Catégorisation sémantique fine des expressions d'opinion pour la détection de consensus

Farah Benamara, Véronique Moriceau, Yvette Yannick Mathieu

Catégorisation sémantique fine des expressions d'opinion pour la détection de consensus

Farah Benamara[1] Véronique Moriceau[2] Yvette Yannick Mathieu[3]
(1) IRIT-CNRS, Université Paul Sabatier, 31062 Toulouse
(2) LIMSI-CNRS, Université Paris-Sud, 91403 Orsay
(3) LLF-CNRS, Université Paris Diderot, 75013 Paris
benamara@irit.fr, moriceau@limsi.fr, yannick.mathieu@linguist.jussieu.fr

Résumé. Dans cet article, nous présentons notre participation aux tâches 2 et 3 de DEFT 2014. Ces tâches consistaient respectivement à évaluer la qualité littéraire de nouvelles courtes en prédisant la note que donnerait un juge humain et à déterminer, pour chacune des nouvelles, si elle est consensuelle auprès des différents relecteurs. Pour ces tâches, nous avons utilisé une approche par apprentissage automatique qui s'appuie sur un lexique d'opinions fournissant une catégorisation sémantique fine des expressions d'opinion en français.

Abstract. In this paper, we present our participation to the tasks 2 and 3 of DEFT 2014. These tasks consisted in evaluating the quality of short stories by predicting the score that a human reviewer would give and in determining, for each short story, if it is consensual for the different reviewers. For these tasks, we used a machine-learning approach based on a French lexicon of opinion expressions where each entry is associated with a fine-grained semantic categorization.

Mots-clés : Lexique d'opinion, classification d'opinion multi-échelle, détection de consensus.

Keywords: Lexicon-based opinion mining, multi-scale rating, consensus detection.

1 Introduction

L'édition 2014 du Défi Fouille de Textes (DEFT) était consacrée en partie à l'analyse de textes littéraires, à savoir des nouvelles courtes. Dans un premier temps, la tâche 2 consistait à évaluer la qualité littéraire de nouvelles courtes en prédisant les notes données par des juges humains (relecteurs), ces notes se trouvant sur une échelle allant de 1 (meilleure note) à 5 (moins bonne note). Ensuite, une fois ces notes prédites, la tâche 3 consistait à déterminer, pour chaque nouvelle, si elle fait consensus auprès des différents relecteurs. Une nouvelle est considérée comme consensuelle si les notes attribuées par les différents relecteurs ne varient pas au-delà d'un écart de 1 point.

Pour ces deux tâches, nous avons utilisé une approche par apprentissage automatique. Afin de prédire les notes données à chaque nouvelle par les différents relecteurs, nous avons tout d'abord projeté un lexique d'opinion sur les commentaires textuels (relectures) écrits par les relecteurs. Cette projection du lexique nous a permis de définir un certain nombre de traits que nous avons utilisés pour l'apprentissage automatique.

Dans cet article, nous commençons par décrire le lexique d'opinion que nous avons utilisé puis comment nous l'avons projeté sur les différentes relectures du corpus. Nous présentons ensuite en détail les différents traits que nous avons définis pour l'apprentissage automatique afin de prédire les notes des relecteurs et le consensus entre eux. Enfin, nous présentons les résultats obtenus lors du défi pour les tâches 2 et 3.

2 Utilisation d'un lexique d'opinion

2.1 Présentation du lexique

Le lexique utilisé est un lexique d'expressions d'opinion désambiguisées en français construit manuellement. Ce lexique a été élaboré à partir d'un premier travail réalisé par (Asher *et al.*, 2008) à partir de l'étude de corpus variés (articles de presse, commentaires web et courrier des lecteurs) puis augmenté dans le cadre du projet DGA-RAPID CASOAR [1]. Dans ce lexique, en plus des informations classiques de polarité et d'intensité, chaque entrée a été classée dans des catégories sémantiques qui sont indépendantes d'une langue donnée et définies dans (Asher *et al.*, 2008). L'approche pour catégoriser les opinions utilise les recherches en sémantique lexicale de (Levin, 1993) et (Wierzbicka, 1987) pour l'anglais et de (Mathieu, 1999) (Mathieu, 2005) pour le français.

Le lexique est composé de verbes, de noms, d'adverbes, d'adjectifs et d'interjections. Deux types de verbes ont été sélectionnés : des verbes qui introduisent des expressions d'opinion et qui reflètent le degré d'implication de la personne qui émet l'opinion (comme *dire, se demander, insister,* etc.), et des verbes qui expriment explicitement et directement une opinion (comme *aimer, blâmer, recommander*).

Chaque entrée du lexique (sauf les adverbes) est associée à l'une des quatre catégories sémantiques de haut niveau suivantes :
- **Reportage** : entrées qui permettent de relater ou d'introduire les opinions des autres ou les siennes, et qui fournissent une évaluation du degré d'implication ou de l'engagement à la fois de la personne qui exprime l'opinion et de son objet, comme le verbe *estimer* dans *les routiers français estiment souffrir d'une fiscalité désavantageuse comparé à leurs rivaux européens* ;
- **Jugement** : entrées qui expriment des évaluations normatives d'objets et d'actions, à l'intérieur desquelles on peut distinguer des jugements reliés aux normes sociales, par exemple les verbes *approuver* et *critiquer* dans *Laurence Parisot approuve la réforme mais critique la méthode*, et des jugements reliés à des normes personnelles comme *C'est un pur chef d'oeuvre* ;
- **Sentiment-appréciation** : entrées qui expriment un sentiment ou une émotion ressentie par une personne, comme *J'ai adoré ce film* ;
- **Conseil** : entrées qui enjoignent de faire ou penser quelque chose, par exemple *Un excellent film à ne pas manquer*.

Chacune de ces catégories est également découpée en 24 sous-catégories. Par exemple, la catégorie *Sentiment-appréciation* regroupe des expressions d'opinion exprimant la colère, l'étonnement, la haine ou la déception (voir (Asher *et al.*, 2009) pour une présentation détaillée des sous-catégories).

Les adverbes sont quant à eux associés à l'une des catégories sémantiques suivantes :
- **Négation** : adverbes qui indiquent principalement des mots de négation. Nous considérons deux types d'adverbes de négation : les mots de négations (comme *ne, sans*) et les quantifieurs de négation (comme *jamais, personne*). Nous considérons également les négations lexicales qui sont des verbes ou des noms qui expriment des négations (comme *manquer de, absence*). Dans ce dernier cas, nous indiquons explicitement si l'entrée lexicale a un emploi de négation.
- **Affirmation** : adverbes qui indiquent une affirmation, comme *absolument, impérativement,* etc. ;
- **Doute** : adverbes qui expriment le doute, comme *probablement, peut-être,* etc. ;
- **Manière** : adverbes qui apportent une indication de manière, comme *admirablement, horriblement,* etc. ;
- **Intensifieur** : adverbes qui regroupent les adverbes de quantité et quelques adverbes de temps comme *toujours, beaucoup, très, peu,* etc.

Seuls les adverbes de manière expriment des opinions. Les autres catégories sont utilisées pour augmenter, diminuer ou inverser la force ou la polarité d'un mot ou d'une expression d'opinion. C'est pourquoi ce type d'adverbe est aussi associé aux mêmes catégories sémantiques que celles utilisées pour les entrées d'autres catégories grammaticales.

En plus de la catégorie sémantique, chaque entrée du lexique est associée à une polarité et une force. La polarité peut avoir trois valeurs : positive, négative ou neutre. Il est important de noter que la polarité neutre ne signifie pas que l'entrée associée est objective mais qu'elle possède une polarité ambiguë, qui peut être positive ou négative selon le contexte (par exemple *froid, ahurissant, bouleversant, délicat*, etc.). Pour la force, les valeurs possibles sont 1, 2 et 3, du plus faible au plus fort. Ainsi pour *bon* la force est de 1, pour *excellent*, elle est de 2, et pour *extraordinaire*, la force est de 3. Ces valeurs s'appliquent également aux expressions de polarité négative. Par exemple *décevant, nul* et *nullissime* ont respectivement

1. projetcasoar.wordpress.com

une force de 1, 2 et 3.

Une entrée peut être un mot (*grotesque, succès*), une expression figée (*bon enfant, haut de gamme*) ou non (*politiquement correct*). Une entrée peut également avoir un seul sens (et donc une polarité et une intensité unique) mais peut aussi avoir plusieurs sens dépendants du contexte : dans ce cas, une entrée peut appartenir à plusieurs catégories sémantiques et avoir des polarités et intensités différentes. Par exemple :

– l'adjectif *acide* a deux sens :
  1. aigre (comme dans *un fruit acide*) : dans ce cas, il appartient à la catégorie *jugement* et a une polarité négative ;
  2. blessant (comme dans *ils ont échangé des propos acides*) : dans ce cas, il appartient à la catégorie *sentiment-appréciation* et a aussi une polarité négative.

– l'adjectif *rigoureux* a deux sens :
  1. qui fait preuve de rigueur (comme dans *un juge rigoureux*) : dans ce cas, il appartient à la catégorie *jugement* et a une polarité positive ;
  2. pénible (comme dans *un hiver rigoureux*) : dans ce cas, il appartient aussi à la catégorie *jugement* mais a une polarité négative.

Le lexique compte au total 2830 entrées lexicales dont 297 expressions composées de plusieurs mots. Le tableau 1 décrit la répartition des entrées lexicales selon la catégorie grammaticale. Le tableau 2 décrit la répartition des entrées lexicales selon la catégorie sémantique.

Catégorie grammaticale	**Nombre d'entrées**
Adjectif	1142
Adverbe	605
Nom	415
Verbe	308
Expression	292
Interjection	62
Conjonction, préposition, pronom	6
TOTAL	2830

TABLE 1 – Répartition des entrées lexicales selon la catégorie grammaticale.

Toutes les catégories grammaticales		**Uniquement les adverbes** (hors *manière*)	
Catégorie sémantique	**Nombre d'entrées**	**Catégorie sémantique**	**Nombre d'entrées**
Reportage	65	Doute	17
Jugement	2234	Intensité	107
Sentiment	407	Négation	23
Conseil	26	Affirmation	37
TOTAL	2732	**TOTAL**	184

TABLE 2 – Répartition des sens des entrées lexicales selon la catégorie sémantique.

2.2 Projection du lexique sur les documents

Chaque document du corpus pour les tâches 2 et 3 est composé d'une nouvelle courte et d'une ou plusieurs relectures sous forme d'un commentaire textuel et d'une note de 1 (meilleure note) à 5 (moins bonne note). Les notes ne sont fournies que pour le corpus d'apprentissage, ce sont ces notes qu'il faut prédire lors de la phase de test. Le format des relectures est donné dans la figure 1.

Pour prédire les notes données par les relecteurs, nous avons choisi de nous intéresser aux commentaires textuels, en particulier aux opinions qu'ils véhiculent. Pour cela, nous avons projeté le lexique décrit précédemment sur les commentaires.

Nous avons dans un premier temps utilisé l'analyseur syntaxique XIP (Aït-Mokhtar *et al.*, 2002) pour lemmatiser les commentaires et ainsi projeter les entrées lemmatisées du lexique. Grâce aux dépendances syntaxiques fournies, nous

```
<reviews>
<review>
<id>1</id>
<uid>12100</uid>
<content>intéressant au départ, décevant au final</content>
<note>4</note>
</review>
</reviews>
```

FIGURE 1 – Format d'une relecture dans le corpus.

avons aussi récupéré les différentes négations et les éléments qui se trouvent dans leur portée (*ne... pas, ne... plus, ne... jamais, ne... rien, sans, aucun*). Ceci va permettre d'inverser les polarités d'expressions d'opinion qui se trouvent dans la portée d'une négation [2].

Pour les expressions composées de plusieurs mots, nous avons toléré l'insertion de 15 caractères entre deux mots afin de permettre la reconnaissance d'expressions non figées (par exemple, *sans (aucun | l'ombre d'un) doute*).

Même si XIP fournit les catégories morphosyntaxiques des mots et qu'elles sont aussi disponibles dans le lexique, nous avons choisi de ne pas en tenir compte étant donné le nombre d'erreurs possibles lors de l'analyse syntaxique.

Finalement, ont été projetés sur les lemmes des commentaires appartenant au lexique leur catégorie sémantique et leur polarité. Si un lemme est associé à plusieurs sens dans le lexique, nous avons projeté toutes les polarités des différents sens possibles.

Les exemples suivants montrent le résultat de la projection sur différents commentaires, contenant entre autres des négations et des mots à polarité ambiguë.

```
<content>frais, efficace et sympathique</content>
<phrase id="1">
 <lemme pos='ADJ' neg='0' category='évaluation' type='jugement' pol='pos' strength='1'>
  frais
 </lemme>
 <lemme pos='ADJ' neg='0' category='évaluation' type='jugement' pol='pos' strength='1'>
  efficace
 </lemme>
 <lemme pos='CONJ' neg='0'>et</lemme>
 <lemme pos='ADJ' neg='0' category='évaluation' type='jugement' pol='pos' strength='1'>
  sympathique
 </lemme>
</phrase>
```

```
<content>Aucune qualité d'écriture.</content>
<phrase id="1">
 <lemme pos='PRO' neg='1' category='négation' subcategory='nég' type='shifter'>aucun</lemme>
 <lemme pos='NOM' neg='1' category='évaluation' type='jugement' pol='pos' strenght='1'>
  qualité
 </lemme>
 <lemme pos='PREP' neg='0'>de</lemme>
 <lemme pos='NOM' neg='0'>écriture</lemme>
</phrase>
```

2. Notre traitement des négations est un simple renversement de polarité. Ceci n'est évidemment pas satisfaisant pour des expressions d'opinion forte, comme *excellent* où la négation n'exprime pas une opinion négative. Voir (Benamara *et al.*, 2012) pour une analyse fine des effets de la négation sur les expressions d'opinion.

```
<content>Je n'ai pas tout compris mais j'ai été sensible à la folie de ce poème.</content>
<phrase id="1">
 <lemme pos='PRO' neg='0'>je</lemme>
 <lemme pos='ADV' neg='1' category='négation' subcategory='nég' type='shifter'>ne</lemme>
 <lemme pos='V' neg='1'>avoir</lemme>
 <lemme pos='ADV' neg='1' category='négation' subcategory='nég' type='shifter'>pas</lemme>
 <lemme pos='ADV' neg='0'>tout</lemme>
 <lemme pos='V' neg='0' category='louer' type='jugement' pol='pos' strength='1'>
  comprendre
 </lemme>
 <lemme pos='CONJ' neg='0'>mais</lemme>
 <lemme pos='PRO' neg='0'>je</lemme>
 <lemme pos='V' neg='1'>avoir</lemme>
 <lemme pos='V' neg='0'>été</lemme>
 <lemme pos='ADJ' neg='0' category='évaluation' type='jugement' pol='pos_neutre_neg'
 strength='1'>
  sensible
 </lemme>
 <lemme pos='PREP' neg='0'>à</lemme>
 <lemme pos='DET' neg='0'>le</lemme>
 <lemme pos='NOM' neg='0' category='évaluation' type='jugement' pol='neg' strength='1'>
  folie
 </lemme>
 <lemme pos='PREP' neg='0'>de</lemme>
 <lemme pos='PRO' neg='0'>ce</lemme>
 <lemme pos='NOM' neg='0'>poème</lemme>
</phrase>
```

3 Prédiction des notes attribuées par les relecteurs et consensus

Nous avons utilisé une approche par apprentissage automatique pour prédire les notes des relectures (tâche 2) en s'appuyant sur les informations fournies par la projection du lexique sur les commentaires. Pour le calcul du consensus, nous avons considéré, comme cela était indiqué dans la définition de la tâche 3, qu'une nouvelle fait consensus auprès des relecteurs si les notes prédites lors de la tâche 2 ne varient pas au-delà d'un écart de 1 point.

3.1 Traits et classifieurs utilisés

A partir des informations fournies par la projection du lexique sur les commentaires, nous avons défini un certain nombre de traits de plusieurs types pour chaque relecture :

- **les traits stylistiques** :
  - présence et nombre de ponctuations (?, !, "", ...), de répétition de caractères (comme *supeeerrr*) et de mots en majuscule ;

- **les traits lexicaux** :
  - nombre de phrases,
  - nombre de mots,
  - nombre de négations,
  - présence et nombre de marqueurs de contrastes (comme *mais, en revanche, cependant, etc.*). Les connecteurs discursifs de contraste ont été extraits du lexique LEXCONN (Roze *et al.*, 2012),
  - nombre de modalité. Nous considérons ici différents types de modalité : les adverbes de doute et les adverbes d'affirmation ainsi que des verbes à emploi modal comme *espérer, pouvoir, devoir, supposer, croire, etc.* ;

– **les traits concernant directement les opinions** :
  – présence et nombre de mots ou expressions de type *reportage*,
  – nombres de polarités positives/négatives/neutres/ambiguës. En plus de la projection des mots du lexique, nous avons considéré des heuristiques simples à base de règles pour renverser la polarité de mots sous la portée d'une négation ou pour prendre en compte de nouveaux mots non présents dans le lexique. Par exemple, soit m un mot non présent dans le lexique et soit o_+ un mot d'opinion de polarité positive présent dans le lexique. Nos règles permettent de gérer des cas comme :
  si $neg(o_+)$ alors renverser la polarité de o
  si $(o_+)\ CONJ\ m$ avec $CONJ = \{et,\ ,\ , etc.\}$ alors m est un mot d'opinion positif
  si $(o_+)\ CONJ\ m$ avec $CONJ = \{mais, cependant, etc.\}$ alors m est un mot d'opinion négatif

– **les traits indiquant des modifications de polarité des opinions** :
  – présence et nombre d'adverbes d'intensité (intensifieur ou atténuateur),
  – nombres de polarités positives/négatives/neutres/ambiguës modifiées par un intensifieur,
  – nombres de polarités positives/négatives/neutres/ambiguës modifiées par un atténuateur,
  – nombres de polarités neutres/ambiguës dans la portée d'une négation,
  – nombre de mots ou expressions d'opinion ayant une force maximale (i.e. strenght=3) de polarité positive, négative ou neutre.

3.2 Phase d'apprentissage

Le tableau 3 présente la répartition de chaque classe dans le corpus d'apprentissage pour chaque tâche.

Classes pour la tâche 2	**Nombre d'instances**
1	405
2	2846
3	9068
4	13809
5	12622
Classes pour la tâche 3	**Nombre d'instances**
Consensus (1)	5331
Absence de consensus (0)	4020

TABLE 3 – Répartition de chaque classe dans le corpus d'apprentissage pour chaque tâche.

Durant la phase d'apprentissage, nous avons divisé le corpus fourni en un corpus d'entraînement (environ 2/3 du corpus) et un corpus de test.

Pour la tâche de prédiction des notes, nous avons testé plusieurs classifieurs de Weka (Hall *et al.*, 2009) avec plusieurs combinaisons de traits. Le classifieur ayant obtenu les meilleurs résultats est la régression logistique avec les paramètres par défaut.

Nous présentons ici les meilleures combinaisons de traits que nous avons testées ainsi que les résultats obtenus sur les données d'entraînement.

3.2.1 Combinaison 1

Les traits utilisés pour ce test sont :
– **traits lexicaux** : nombre de mots, nombre de négations, nombre de contrastes ;
– **traits concernant directement les opinions** : nombres de polarités positives et négatives ;
– **traits indiquant des modifications de polarité des opinions** : nombres de polarités positives/négatives modifiées par un intensifieur, nombres de polarités positives/négatives modifiées par un atténuateur.

Avec ces traits, l'accuracy obtenue sur la tâche 2 est de 45,64 %. Le tableau 4 montre les résultats détaillés sur chacune des classes à prédire.

Note	Rappel	Précision	F-mesure
1	0.005	1	0.01
2	0.06	0.442	0.106
3	0.463	0.435	0.448
4	0.276	0.422	0.334
5	0.729	0.482	**0.58**
Moyenne	0.456	0.453	0.425

TABLE 4 – Résultats obtenus sur les données d'entraînement pour la combinaison de traits 1.

Pour la tâche 3, la précision est de 63 %.

3.2.2 Combinaison 2

Les traits utilisés ici sont :
– **traits lexicaux** : nombre de négations, nombre de contrastes ;
– **traits concernant directement les opinions** : nombres de polarités positives/négatives/neutres ;
– **traits indiquant des modifications de polarité des opinions** :
  – nombres de polarités positives/négatives/neutres modifiées par un intensifieur,
  – nombres de polarités positives/négatives/neutres modifiées par un atténuateur,
  – nombre d'expression de force maximale de polarités positives/négatives/neutres.

Avec ces traits, l'accuracy obtenue sur la tâche 2 est de 46,47 %. Le tableau 5 montre les résultats détaillés sur chacune des classes à prédire.

Note	Rappel	Précision	F-mesure
1	0	0	0
2	0.06	0.462	0.106
3	0.464	0.438	0.451
4	0.317	0.426	0.363
5	0.712	0.498	**0.586**
Moyenne	0.465	0.452	0.438

TABLE 5 – Résultats obtenus sur les données d'entraînement pour la combinaison de traits 2.

Pour la tâche 3, la précision est de 64 %.

3.2.3 Combinaison 3

Tous les traits sont utilisés ici. Avec ces traits, l'accuracy obtenue sur la tâche 2 est de 45,23 %. Le tableau 6 montre les résultats détaillés sur chacune des classes à prédire.

Note	Rappel	Précision	F-mesure
1	0.021	0.5	0.04
2	0.068	0.413	0.117
3	0.473	0.457	**0.708**
4	0.372	0.431	0.399
5	0.66	0.466	0.546
Moyenne	0.452	0.447	0.43

TABLE 6 – Résultats obtenus sur les données d'entraînement pour la combinaison de traits 3.

Pour la tâche 3, la précision est de 62 %.

4 Résultats

Pour la phase de test, nous avons appris les modèles sur l'intégralité du corpus d'entraînement et testé le classifieur *régression logistique* avec les 3 combinaisons de traits sur le corpus de test. Nous avons ainsi soumis 3 runs pour chacune des deux tâches.

Le tableau 7 présente la répartition de chaque classe dans le corpus de test pour chaque tâche. La répartition dans le corpus de test est semblable à celle observée dans le corpus d'entraînement.

Classes pour la tâche 2	**Nombre d'instances**
1	179
2	1213
3	3850
4	5981
5	5562
Classes pour la tâche 3	**Nombre d'instances**
Consensus (1)	2150
Absence de consensus (0)	1854

TABLE 7 – Répartition de chaque classe dans le corpus de test pour chaque tâche.

4.1 Tâche 2

Pour la tâche 2, la mesure d'évaluation utilisée est l'EDRM (Exactitude en Distance Relative à la solution Moyenne) (Grouin *et al.*, 2013). Cette mesure permet par exemple de pénaliser plus un système qui prédirait une note de 1 au lieu de 5 qu'un système qui prédirait une note de 4 au lieu de 5. Les résultats obtenus sont les suivants :
– combinaison de traits 1 : 0,8193
– combinaison de traits 2 : 0.8217
– combinaison de traits 3 : **0.8267**
C'est donc la combinaison 3 qui obtient les meilleurs résultats pour cette tâche.

D'après les résultats officiels, l'EDRM sur la médiane obtenue par le deuxième meilleur participant est de 0,3975, ce qui nous place donc en première position du défi pour la tâche 2.

4.2 Tâche 3

Pour la tâche 3, la mesure d'évaluation utilisée est la précision. Les résultats obtenus sont les suivants :
– combinaison de traits 1 : 0,6453
– combinaison de traits 2 : **0.6473**
– combinaison de traits 3 : 0.6401
C'est donc la combinaison 2 qui obtient les meilleurs résultats pour cette tâche.

D'après les résultats officiels, la précision obtenue par le deuxième meilleur participant est de 0,3776, ce qui nous place donc aussi en première position du défi pour la tâche 3.

Les résultats pour les deux tâches sont comparables à ceux obtenus lors de la phase d'entraînement.

5 Conclusion

Dans cet article, nous avons présenté les expériences et tests effectués dans le cadre du défi DEFT 2014 pour les tâches 2 et 3. Nous avons utilisé une approche par apprentissage automatique qui permet, à partir d'informations sémantiques fines fournies par un lexique d'expressions d'opinion, de prédire les notes que donneraient des relecteurs à partir des commentaires textuels qu'ils ont rédigés. Nous avons obtenu des résultats très encourageants mais n'avons pas exploité

toutes les possibilités offertes par le lexique. En effet, pour améliorer ces résultats, nous pouvons envisager de tenir compte les catégories morphosyntaxiques des mots mais surtout d'utiliser l'intensité des expressions d'opinions en plus de leur polarité afin de prédire les différentes classes de notes plus finement. Enfin, nous envisageons de tester cette approche sur d'autres domaines pour vérifier la généricité du lexique.

Références

AÏT-MOKHTAR S., CHANOD J.-P. & ROUX C. (2002). Robustness beyond Shallowness : Incremental Deep Parsing. *Natural Language Engineering*, **8**, 121–144.

ASHER N., BENAMARA F. & MATHIEU Y. (2009). Appraisal of Opinion Expressions in Discourse. *Linguisticae Investigationes 32 :2*.

ASHER N., BENAMARA F. & MATHIEU Y. Y. (2008). Categorizing Opinions in Discourse. In *Actes de ECAI*.

BENAMARA F., CHARDON B., MATHIEU Y., POPESCU V. & ASHER N. (2012). How do Negation and Modality Impact on Opinions ? (regular paper). In *Extra-propositional aspects of meaning in computational linguistics - Workshop at ACL 2012, Jeju Island, Korea, 13/07/2012* : Association for Computational Linguistics (ACL).

GROUIN C., ZWEIGENBAUM P. & PAROUBEK P. (2013). DEFT2013 se met à table : présentation du défi et résultats. In *Actes du 9ème Défi Fouille de Texte, DEFT2013*, Les Sables d'Olonne, France.

HALL M., FRANK E., HOLMES G., PFAHRINGER B., REUTEMANN P. & WITTEN I. H. (2009). The WEKA Data Mining Software : An Update. *SIGKDD Explorations*, **11, Issue 1**.

LEVIN B. (1993). *English Verb Classes and Alternations : A Preliminary Investigation*. University of Chicago Press.

MATHIEU Y. Y. (1999). Sémantique lexicale et grammaticale. *Langage*, **136**.

MATHIEU Y. Y. (2005). A Computational Semantic Lexicon of French Verbs of Emotion. In *Computing Attitude and Affect in Text : Theory and Applications*, Dordrecht, The Netherlands.

ROZE C., DANLOS L. & MULLER P. (2012). LEXCONN : a French Lexicon of Discourse Connectives. *Discours*, **10**.

WIERZBICKA A. (1987). *Speech Act Verbs*. Sydney Academic Press.