

Un système hybride pour l'analyse de sentiments associés aux aspects

Caroline Brun¹ Diana Nicoleta Popa¹ Claude Roux¹
(1) XRCE, 6 chemin de Maupertuis, 38240 Meylan France
{Caroline.Brun,Diana.Popa,Claude.Roux}@xrce.xerox.com

Résumé. Cet article présente en détails notre participation à la tâche 4 de SemEval2014 (Analyse de Sentiments associés aux Aspects). Nous présentons la tâche et décrivons précisément notre système qui consiste en une combinaison de composants linguistiques et de modules de classification. Nous exposons ensuite les résultats de son évaluation, ainsi que les résultats des meilleurs systèmes. Nous concluons par la présentation de quelques nouvelles expériences réalisées en vue de l'amélioration de ce système.

Abstract.

An Hybrid System for Aspect-Based Sentiment Analysis

This paper details our participation to the SemEval2014 task 4 (Aspect Based Sentiment Analysis). We present the shared task, and then describe precisely our system, which is a combination of natural processing components and classification modules. We also present its evaluation results and the best system results. We finally expose some new experiments aiming at improving the system.

Mots-clés : Analyse de sentiments associés aux aspects, SemEval2014, système hybride.

Keywords: Aspect Based Sentiment Analysis, SemEval2014, hybrid system.

1 Introduction

L'analyse de sentiments associés aux aspects vise à découvrir les sentiments exprimés par un utilisateur sur les différents aspects d'une entité donnée, (Hu & Liu, 2004), (Liu, 2012). Récemment, la campagne d'évaluation "SemEval" a proposé une tâche d'analyse de sentiments associés aux aspects, (Pontiki *et al.*, 2014), à laquelle nous avons participé. Nous présentons le système que nous avons développé et les résultats obtenus dans le cadre de la campagne (Brun *et al.*, 2014). Pour cet article, en plus de présenter la tâche 4 de la campagne SemEval 2014, nous avons voulu reprendre le travail que nous avons fait pour SemEval afin d'évaluer les points forts et les points faibles de notre système, et de le comparer aux autres systèmes présentés, avec comme objectif l'amélioration des résultats. Pour cela, nous avons réalisé un nouvel ensemble d'expériences pour déterminer les parties les plus pertinentes à modifier.

2 Présentation de la tâche 4 de SemEval 2014

L'analyse de sentiments est une tâche qui connaît actuellement un engouement certain. Cependant, la majorité des approches actuelles tente de détecter la polarité globale d'une phrase ou d'un document dans son ensemble, indépendamment des entités mentionnées (les concepts du domaine visé, par exemple, ordinateurs portables, restaurants, films) et leurs aspects (par exemple la batterie, l'écran, la nourriture, le service, ...). L'objet de la tâche 4 de la campagne d'évaluation SemEval2014 était précisément de s'atteler à l'analyse d'opinions associées aux aspects, c'est-à-dire à détecter les aspects cibles des opinions. Les corpus fournis pour cette tâche étaient constitués de commentaires d'internautes annotés en aspects et polarités. Plus exactement, la tâche était subdivisée en 4 sous-tâches :

1. Extraction des termes dénotant les aspects : par exemple, dans la phrase *I liked the service and the staff, but not the food*, les termes à détecter sont *service*, *staff* et *food*.

2. Extraction des catégories sémantiques dénotant les aspects : étant donné un ensemble prédéfini de catégories par domaine ({"price", "food", "service", "ambiance", et "anecdote"} dans le domaine des restaurants), associer ces catégories aux phrases ; leur granularité est moins fine que celle des termes de la sous-tâche précédente et elles ne sont pas nécessairement associées à la présence de termes dans la phrase. Par exemple :
 - “The restaurant was too expensive” → {price}
 - “The restaurant was expensive, but the menu was great” → {price, food}
3. Extraction de la polarité associée aux termes précédemment détectés : la polarité prend ici 4 valeurs :{positif, négatif, neutre et conflit}.
 - “I loved their **fajitas**” → {fajitas : **positive**}
 - “I hated their **fajitas**, but their **salads** were great” → {fajitas : **negative**, salads : **positive**}
 - “The **fajitas** are their first plate” → {fajitas : **neutral**}
 - “The **fajitas** were great to taste, but not to see” → {fajitas : **conflit**}
4. Extraction de la polarité associée aux catégories précédemment détectées : Par exemple :
 - “The restaurant was too expensive” → {price : **negative**}
 - “The restaurant was expensive, but the menu was great” → {price : **negative**, food : **positive**}

Deux jeux de données étaient disponibles, l’un concernant des revues d’ordinateurs portables, annotées seulement pour les sous-tâches 1 et 3, l’autre concernant des revues de restaurants, annotées pour les 4 sous-tâches. Nous nous sommes concentrés sur ce dernier domaine afin d’aborder l’ensemble des sous-tâches. Le corpus d’entraînement comprenait 3044 phrases annotées avec 3700 termes et leur polarité et 3715 catégories sémantiques et leur polarité.

3 Notre système d’analyse de sentiments associés aux aspects

3.1 Le système pré-existant

Nous avons d’abord adapté le système d’analyse d’opinions que nous avons développé préalablement (Brun, 2011, 2012). Ce système est fondé sur un analyseur syntaxique robuste XIP, (Ait-Mokhtar *et al.*, 2001), utilisé pour calculer des relations d’opinions, en combinant dépendances syntaxiques et informations lexicales sur la polarité des mots et la sous-catégorisation des prédicats. Ce système génère des dépendances sémantiques appelées SENTIMENT qui sont soit binaires, c’est-à-dire reliant les prédicats de polarité et les cibles des opinions, soit unaires, lorsque la cible de l’opinion n’a pu être détectée. Par exemple, l’analyse de *I was very disappointed by the food and the service*, produit :

SUBJ_N(disappointed, food), SUBJ_N(disappointed, service), OBJ_N(disappointed,I), MANNER_PRE(disappointed, very), SENTIMENT_NEGATIVE(disappointed, service), SENTIMENT_NEGATIVE(disappointed, food)

Ce système préexistant n’extrait pas explicitement les termes dénotant les aspects du domaine, cependant les arguments cibles de la dépendance "SENTIMENT" désignent potentiellement des aspects. De plus, ce système n’extrait que les opinions positives ou négatives, mais ne couvre pas la polarité neutre, ni le conflit. Le système utilisé pour la campagne SemEval est celui de l’anglais, mais un système équivalent est également disponible pour le français.

3.2 Adaptation pour SemEval2014

Ce système a été adapté pour les besoins de SemEval2014 selon deux axes : acquisition lexicale afin de couvrir les termes du domaine, et développement de règles de détection des termes multi-mots et de règles d’extraction d’opinions.

3.2.1 Acquisition lexicale et détection de termes

Le lexique de notre système inclut préalablement du vocabulaire de polarité, tandis que la tâche de détection de termes de SemEval implique des connaissances lexicales du domaine. Nous avons donc encodé du vocabulaire de domaine concernant les revues de restaurant, en utilisant le corpus d’entraînement, les termes y étant explicitement balisés, et en lui associant les traits sémantiques correspondant aux aspects (*food, service, ambiance, price, anecdote*). Cette liste a

ensuite été étendue par filtrage de vocabulaire extrait du portail Wikipedia dédié à la nourriture¹. Cela nous a permis d'encoder 761 termes reliés à la nourriture, 31 termes reliés au prix, 105 termes reliés à l'ambiance, et 42 termes reliés au service. Pour détecter les termes complexes, des grammaires locales ont été construites sur la base de ce vocabulaire. Les règles, de type "expressions régulières", détectent les termes à mot multiples, comme *Grilled Chicken special with Edamame Puree* ou encore *staff members* et leur associent une catégorie sémantique.

3.2.2 Adaptation de la grammaire pour la détection de la polarité

La grammaire a été également modifiée pour détecter les opinions liées aux termes et aux aspects sémantiques : si un terme est argument d'une dépendance SENTIMENT, deux nouvelles dépendances sont créées, une associée au terme (OPINION_ON_TERM) et une associée à la catégorie sémantique correspondante (OPINION_ON_CATEGORY) Elles héritent de la polarité de la dépendance SENTIMENT. Si deux dépendances SENTIMENT ciblent un même terme avec des polarités opposées, elles sont alors créées avec le trait de polarité CONFLIT. Si un terme est détecté sans être cible de la dépendance SENTIMENT, elles sont créées avec le trait de polarité "neutre". Enfin, si aucun terme n'est détecté dans la phrase, OPINION_ON_CATEGORY est créée avec la cible "anecdote" (= aspect "fourre-tout"), héritant de la polarité calculée dans la phrase, ou bien du neutre. La dépendance OPINION_ON_TERM relie les termes et leur polarité et constitue la base des sous-tâches 1 et 3.

3.3 Classification

Tandis que la détection des termes et des polarités qui leur sont associées est effectuée par la grammaire, la détection des catégories sémantiques (aspects) et de leur polarité est réalisée par classification automatique via la bibliothèque "LibLinear", (Fan *et al.*, 2008). Nous entraînons tout d'abord un classifieur unique pour associer leurs catégories sémantiques aux phrases, puis pour chaque catégorie, nous entraînons un classifieur de polarité. Dans les deux cas, nous appliquons une validation croisée ("10-fold cross-validation"), qui nous a permis de sélectionner le classifieur de meilleure performance en fonction de la sous-tâche. Nous avons retenu cette méthode pour pallier les erreurs possibles de la grammaire. De cette façon, les informations issues de la grammaire sont utilisées comme des attributs à la fois dans la phase d'entraînement et de prédiction, mais ne sont retenues que les valeurs renvoyées par les classifieurs comme résultats finaux.

3.3.1 Classification des aspects

Ce module associe une ou plusieurs catégories sémantiques aux phrases des revues de restaurant. Pour chaque phrase, le module utilise comme traits le "sac de mots" de la phrase ainsi que des informations fournies par l'analyseur syntaxique. Dans l'étape de prétraitement, les mots outils (déterminants, conjonctions) sont éliminés. Nous utilisons la régression logistique ("L2-regularized") de LibLinear, pour créer le modèle. Les traits considérés sont les lemmes des mots avec leurs fréquences. De plus, l'information fournie par l'analyseur syntaxique est utilisée pour incrémenter la fréquence des termes appartenant aux catégories détectées. Cette information se compose des dépendances associant leurs catégories aux termes et des dépendances d'opinions concernant catégories et termes. Par exemple, pour la phrase suivante : *"Fabulous service, fantastic food, and a chilled out atmosphere and environment"*, les dépendances retenues sont :

SERVICE(service), FOOD(food) AMBIENCE(atmosphere), AMBIENCE(environment)
 OP_CAT_POS(food), OP_CAT_POS (service), OP_CAT_POS (ambiance),
 OP_TERM_POS(food), OP_TERM_POS(service), OP_TERM_POS(atmosphere), OP_TERM_POS(environment).

Ceci nous conduit à augmenter les fréquences de la façon suivante : food (+3), service (+3), atmosphere (+2), environment (+2), ambiance (+1). La régression logistique associe alors chaque catégorie à une certaine probabilité. Nous imposons un seuil concernant les valeurs des probabilités des catégories à retenir ; après plusieurs expérimentations le seuil (borne inférieure) que nous avons retenu est de 0,25.

1. http://en.wikipedia.org/wiki/Category:Listes_of_foods

3.3.2 Classification des polarités associées aux catégories

L'approche pour le calcul des polarités est assez similaire, à quelques détails près. Nous utilisons également les traits correspondant au sac de mots, et la polarité fournie par les dépendances suivantes : OPINION_ON_CATEGORY et SENTIMENT. Lorsque ces dépendances sont détectées, un trait de la forme `polarity_category` est ajouté pour la classification. Ainsi pour l'exemple précédent : "*Fabulous service, fantastic food, and a chilled out atmosphere and environment*", les dépendances supplémentaires considérées sont :

SENTIMENT_POSITIVE(atmosphere, chilled out), SENTIMENT_POSITIVE(food, fantastic),
SENTIMENT_POSITIVE(service, fabulous).

Après appariement des catégories avec les termes correspondants, les traits supplémentaires sont : *positive_ambience*, *positive_food* et *positive_service*. Puisque la dépendance OPINION_ON_CATEGORY est également détectée par l'analyseur, chacun des traits mentionnés ci-dessus aura une fréquence de 2. De plus, la polarité seule est également ajoutée comme trait. L'entraînement des modèles est réalisé avec le solveur SVM "L2-regularized L2-loss" de la bibliothèque Liblinear. Un modèle de classification en polarité est produit pour chaque catégorie. Ainsi, selon les catégories préalablement associées à une phrase donnée, le modèle correspondant est employé pour en prédire la polarité.

3.3.3 Correction de la polarité des termes

Comme précédemment évoqué, les termes et leur polarité sont détectés par la grammaire, la polarité *neutre* étant affectée par celle-ci lorsque ni *positif* ou *negatif* n'est associé. Nous avons utilisé les résultats de la classification des polarités des catégories pour corriger celle des termes, dans le cas des neutres : ainsi, si la classification associe à la catégorie *food* la polarité *positif* dans une phrase pour laquelle seuls des termes de polarité *neutre* sont détectés, ils sont corrigés et prennent la polarité *positif*. Cette correction *a posteriori*, possible du fait que les termes sont associés à leur catégories sémantiques par notre grammaire, s'est montré extrêmement performante (voir table 3).

3.4 Evaluation

Le corpus de test de SemEval14 contient 800 phrases, 1134 occurrences de termes et leur polarité (555 termes distincts), 1025 catégories et leur polarité. L'évaluation s'est déroulée en deux phases : la phase A consistait à annoter les termes et catégories, puis un corpus corrigé pour les termes et les catégories était renvoyé pour la phase B, concernant l'annotation des polarités.

3.4.1 Détection des termes et des catégories (Phase A)

Les mesures considérées pour ces deux sous-tâches étaient les mesures classiques de précision, rappel et F-mesure. Les tableaux 1 et 2 présentent les résultats de notre système (XABSA) pour la détection des termes et des catégories. Les mesures de référence ("baseline") sont décrites en détails dans l'article présentant la tâche (Pontiki *et al.*, 2014). La "baseline" pour la détection des termes consiste simplement à contruire un lexique des termes présents dans le corpus d'entraînement et à les repérer dans le corpus de test. Pour la détection des des catégories, chaque phrase du test est associée au k-phrases les plus proches du corpus d'entraînement (la similarité étant calculée avec le coefficient de Dice). Les catégories les plus fréquentes de ces k phrases sont ensuite associées aux phrases du test.

Méthode	Précision	Rappel	F-Mesure
Baseline	0,627329	0,376866	0,470862
XABSA	0,862453	0,818342	0,839818

TABLE 1: Détection des termes

Méthode	Précision	Rappel	F-Mesure
Baseline	0,637500	0,483412	0,549865
Sac de mots	0,77337	0,799024	0,785988
XABSA	0,832335	0,813658	0,822890

TABLE 2: Détection des catégories

On constate que la combinaison de l'approche "sac-de-mots" avec les sorties de l'analyseur syntaxique permet une amélioration notable des performances pour la détection des catégories. Pour les deux tâches de détection des termes et des

catégories, notre système surpasse largement la référence, se classant parmi les 3 premiers de la compétition pour le corpus des revues de restaurants.

3.4.2 Détection de la polarité (Phase B)

Similairement, les tableaux 3 et 4 décrivent les résultats du système sur la détection de la polarité pour les termes et les catégories. Ici la mesure utilisée est l'exactitude ("accuracy"). Les méthodes de calcul de la référence (baseline) sont également décrites en détails dans (Pontiki *et al.*, 2014). Pour les polarités des termes et des catégories, la polarité la plus fréquente des k-phrases les plus proches du corpus d'entraînement est associée aux phrases du test.

Méthode	Exactitude
Baseline	0,55
XABSA sans correction des termes	0,66
XABSA	0,78

TABLE 3: Polarité des termes

Méthode	Exactitude
Baseline	0,56
Sac de mots	0,68
XABSA	0,78

TABLE 4: Polarité des catégories

A nouveau, notre système s'est bien classé dans la compétition, avec une exactitude globale de 0,78 pour la détection de la polarité des termes et de 0,78 pour la détection de la polarité des catégories. Pour cette dernière sous-tâche, l'utilisation de traits provenant de l'analyse syntaxique combinés au sac de mots se montre là aussi très performante.

3.4.3 Comparaison avec les autres systèmes

32 équipes ont participé à cette campagne d'évaluation. Concernant la détection des termes, le meilleur système, DLIREC (Toh & Wang, 2014) est un tagger CRF, avec une F-mesure de 0,8401, juste devant notre système. Concernant la détection des aspects, le meilleur système, proposé par NRC-Canada (Kiritchenko *et al.*, 2014), obtient la F-mesure de 0,8857 avec 5 classifieurs SVM binaires (1 par aspect), entraînés avec divers n-grams et de l'information lexicale apprise sur le corpus YELP de revues de restaurants. Notre système obtient ici le 3ème F-score. Pour la détection des polarités des termes, DCU (Wagner *et al.*, 2014) et NRC-Canada obtiennent les meilleures exactitudes, 80,95 et 80,15 respectivement. Ces deux systèmes utilisent à nouveau des classifieurs SVM enrichis avec des lexiques de polarité publiquement disponibles, de l'information syntaxique. Là aussi, notre système obtient le 3ème score. Enfin pour la détection des polarités associées aux aspects, NRC-Canada obtient le meilleur score avec une exactitude de 82,92 en utilisant un SVM enrichi pour capturer l'information relative à chaque aspect. Notre système est ici second.

D'une manière générale, les méthodes les plus performantes combinent des méthodes de classification "état de l'art" avec de l'information linguistique plus ou moins sophistiquée, certains systèmes se restreignant aux ressources fournies (systèmes contraints), d'autre utilisant des corpus additionnels (systèmes non contraints).

3.4.4 Analyse des erreurs

Si notre système a obtenu de bonnes performances, et ce, sur l'ensemble des 4 sous-tâches, les résultats détaillés en montrent les forces et les faiblesses, en particulier pour le calcul de la polarité, c.f. tableaux 5 et 6. On constate tout d'abord

Polarité	Préc.	Rappel	F-Mes.
Conflit	NaN (0/0)	NaN (0/0)	NaN
Négatif	0,78(143/182)	0,73 (143/196)	0,76
Positif	0,79 (675/845)	0,92 (675/728)	0,86
Neutre	0,58 (62/107)	0,31 (62/196)	0,41

TABLE 5: Résultats par polarité (termes)

Polarité	Préc.	Rappel	F-Mes.
Conflit	0,5(7/14)	0,13(7/52)	0,21
Négatif	0,73(151/208)	0,68(151/222)	0,70
Positif	0,83(599/720)	0,91(599/657)	0,87
Neutre	0,50(42/83)	0,45(42/94)	0,47

TABLE 6: Résultats par polarité (catégories)

que la polarité *conflit* est particulièrement difficile à détecter. En effet, elle est à la fois mal couverte par notre grammaire, car les conflits mettent souvent en jeu des éléments à longue distance, et également assez peu représentée dans le corpus, ce qui pose problème pour l'apprentissage. La polarité *neutre* présente également des résultats relativement faibles : nous pensons que c'est lié au fait que la grammaire l'affecte par défaut, si *positif* ou *négatif* ne sont pas associés à un aspect.

4 Nouvelles expériences

À la suite de SemEval2014, nous avons voulu améliorer les performances de notre système, en particulier pour la phrase B pour lesquelles les résultats étaient un peu plus faibles pour la détection de la polarité des termes et des catégories. Nous avons donc mis en place de nouvelles expériences en intégrant les éléments les plus performants utilisés par les autres équipes lors de la campagne d'évaluation.

Pour les premières expériences, nous avons rajouté le "NRC emotion lexicon", (Mohammad & Turney, 2010) utilisé parmi d'autres par NRC-Canada lors de la campagne, comme lexique supplémentaire de polarité. Nous avons aussi tenté d'intégrer les "ngrams" dans notre système, à l'instar de nombres d'équipes qui ont eu de bons résultats. Nous nous sommes aussi intéressés à la notion de distance entre un mot et un terme de polarité (positif ou négatif) ainsi que la notion de distance entre un terme du domaine (*food, service,...*) et un terme de polarité (positif ou négatif).

Les tableaux 7 et 8 synthétisent les résultats de ces expériences pour la polarité des termes et des catégories sémantiques. Les résultats étant très proches, nous donnons, en plus de l'exactitude, le nombre d'occurrences correctes par rapport au nombre total.

Méthode	Exactitude
E0 : XABSA	0,787 (893/1134)
E1 : XABSA+lexique	0,76 (863/1134)
E2 : XABSA+bigrams	0,789 (895/1134)
E3 : XABSA+bigrams+distance	(895/1134)

TABLE 7: Polarité des termes

Méthode	Exactitude
E0 : XABSA	0,787 (807/1025)
E1 : XABSA+lexique	0,78 (788/1025)
E2 : XABSA+bigrams	0,798 (818/1025)
E3 : XABSA+bigrams +distance	0,80 (820/1025)

TABLE 8: Polarité des catégories

L'intégration du nouveau lexique dégrade nos résultats. Ce lexique a été créé via "crowdsourcing", et par exemple, associe le mot *food* à la polarité *positif*. Pour cela, il s'avère être difficilement intégrable directement à notre composant symbolique. L'utilisation des "ngrams" est une piste plus prometteuse, le meilleur résultat correspondant à l'intégration des "bigrams", pour lesquels on observe une nette amélioration des résultats. Concernant la notion de distance, nous rapportons également les meilleurs résultats qui correspondent à la distance entre termes du domaines et mots positifs ou négatifs. Ils améliorent légèrement l'expérience précédente. Ces expériences préliminaires montrent que des améliorations sont possibles, même si l'incrément est relativement réduit.

5 Conclusion

Dans cet article, nous sommes revenus en détails sur notre participation à la campagne d'évaluation SemEval14, pour la tâche 4 concernant l'analyse de sentiments associés aux aspects. Nous avons décrit en détails le système que nous avons conçu : il est fondé sur l'utilisation d'un composant symbolique permettant de détecter les termes du domaine et leur polarité et d'un composant de classification, qui classe les phrases selon leurs catégories sémantiques (aspects) et dans un second temps, leur associe une polarité. Nous montrons également comment les résultats de la classification de la polarité des catégories sémantiques peuvent améliorer a posteriori la polarité des termes. Ce système s'est montré performant, mais nous avons cherché à l'améliorer, en particulier pour certaines polarités, (*neutre* et *conflit*). Pour cela, nous avons réalisé quelques expériences préliminaires qui montrent qu'une amélioration est possible. Ceci constitue une des pistes de recherche future, mais nous souhaitons également travailler sur l'adaptation de ce système au français, pour lequel très peu de données annotées existent pour cette tâche, ainsi que sur l'application d'une méthodologie similaire pour la détection des émotions.

Références

- AIT-MOKHTAR S., CHANOD J.-P. & ROUX C. (2001). A multi-input dependency parser. In T. U. PRESS, Ed., *IWPT*.
- BRUN C. (2011). Detecting opinions using deep syntactic analysis. In *Proceedings of RANLP*, Hissar, Bulgaria.
- BRUN C. (2012). Learning opinionated patterns for contextual opinion detection. In *COLING 2012, 24th International Conference on Computational Linguistics, Proceedings of the Conference : Posters, 8-15 December 2012, Mumbai, India*, p. 165–174.

- BRUN C., POPA D. N. & ROUX C. (2014). Xrce : Hybrid classification for aspect-based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, p. 838–842, Dublin, Ireland : Association for Computational Linguistics and Dublin City University.
- FAN R.-E., CHANG K.-W., HSIEH C.-J., WANG X.-R. & LIN C.-J. (2008). Liblinear : A library for large linear classification. *Journal of Machine Learning Research*, **9**, 1871–1874.
- HU M. & LIU B. (2004). Mining and summarizing customer reviews. In *KDD*, p. 168–177.
- KIRITCHENKO S., ZHU X., CHERRY C. & MOHAMMAD S. (2014). Nrc-canada-2014 : Detecting aspects and sentiment in customer reviews. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, p. 437–442, Dublin, Ireland : Association for Computational Linguistics and Dublin City University.
- LIU B. (2012). *Sentiment Analysis and Opinion Mining*. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers.
- MOHAMMAD S. M. & TURNEY P. D. (2010). Emotions evoked by common words and phrases : Using mechanical turk to create an emotion lexicon. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, CAAGET '10, p. 26–34, Stroudsburg, PA, USA : Association for Computational Linguistics.
- PONTIKI M., GALANIS D., PAVLOPOULOS J., PAPAGEORGIOU H., ANDROUTSOPOULOS I. & MANANDHAR S. (2014). Semeval-2014 task 4 : Aspect based sentiment analysis. In *International Workshop on Semantic Evaluation (SemEval)*.
- TOH Z. & WANG W. (2014). Dlirec : Aspect term extraction and term polarity classification system. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, p. 235–240 : Association for Computational Linguistics.
- WAGNER J., ARORA P., CORTES S., BARMAN U., BOGDANOVA D., FOSTER J. & TOUNSI L. (2014). Dcu : Aspect-based polarity classification for semeval task 4. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, p. 223–229, Dublin, Ireland : Association for Computational Linguistics and Dublin City University.