language Processing)*, pages 79–86.
- PAROUBEK P., BERTHELIN J.B., EL AYARI S., GROUIN C., HEITZ T., HURAUULT-PLANTET M., JARDINO M., KHALIS Z., LASTES M. (2007). Résultats de l'édition 2007 du DÉfi Fouille de Textes, Actes du *Défi Fouille de Textes, Grenoble*.
- POPESCU, A. ETZIONI, O. (2005). Extracting product features and opinions from reviews. Actes de *EMNLP (Conférence on Empirical Methods in Natural Language Processing)*.
- VERNIER M., MONCEAUX L., DAILLE B., DUBREIL E. (2009) Catégorisation des évaluations dans un corpus de blogs multi-domaine. *Revue des nouvelles technologies de l'information (RNTI)*.
- VEGNADUZZO, S. (2004). Acquisition of subjective adjectives with limited resources. Actes de *AAAI spring symposium on exploring attitude and affect in text: Theories and applications*, Stanford, USA.
- YI, J. NASUKAWA T., VALERIO A., ZHANG H. (2003). Sentiment analyzer: Extracting sentiments about a given topic using NLP techniques. Actes d'ICDM'03: *3<sup>rd</sup> IEEE International conference on Data Mining*.