

Bilingual Experiments on an Opinion Comparable Corpus

E. Martínez-Cámara	M. T. Martín-Valdivia	M. D. Molina-González	L. A. Ureña-López
SINAI research group	SINAI research group	SINAI research group	SINAI research group
University of Jaén	University of Jaén	University of Jaén	University of Jaén
E-23071, Jaén (Spain)	E-23071, Jaén (Spain)	E-23071, Jaén (Spain)	E-23071, Jaén (Spain)
emcamara@ujaen.es	maite@ujaen.es	mdmolina@ujaen.es	laurena@ujaen.es

Abstract

Up until now most of the methods published for polarity classification are applied to English texts. However, other languages on the Internet are becoming increasingly important. This paper presents a set of experiments on English and Spanish product reviews. Using a comparable corpus, a supervised method and two unsupervised methods have been assessed. Furthermore, a list of Spanish opinion words is presented as a valuable resource.

1 Introduction

Opinion Mining (OM) is defined as the computational treatment of opinion, sentiment, and subjectivity in text. The OM discipline combines Natural Language Processing (NLP) with data mining techniques and includes a large number of tasks (Pang and Lee, 2008). One of the most studied tasks is polarity classification of reviews. This task focuses on determining which is the overall sentiment-orientation (positive or negative) of the opinions contained within a given document.

Two main approaches are followed by researchers to tackle the OM task. On the one hand, the Machine Learning (ML) approach (also known as the supervised approach) is based on using a collection of data to train the classifiers (Pang et al., 2002). On the other hand, (Turney, 2002) proposed an unsupervised method based on the semantic orientation of the words and phrases in the reviews. Both methodologies have their advantages and drawbacks. For example, the ML approach depends on the availability of labelled data sets (training data), which

in many cases are impossible or difficult to achieve, partially due to the novelty of the task. On the contrary, the unsupervised method requires a large amount of linguistic resources which generally depend on the language, and often this approach obtains lower recall because it depends on the presence of the words comprising the lexicon in the document in order to determine the polarity of opinion.

Although opinions and comments on the Internet are expressed in any language, most of research in OM is focused on English texts. However, languages such as Chinese, Spanish or Arabic, are ever more present on the web. Thus, it is important to develop resources for these languages. The work presented herein is mainly motivated by the need to develop polarity classification systems and resources in languages other than English. We present an experimental study over the SFU Review Corpus (Taboada, 2008), a comparable corpus that includes opinions of several topics in English and in Spanish. We have followed this line of work: Firstly, we have taken as baseline a supervised experiment using Support Vector Machine (SVM). Then we have tried different unsupervised strategies. The first one uses the method presented in (Montejo-Ráez et al., 2012). This approach combines SentiWordNet scores with a random walk analysis of the concepts found in the text over the WordNet graph in order to determine the polarity of a tweet. This method obtained very good results in short texts (tweets) and so, we want to try it using larger document. Although we have carried out several experiments using different parameters and modifications, the results are not as good as we hoped. For this, we have

tried a very simple experiment using a list of opinionated words in order to classify the polarity of the reviews. For English we have used the Bin Liu English lexicon (BLEL) (Hu and Liu, 2004) and for Spanish we have automatically translated the BLEL lexicon into Spanish. In addition, we have also checked manually and improved the Spanish list.

The paper is organized as follows: Section 2 briefly describes papers that study non-English sentiment polarity classification and, specifically work related to Spanish OM. In Section 3 we explain the resources used in the unsupervised methods assessed. Section 4 presents the experiments carried out and discusses the main results obtained. Finally, we outline conclusions and further work.

2 Related Work

There are some interesting papers that have studied the problem using non-English collections. Dencke (2008) worked on German comments collected from Amazon. These reviews were translated into English using standard machine translation software. Then the translated reviews were classified as positive or negative, using three different classifiers: LingPipe7, SentiWordNet (Baccianella et al., 2010) with classification rule, and SentiWordNet with machine learning. Ghorbel and Jacot (2011) used a corpus with movie reviews in French. They applied a supervised classification combined with SentiWordNet in order to determine the polarity of the reviews. In (Rushdi-Saleh et al., 2011a) a corpus of movies reviews in Arabic annotated with polarity was presented and several supervised experiments were performed. Subsequently, they generated the parallel EVOCA corpus (English version of OCA) by translating the OCA corpus automatically into English. The results showed that they are comparable to other English experiments, since the loss of precision due to the translation process is very slight, as can be seen in (Rushdi-Saleh et al., 2011b).

Regarding Spanish, there are also some interesting studies. Banea et al. (2008) showed that automatic translation is a viable alternative for the construction of resources and tools for subjectivity analysis in a new target language. In (Brooke et al., 2009) several experiments are presented dealing with Spanish and English resources. They con-

clude that although the ML techniques can provide a good baseline performance, it is necessary to integrate language-specific knowledge and resources in order to achieve an improvement. Cruz et al. (2008) manually recollected the MuchoCine (MC) corpus to develop a sentiment polarity classifier based on the semantic orientation of the phrases and words. The corpus contains annotated Spanish movie reviews from the MuchoCine website. The MC corpus was also used in (Martínez-Cámara et al., 2011) to carry out several experiments with a supervised approach applying different ML algorithms. Finally, (Martín-Valdivia et al., 2012) also dealt with the MC corpus to present an experimental study of supervised and unsupervised approaches over a Spanish-English parallel corpus.

3 Resources for the unsupervised methods

In order to tackle the unsupervised experiments we have chosen several well-known resources in the OM research community. In addition, we have also generated a new Spanish linguistic resource.

Comparable corpora are those consisted of texts in two or more languages about the same topic, but they are not the translated version of the texts in the source language. For the experiments, we chose the comparable corpus SFU Review Corpus. The SFU Review Corpus is composed of reviews of products in English and Spanish. The English version (Taboada and Grieve, 2004) has 400 reviews (200 positive and 200 negative) of commercial products downloaded in 2004 from the Epinions web which are divided into eight categories: books, cars, computers, cookware, hotels, movies, music and phones. Each category includes 25 positive reviews and 25 negative reviews. Recently, the authors of SFU Review Corpus have made available the Spanish version of the corpus¹. The Spanish reviews are divided into the same eight categories, and also each category has 25 positive and 25 negative reviews.

In the unsupervised experiments we have analysed the performance of two approaches, the first one is based on lexicon and the other one in a graph-based method. We have selected the BLEL lexicon (Hu and Liu, 2004) to carry out the experiment based

¹<http://www.sfu.ca/~mtaboada/download/downloadCorpusSpa.html>

on lexicon on the English version of the corpus. The lexicon is composed by 6,787 opinion words that indicate positive or negative opinions, which 2,005 are positive and 4,782 are negative. With the aim of following the same approach over the Spanish version, firstly we have translated the BLEL lexicon with the Reverso machine translator, and then we have checked manually the resultant list. The Spanish Opinion Lexicon² (SOL) is composed by 2,509 positive and 5,627 negative words, thus in total SOL has 8,136 opinion words. If a review has more or the same positive words than negative the polarity is positive, otherwise negative.

The graph-based method is a modular system which is made up of different components and technologies. The method was first presented in (Montejo-Ráez et al., 2012) with a good performance over a corpus of English tweets. The main idea of the algorithm is to represent each review as a vector of polarity scores of the senses in the text and senses related to the context of the first ones. Besides, the polarity score is weighted with a measure of importance. Taking a review as input, the workflow of the algorithm is the following:

1. Disambiguation: To get the corresponding sense of the words that are in the text is required to disambiguate them. Thus, the output of this first step is one unique synset from WordNet³ (Miller, 1995) for each term. The input of the algorithm is the set of words with a POS-Tag allowed in WordNet. The graph nature of the WordNet structure is the basis for the UKB disambiguation method proposed by (Agirre and Soroa, 2009). The UKB disambiguation algorithm apply PageRank (Page et al., 1999) on the WordNet graph starting from term nodes, where each term node points to all its possible senses or synsets. The output of the process is a ranked list of synsets for each input word, and the highest rank synset is chosen as candidate sense.

For the Spanish disambiguation process we have chosen the Spanish WordNet version offered by the project Multilingual Central

²<http://sinai.ujaen.es/wiki/index.php/SOL>

³We have used the 3.0 release of WordNet.

Repository (MCR) (Gonzalez-Agirre et al., 2012). The Spanish WordNet of MCR has 38,702 synsets while WordNet has 117,659, i.e. the MCR covers the 32.89% of WordNet.

2. PPV: Once the synsets for the reviews are computed, the following step performs a second run of PageRank described in (Agirre and Soroa, 2009). Using the *Personalized PageRank*, a set of Personalized PageRank Vectors (PPVs) is obtained. This vector is a list of synsets with their ranked values. The key of this approach is to take from this vector additional synsets not related directly to the set of synsets disambiguated in the first step. The result is a longer list of pair $\langle \text{synset}, \text{weight} \rangle$ where the weight is the rank value obtained by the propagation of the weights of original synsets across the WordNet graph.
3. Polarity: The following step is to calculate the polarity score. For this purpose it is necessary a semantic resource to take the polarity score for each retrieved synset in the two previous steps. The semantic resource selected is SentiWordNet (Baccianella et al., 2010). According to these values, the three following equations have been applied to obtain the final polarity value:

$$p(r) = \frac{1}{|r|} \sum_{s \in r} \frac{1}{|s|} \sum_{i \in s} (p_i^+ - p_i^-) w_i \quad (1)$$

$$p(r) = \frac{1}{|r|} \sum_{s \in r} \frac{1}{|s|} \sum_{i \in s} f(p_i) \quad (2)$$

$$f(p_i) = \begin{cases} p_i^+ & \text{if } p_i^+ > p_i^- \\ p_i^- & \text{if } p_i^+ \leq p_i^- \end{cases}$$

$$p(r) = \frac{1}{|r|} \sum_{s \in r} \frac{1}{|s|} \sum_{i \in s} f(p_i) \quad (3)$$

$$f(p_i) = \begin{cases} 1 & \text{if } i \in [\text{positive words}] \\ -1 & \text{if } i \in [\text{negative words}] \\ p_i^+ & \text{if } p_i^+ > p_i^- \\ p_i^- & \text{if } p_i^+ \leq p_i^- \end{cases}$$

where $p(r)$ is the polarity of the review; $|r|$ is the number of sentences in the review r ; s is a sentence in r , being itself a set of synsets; i is a synset in s ; p_i^+ is the positive polarity of synset i ; p_i^- is the negative polarity of synset i and w_i is the weight of synset i .

4 Experiments and Results

Systems based on supervised approach are the most successfully in the OM literature. Therefore, we began the set of experiments applying a machine learning algorithm to the SFU corpus. Also, we have carried out a set of unsupervised experiments following a lexicon-based approach and a graph-based algorithm. For all the experiments the evaluation measures have been: precision, recall, F1 and Accuracy (Acc.). The validation approach followed for the supervised approach has been the well-known 10-cross-validation.

The algorithm chose for the supervised experiments is SVM (Cortes and Vapnik, 1995) because is one of the most successfully used in OM. LibSVM⁴ (Chang and Lin, 2011) was the implementation selected to carry out several experiments using SVM. We have evaluated unigrams and bigrams as minimum unit of information. Also, the influence of stemmer have been assessed. The weight scheme for representing each unit of information is TF-IDF. The same configuration has been applied to English and Spanish version of SFU corpus. Table 1 and Table 2 show the results for English version and Spanish version respectively.

	Precision	Recall	F1	Acc.
Unigrams	79.07%	78.50%	78.78%	78.50%
Unigrams & stemmer	79.82%	79.50%	79.66%	79.50%
Bigrams	78.77%	78.25%	78.51%	78.25%
Bigrams & stemmer	80.64%	80.25%	80.44%	80.25%

Table 1: SVM results for English SFU corpus

	Precision	Recall	F1	Acc.
Unigrams	73.65%	73.25%	73.45%	73.25%
Unigrams & stemmer	74.10%	73.75%	73.92%	73.75%
Bigrams	74.02%	73.50%	73.76%	73.50%
Bigrams & stemmer	73.90%	73.50%	73.70%	73.50%

Table 2: SVM results for Spanish SFU corpus

The results show one of the differences between the works published in SA, the use of unigrams or

⁴<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

bigrams. In (Pang et al., 2002) is asserted that the reviews should be represented with unigrams, but in (Dave et al., 2003) bigrams and trigrams outperformed the unigrams features. In our case, regarding the results reached without using a stemmer, the use of unigrams as minimum unit of information achieves better result than the use of bigrams when the language is English, but bigrams outperform unigrams when the texts are in Spanish. On the other hand, the best result both in English and Spanish is reached when a stemmer algorithm is applied. So, one conclusion of the supervised experiments is that the use of stemmer enhances the polarity classification in reviews. The following conclusion is that in English the presence of pair of words separate better the positive and negative classes, while in Spanish the use of unigrams is enough to classify the polarity when a stemmer algorithm is used.

The set of unsupervised experiments begins with a lexicon-based method. The method consists of find the presence in the reviews of opinion words which are included in a lexicon of opinion words. BLEL has been used for the English reviews, and SOL for the Spanish reviews. The results are presented in Table 3.

	Precision	Recall	F1	Acc.
BLEL lexicon	69.56%	64.42%	66.89%	64.75%
SOL	66.91%	61.94%	64.33%	62.25%

Table 3: Lexicon-based approach results

The differences in the results between the English and Spanish version of SFU Review Corpus are lower when a lexicon is used instead of a machine learning algorithm is applied. In a lexicon-based method is very important the recall value, because it indicates whether the set of words covers the vocabulary of the corpus. The recall value is upper 60% regarding English and Spanish, although is not an excellent value, is good for the two small and independent-domain lexicons. In the case of Spanish the supervised method is only 15.59% better regarding Accuracy. The results show that may be considered the use of a lexicon-based method for Spanish due to the few computer resources needed. Moreover, it must be highlighted the performance of SOL, so it is the first time that this resource is used to resolve a polarity classification problem.

The graph-based method has been described as a modular and flexible algorithm. Due to its modular nature we have carried out several experiments:

1. **wnet_ant+_eq1_[en|es]**: As baseline, we have run the algorithm with the same configuration as is described in (Montejo-Ráez et al., 2012), i.e. using the equation 1.
2. **wnet_ant-_eq1_[en|es]**: We have assessed the algorithm with a version of WordNet without the antonym relation.
3. **wnet_ant+_eq2_[en|es]**: The equation to calculate the polarity is 2
4. **wnet_ant-_eq2_[en|es]**: The same as wnet_ant+_eq2_[en|es] but the antonym relation is not considered.
5. **wnet_ant+_eq3_[en|es]**: The same as wnet_ant+_eq2_[en|es] but the equation 3 is used to calculate the polarity.
6. **wnet_ant-_eq3_[en|es]**: The same as wnet_ant+_eq3_[en|es] but the antonym relation is not considered.

Furthermore, one of the key elements of the algorithm is the possibility of setting the number of related synsets to get from WordNet. In all of the experiments we have evaluated from an expansion of 0 synsets to 100 synsets. In Table 4 are the best results obtained with the English and the Spanish version of SFU corpus.

Regarding the results, only for English is evident that the selection of the right equation to calculate the polarity score is important. On the other hand, the initial assumption that the relation of antonym could complicate the calculation of the final polarity, and the use of a graph of WordNet without antonym could enhance the results cannot be demonstrated because these experiments have reached the same results as the obtained ones using the graph with the relation of antonym. The equation 3, which includes additional information (in this case the BLEL lexicon) to calculate the final polarity score, outperforms the original way to get the polarity score (equation 1). The equation 3 for the English version of the corpus reaches 5.84% and 8.4% better results

than equation 1 regarding F1 and Accuracy respectively.

The results obtained with the Spanish reviews are a bit different. In this case, the results are always improved when the antonym relation is not taking into account. So the first conclusion is the relation of antonym is not convenient for the calculation of the polarity value on Spanish texts. The process of expansion with related senses has not been relevant for the final results on the English reviews, but when the language of the reviews is Spanish the expansion is more decisive. For the *wnet_ant-_eq3_es* experiment the best result has been reached considering 71 related senses, so we can conclude that for Spanish the context should be considered. Although the best results is obtained with the configuration *wnet_ant+_eq3_es*, it must be highlighted the precision value of 68.03% reached by the configuration *wnet_ant+_eq2_es*. In some OM experiments is more important the precision of the system than the recall or other evaluation measures, so for Spanish reviews should be taken account this configuration too.

Regarding English and Spanish results, Table 4 shows similar performance, i.e. the graph-based algorithm obtained better results when the antonym is not considered and the use of a lexicon of opinion words enhances considerably the results.

The supervised approach clearly outperforms the two unsupervised approaches. The results obtained by the two unsupervised approaches are closer. The lexicon based method has a better performance on English reviews regarding the four different evaluation measures considered. Thus, the lexicon method not only has better results but also it is simpler, faster and more efficient than the graph-based method. Nevertheless, the graph-based method on Spanish reviews outperforms in precision regarding the configuration *wnet_ant+_eq2_es* and in the other three measures take into account the configuration *wnet_ant+_eq3_es*. However, the graph-based method is only 1.64% better regarding the precision value, and 0.54% better regarding F1. Therefore, we also considered the lexicon-based approach as the more suitable approach than the graph-based one.

	Expansion	Precision	Recall	F1	Accuracy
wnet_ant+_eq1_en	2	66.86%	57.25%	61.68%	57.25%
wnet_ant-_eq1_en	2	66.86%	57.25%	61.68%	57.25%
wnet_ant+_eq2_en	0	65.27%	55.5%	59.99%	55.50%
wnet_ant-_eq2_en	0	65.27%	55.5%	59.99%	55.50%
wnet_ant+_eq3_en	3	68.83%	62.50%	65.51%	62.50%
wnet_ant-_eq3_en	3	68.83%	62.50%	65.51%	62.50%
wnet_ant+_eq1_es	0	65.42%	54.5%	59.46%	54.5%
wnet_ant-_eq1_es	19	64.39%	57.75%	60.89%	57.75%
wnet_ant+_eq2_es	0	68.03%	52.75%	59.42%	52.75%
wnet_ant-_eq2_es	70	64.62%	58.00%	61.13%	58.00%
wnet_ant+_eq3_es	71	65.91%	63.50%	64.68%	63.05%
wnet_ant-_eq3_es	71	65.91%	63.50%	64.68%	63.05%

Table 4: Results of the graph-based algorithm

5 Conclusion and future work

In this work, we have presented a set of experiments with a comparable corpora in English and Spanish. As it is usual, the supervised experiment has outperforms the unsupervised ones. The unsupervised experiments have included the evaluation of two different approaches: lexicon-based and graph-based. In the lexicon-based approach we have presented a new resource for the Spanish OM research community, being an important contribution of this paper. The results reached with SOL are very closed to the ones obtained with graph-based methods. Although, for short texts the graph-based method performed well, for the kind of reviews used in these experiments is not as good. Due to the fact that for English the BLEL lexicon has reached better results, for Spanish the results of SOL are nearly the same ones obtained by the graph method, and the use of a lexicon is more efficient, we conclude that the lexicon-based method is most suitable.

Currently we are improving the SOL lexicon, and also we are adding domain information to the words in SOL. Furthermore, one of our main objectives is the treatment of the negation because we considered that is essential for OM.

Acknowledgments

This work has been partially supported by a grant from the Fondo Europeo de Desarrollo Regional (FEDER), TEXT-COOL 2.0 project (TIN2009-13391-C04-02) and ATTOS project (TIN2012-38536-C03-0) from the Spanish Government. Also, this paper is partially funded by the European

Commission under the Seventh (FP7 - 2007-2013) Framework Programme for Research and Technological Development through the FIRST project (FP7-287607). This publication reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

References

- Eneko Agirre and Aitor Soroa. 2009. Personalizing pagerank for word sense disambiguation. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics, EACL '09*, pages 33–41, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner, and Daniel Tapias, editors, *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta, may. European Language Resources Association (ELRA).
- Carmen Banea, Rada Mihalcea, Janyce Wiebe, and Samer Hassan. 2008. Multilingual subjectivity analysis using machine translation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP '08*, pages 127–135, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Julian Brooke, Milan Tofiloski, and Maite Taboada. 2009. Cross-linguistic sentiment analysis: From english to spanish. In *Proceedings of the International Conference RANLP-2009*, pages 50–54, Borovets,

- Bulgaria, September. Association for Computational Linguistics.
- Chih-Chung Chang and Chih-Jen Lin. 2011. Libsvm: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.*, 2(3):27:1–27:27, May.
- Corinna Cortes and Vladimir Vapnik. 1995. Support-vector networks. *Machine Learning*, 20:273–297.
- Fermín L. Cruz, Jose A. Troyano, Fernando Enriquez, and Javier Ortega. 2008. Clasificación de documentos basada en la opinión: experimentos con un corpus de críticas de cine en español. *Procesamiento del Lenguaje Natural*, 41:73–80.
- Kushal Dave, Steve Lawrence, and David M. Pennock. 2003. Mining the peanut gallery: opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th international conference on World Wide Web, WWW '03*, pages 519–528, New York, NY, USA. ACM.
- Kerstin Denecke. 2008. Using sentiwordnet for multilingual sentiment analysis. In *Data Engineering Workshop, 2008. ICDEW 2008. IEEE 24th International Conference on*, pages 507–512. IEEE.
- Hatem Ghorbel and David Jacot. 2011. Sentiment analysis of french movie reviews. *Advances in Distributed Agent-Based Retrieval Tools*, pages 97–108.
- Aitor Gonzalez-Agirre, Egoitz Laparra, and German Rigau. 2012. Multilingual central repository version 3.0. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, may. European Language Resources Association (ELRA).
- Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '04*, pages 168–177, New York, NY, USA. ACM.
- Eugenio Martínez-Cámara, M. Teresa Martín-Valdivia, and L. Alfonso Ureña López. 2011. Opinion classification techniques applied to a spanish corpus. In *Proceedings of the 16th international conference on Natural language processing and information systems, NLDB'11*, pages 169–176, Berlin, Heidelberg. Springer-Verlag.
- M. Teresa Martín-Valdivia, Eugenio Martínez-Cámara, Jose M. Perea-Ortega, and L. Alfonso Ureña López. 2012. Sentiment polarity detection in spanish reviews combining supervised and unsupervised approaches. *Expert Systems with Applications*. In press.
- George A. Miller. 1995. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41.
- Arturo Montejo-Ráez, Eugenio Martínez-Cámara, M. Teresa Martín-Valdivia, and L. Alfonso Ureña López. 2012. Random walk weighting over sentiwordnet for sentiment polarity detection on twitter. In *Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis*, pages 3–10, Jeju, Korea, July. Association for Computational Linguistics.
- Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. The pagerank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab, November. Previous number = SIDL-WP-1999-0120.
- Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Found. Trends Inf. Retr.*, 2(1-2):1–135, January.
- Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up?: Sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing - Volume 10, EMNLP '02*, pages 79–86, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Mohammed Rushdi-Saleh, M. Teresa Martín-Valdivia, L. Alfonso Ureña López, and José M. Perea-Ortega. 2011a. OCA: Opinion corpus for Arabic. *Journal of the American Society for Information Science and Technology*, 62(10):2045–2054, October.
- Mohammed Rushdi-Saleh, Maria Teresa Martn-Valdivia, Luis Alfonso Urea-Lpez, and Jos M. Perea-Ortega. 2011b. Bilingual Experiments with an Arabic-English Corpus for Opinion Mining. In Galia Angelova, Kalina Bontcheva, Ruslan Mitkov, and Nicolas Nicolov, editors, *RANLP*, pages 740–745. RANLP 2011 Organising Committee.
- Maite Taboada and Jack Grieve. 2004. Analyzing appraisal automatically. In *Proceedings of AAAI Spring Symposium on Exploring Attitude and Affect in Text (AAAI Technical Report SS# 04# 07)*, Stanford University, CA, pp. 158q161. AAAI Press.
- Maite Taboada. 2008. Sfu review corpus. http://www.sfu.ca/~mtaboada/research/SFU_Review_Corpus.html.
- Peter D. Turney. 2002. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, ACL '02*, pages 417–424, Stroudsburg, PA, USA. Association for Computational Linguistics.