Information Processing and Management 51 (2015) 520–531

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/infoproman

# A Spanish semantic orientation approach to domain adaptation for polarity classification

M. Dolores Molina-González, Eugenio Martínez-Cámara *, M. Teresa Martín-Valdivia, L. Alfonso Ureña-López

*SINAI Research Group, University of Jaén, Campus Las Lagunillas, E-23071 Jaén, Spain*

ARTICLE INFO

*Article history:*
Received 14 August 2013
Received in revised form 3 October 2014
Accepted 5 October 2014
Available online 8 November 2014

*Keywords:*
Spanish opinion mining
Sentiment lexicon
Domain adaptation

ABSTRACT

One of the problems of opinion mining is the domain adaptation of the sentiment classifiers. There are several approaches to tackling this problem. One of these is the integration of a list of opinion bearing words for the specific domain. This paper presents the generation of several resources for domain adaptation to polarity detection. On the other hand, the lack of resources in languages different from English has orientated our work towards developing sentiment lexicons for polarity classifiers in Spanish. The results show the validity of the new sentiment lexicons, which can be used as part of a polarity classifier.



## 1. Introduction

Opinion Mining (OM) is defined as the computational treatment of opinion, sentiment, and subjectivity in text. This new area of research is becoming more and more important mainly due to the growth of social media where users con- tinually post contents on the web in the form of comments, opinions, emotions, etc. The OM discipline combines Natural Language Processing (NLP) with data mining techniques and includes a large number of tasks (Pang & Lee, 2008). One of the most widely studied tasks is the polarity classification of reviews. This task focuses on determining the overall sentiment-orientation (positive or negative) of the opinions contained within a given document.

Although different approaches have been applied to polarity classification, the mainstream basically consists of two major methodologies. 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, Lee, & Vaithyanathan, 2002). On the other hand, the approach based on Semantic Orientation (SO) does not need prior training, but takes into account the positive or negative orientation of words (Turney et al., 2002). This method, also known as the unsupervised approach, makes use of lexical resources like lists of opinion words, lexicons, dictionaries, etc. 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. On the contrary, the SO strategy 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 orientation of opinion. In this paper we focus on the generation of linguistic resources to tackle the problem of polarity classification using an unsupervised approach.

* Corresponding author.
*E-mail addresses:* mdmolina@ujaen.es (M. Dolores Molina-González), emcamara@ujaen.es (E. Martínez-Cámara), maite@ujaen.es (M. Teresa Martín-Valdivia), laurena@ujaen.es (L. Alfonso Ureña-López).

http://dx.doi.org/10.1016/j.ipm.2014.10.002
0306-4573/

While opinions and comments on the Internet are expressed in any language, most research in OM is focused on English texts. However, languages such as Chinese, Spanish or Arabic, are even more present on the web. Thus it is important to develop resources to help researchers to work with these languages. The work presented herein is mainly motivated by the need to develop polarity classification systems and resources in languages other than English. Specifically, in this paper we deal with Spanish reviews. We present an experimental study over the SFU Review Corpus[1] (Brooke, Tofiloski, & Taboada, 2009), which is a comparable corpus that includes opinions of several topics in English and in Spanish in different domains.

One of the open problems in OM is that of domain adaptation. Although movie reviews have been the most studied domain in sentiment analysis, a wide range of areas are being investigated such as political debates, hotels or music. However, when we train a classifier using a specific domain we need to adapt it in order to apply it to another domain. For example, the sentence "Definitively, you should read the book" most likely refers to positive polarity for Book reviews but negative sentiment for Movie reviews.

Thus the problem of domain adaptation is attracting more and more attention in OM. In this paper we carry out an experimental study of domain adaptation of linguistic resources for Spanish reviews in different domains. We have used the Spanish version of SFU, which includes 400 reviews for 8 different domains. We have generated several lists of opinionated words integrating knowledge from the different domains and we have compared the results obtained. A corpus-based approach is followed with the aim of adapting a general-purpose sentiment lexicon to a specific domain by integrating lists of opinion bearing words. iSOL[2] (Molina-González, Martínez-Cámara, Martín-Valdivia, & Perea-Ortega, 2013) is the general-purpose sentiment lexicon chosen. The Spanish version of the SFU corpus was the corpus selected for the adaptation process due mainly to the fact that it covers 8 different domains. Following different heuristics, which will be described later, the most frequent opinion bearing words are appended to iSOL. Several experiments were carried out with the goal of assessing the new domain-specific sentiment lexicons. The analysis of the results shows the validity of the new lists.

The paper is organised as follows: Section 2 briefly describes other papers that study non-English sentiment polarity classification and, specifically work related to Spanish OM. In addition, we include some papers studying the domain adaptation problem. In Section 3 we explain the different resources used. Sections 4 and 5 present the experiments carried out and discusses the main results obtained. Finally, we outline conclusions and further work.

## 2. Related work

In this study we focus on two open problems in opinion mining: non-English polarity classification and the domain adaptation problem. Next, we will comment on some papers that have inspired our work.

### 2.1. Non-English polarity classification

There are some interesting papers that have studied the problem using non-English collections. For example, Denecke (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: LingPipe3, SentiWordNet (Esuli & Sebastiani, 2006) with classification rule, and SentiWordNet with machine learning. In (Zhang, Zeng, Li, Wang, & Zuo, 2009) Chinese sentiment analysis is applied on two datasets. In the first one euthanasia reviews were collected from different web sites, while the second dataset was about six product categories collected from Amazon (Chinese reviews). Ghorbel and Jacot (2011) used a corpus with movie reviews in French. They applied a supervised classification combined with SentiWordNet in order to determinate the polarity of the reviews. In (Agić, Ljubešić, & Tadić, 2010) a manually annotated corpus is presented with news on the financial market in Croatia. In (Rushdi-Saleh, Martín-Valdivia, Ureña López, & Perea-Ortega, 2011) a corpus of movies reviews in Arabic annotated with polarity was presented and several experiments using machine learning techniques were performed.

Regarding Spanish, there are also some interesting studies. For example, Banea, Mihalcea, Wiebe, and Hassan (2008) proposed several approaches to cross lingual subjectivity analysis by directly applying the translations of opinion corpus in English to training an opinion classifier in Romanian and Spanish. This study 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 dealing with Spanish and English resources are presented. They conclude 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. They proposed three approaches: the first one uses Spanish resources generated manually and automatically. The second one applies ML to a Spanish corpus. The last one translates the Spanish corpus into English and then applies the SO-CAL (Semantic Orientation CALculator), a tool developed by themselves (Taboada, Brooke, Tofiloski, Voll, & Stede, 2011). Cruz, Troyano, Enriquez, and Ortega (2008) manually recollected the MuchoCine (MC) corpus in order to develop a sentiment polarity classifier based on semantic orientation. The corpus contains annotated Spanish movie reviews from the MuchoCine website.[3] The MC corpus was also used in

[1] urlhttp://www.sfu.ca/œmtaboada/research/SFU_Review_Corpus.html.
[2] The iSOL resource is freely available for research purpose at urlhttp://sinai.ujaen.es/?p=1202.
[3] urlhttp://www.muchocine.net/.

(Martínez-Cámara, Martín-Valdivia, & Ureña-López, 2011) to carry out several experiments with a supervised approach applying different ML algorithms (SVM, NB,BBR, KNN, C4.5). The results are much better than those obtained with the unsupervised approach proposed by Cruz et al. (2008).

One of the barriers in the study of how to resolve the problem of polarity classification on Spanish reviews is the lack of Spanish linguistic resources for opinion mining. However, some new sentiment linguistic resources, mainly lists of opinion bearing words, have been made available in the last years. Sidorov et al. (2013) provided the Spanish Emotion Lexicon (SEL). SEL is composed of 2,036 words that are associated with the measure of Probability Factor of Affective use (PFA) with respect to at least one basic emotion or category: joy, anger, fear, sadness, surprise, and disgust. Molina-González et al. (2013) describe a new Spanish sentiment lexicon. The authors translated the Bing Liu English Opinion Lexicon (Hu & Liu, 2004) into Spanish. Subsequently, the translated version was manually corrected and improved with Spanish opinion bearing words. The result is the lexicon iSOL, which is composed of 8135 words. iSOL has been also used in (Martínez-Cámara, Martín-Valdivia, & Molina-González, 2013) with promising results.

### 2.2. Domain adaptation for sentiment analysis

Different methods have been proposed for tackling the domain adaptation problem. One of the primary studies in sentiment analysis is (Blitzer, Dredze, & Pereira, 2007). They use Structural Correspondence Learning (SCL) to find correspondences between features from source and target domains through modelling their correlations with pivot features. The proposed approach was successfully tested on review data from 4 domains (DVDs, books, kitchen appliances and electronics). Following the same idea, Pan, Ni, Sun, Yang, and Chen (2010) present the Spectral Feature Alignment (SFA) that uses spectral clustering to align domain-specific and domain- independent words into a set of feature-clusters. The results obtained surpass the SCL. Jiang et al. (2007) describe two distinct needs. On the one hand, instance adaptation takes into account the change of instance probability, e.g., the change of vocabulary or the change of words frequency from one domain to another; On the other hand, labelling adaptation models the changes of the labelling function, since one feature that is positive in the source domain may express the opposite meaning in the target domain. Most studies tackle the instance adaptation problem, while Xia, Zong, Hu, and Cambria (2013) propose a combination taking into account both kinds of adaptation, obtaining good results. In Ponomareva et al. (2012) graph-based approaches are applied. They model the data as a graph of documents, taking into account not only the document content but also document connectivity, which is modelled as document sentiment similarity rather than content similarity.

## 3. Domain adaptation method

Like some of the studies mentioned in the previous section, herein we propose a domain adaptation method for sentiment analysis. However, we focus our study on reviews written in Spanish. In addition, in contrast to the aforementioned methods which mainly focus on machine learning algorithms, we propose a lexicon-based approach to the domain adaptation problem. We follow a very simple strategy by generating lists of opinionated words for each domain in an automatic way. The Spanish version of Taboada corpus SFU is used in our experiments. Firstly we apply a general lexicon to the corpus, taking into account the different domains. Then, four different opinionated word lists are generated for each of the eight different domains and four different word lists for all domains of the corpus. Following a corpus based method, two heuristics are assessed with the aim of integrating into each list the most frequent words used for positive and negative reviews. A subset of the corpus is used to build the lists and the other part to test the new resources. The results obtained show an improvement over the experiments using the general lexicon.

### 3.1. Corpus

In order to carry out the experiment we chose the Spanish part of the comparable SFU Review Corpus. The SFU Review Corpus is composed of reviews of products in English and Spanish. The English version (Taboada & Grieve, 2004) has 400 reviews (200 positive and 200 negative) of commercial products downloaded in 2004 from the Epinions[4] web which are divided into eight categories. Each category includes 25 positive reviews and 25 negative reviews. Subsequently, the authors of the SFU Review Corpus have made available the Spanish version of the corpus[5] with the aim of offering a comparable corpus for the research community. The Spanish reviews are divided into eight similar categories: books, cars, computers, washing machines, hotels, movies, music and phones. Each category also has 25 positive and 25 negative reviews, which are defined as positive or negative based on the number of stars given by the reviewer (1–2 = negative; 4–5 = positive; 3-star reviews are not included). In this case, the reviews are downloaded from the Ciao.es[6] website.

[4] urlhttp://www.epinions.com.
[5] urlhttp://www.sfu.ca/mtaboada/download/downloadCorpusSpa.html.
[6] urlhttp://www.ciao.es/.

### 3.2. Opinion lists generation

We followed a lexicon-based approach to tackle the problem. The iSOL lexicon (Molina-González et al., 2013) was selected to carry out the experiments. This resource was generated from the BLEL lexicon (Hu & Liu, 2004) by automatically translating it into Spanish and obtaining the SOL (Spanish Opinion Lexicon) resource. Then, this resource was manually reviewed in order to improve the final list of words obtaining iSOL (improved SOL). This resource has been successfully evaluated in (Molina-González et al., 2013) using a Spanish corpus of movie reviews called MuchoCine (Cruz et al., 2008). The results showed that the use of an improved list of sentiment words from the same language could be considered as a good strategy for unsupervised polarity classification. Moreover, another list was generated appending the positive and negative words of the MuchoCine corpus. In this way, domain knowledge was added in the lexicon. The result of the process was a new lexicon which is called eSOL (enriched SOL). The experiments with eSOL showed the advantages of using domain knowledge. Thus, the main motivation of this paper is the integration of knowledge from the domain in order to improve the final polarity classification system.

The improved Spanish Opinion Lexicon (iSOL) is composed of 2509 positive and 5626 negative words, thus in total the Spanish lexicon has 8135 opinion words.

As is well-known in the SA research community, the semantic orientation of a word is domain-dependent. Within the approaches followed by research into the compilation of a set of polar words, the most suitable for obtaining domain-dependent opinion words is that known as the corpus-based approach. Hatzivassiloglou et al. (1997) take some adjectives as seeds in order to find additional sentiment adjectives in the corpus. Their method takes advantage of a set of conventions on connectives with the aim of identifying more polar words and their orientation from a sentiment label corpus. On the other hand, Du, Tan, Cheng, and Yun (2010) follow a similar assumption, and they consider that a word should be positive (or negative) if it appears in many positive (or negative) documents.

We follow a more straightforward method which consists of enlarging iSOL with the most frequent words of a sample of the SFU Spanish Review Corpus. The key point of the method is to automatically find domain sentiment words in the different domains of the corpus with the goal of developing a domain specific sentiment lexicon for each domain covered by the SFU Spanish Review Corpus. Four different word lists for each domain of the corpus and four different word lists for the categories (positive and negative) of the corpus are generated. Then, the new resources are assessed over the reviews of the corpus which are not utilised for building the lists. To build the first four lists, we split the 50 reviews for each category into two random groups of 15 and 10 positive reviews and 15 and 10 negative reviews. We used the group of 15 reviews of both polarities (30 reviews in total) to seek the words and integrate them into the new resources. Then we used the group of 20 reviews (10 positive, 10 negative) to test the validity of the new domain specific lexicons.

Taking the general-purpose sentiment lexicon iSOL, we generated our first list of opinion words for each domain of the SFU Spanish Review Corpus. After removing the stop words from the documents, the selection of the polarity domain-dependent words consists of calculating the absolute frequency of each word per class (positive/negative), and then the most-used positive and negative words were appended to iSOL. The new list with domain information is called eSOLdomainLocal (enriched SOL Local) where domain = cars, hotels, washing machine, books, phones, music, computers, movies.

The second way to enrich the iSOL lexicon consists of adding not only the sentiment words but also the most frequently used domain-dependent words. Commercial names and proper nouns were discarded from this selection. Some examples of these discarded words were: Fagor, BMW, Almudena Grandes, Quijote, Siemens, Citroen, Acer, Nokia, Bon Jovi, AC/DC, Bosh, Hannibal Lecter, George Lucas, etc. The most frequent domain-dependent words are selected using the following formula:

$$\text{list}(\text{word}) = \begin{cases} positive & \text{if } (f^- = 0 \wedge f^+ \geqslant 3) \vee \left(\frac{f^+}{f^-} \geqslant 3\right) \\ negative & \text{if } (f^+ = 0 \wedge f^- \geqslant 3) \vee \left(\frac{f^-}{f^+} \geqslant 3\right) \end{cases} \quad (1)$$

where $f^+$ is the frequency of the word in positive reviews and $f^-$ in negative reviews. Thus, those words that satisfy Eq. (1) are appended to the positive or negative list of eSOLdomainLocal. These new resources are called eSOLdomainLocal*. Tables 1 and 2 show some examples of domain-dependent words that have been appended to the lists.

The third and fourth lists generated are similar to the first and second ones. The difference is how to find the most used sentiment and domain-dependent words. In these lists, if one word is used one or more times in one positive or negative review we considered that its frequency is one. That is, although the word appears several times in a specific review, its frequency is one. Therefore in these lists, the highest possible frequency of a word is 15. The new resources are called eSOLdomainGlobal with only sentiment words and eSOLdomainGlobal* including the most frequently used sentiment words and domain-dependent words.

In order to generate the last four lists, we considered all the domains together. Again, we split the corpus into two groups, one for integrating opinion words into the lists and another one for testing the new resources. Thus, we used 120 positive reviews (the same 15 positive reviews per domain used before multiplied by 8 domains) and 120 negative reviews to generate new resources from eSOL, and we carried out the experiment with the rest of the corpus, that is 160 reviews, 80 positive (10 positive reviews for each 8 different domains) and 80 negative.

**Table 1**
Some positive words included in eSOLdoaminLocal*.

| Word | Domain | Freq. in positive reviews | Freq. in negative reviews |
|---|---|---|---|
| *Consumo* (consumption) | Cars | 10 | 1 |
| *Maletero* (boot) | Cars | 6 | 0 |
| *Menú* (menu) | Hotels | 4 | 0 |
| *Minibar* (minibar) | Hotels | 5 | 0 |
| *Temperatura* (temperature) | Washing machine | 12 | 1 |
| *Capacidad* (capacity) | Washing machine | 7 | 2 |
| *Recuerdos* (memories) | Books | 10 | 1 |
| *Introducción* (introduction) | Books | 5 | 1 |
| *Conectividad* (connectivity) | Phones | 6 | 1 |
| *Navegación* (navigation) | Phones | 6 | 0 |
| *Ritmos* (rhythms) | Music | 8 | 1 |
| *Sonidos* (sounds) | Music | 8 | 1 |
| *Rendimiento* (performance) | Computer | 13 | 1 |
| *Plataforma* (platform) | Computer | 11 | 0 |
| *Escena* (scene) | Movies | 19 | 2 |
| *Estreno* (premiere) | Movies | 5 | 0 |

**Table 2**
Some negative words included in eSOLdoaminLocal*.

| Word | Domain | Freq. in positive reviews | Freq. in negative reviews |
|---|---|---|---|
| *Taller* (workshop) | Cars | 2 | 19 |
| *Sensor* (sensor) | Cars | 0 | 5 |
| *Manchas* (spots) | Hotels | 0 | 4 |
| *Moqueta* (fitted carpet) | Hotels | 1 | 6 |
| *Acero* (steel) | Washing machine | 0 | 3 |
| *Cocina* (kitchen) | Washing machine | 1 | 8 |
| *Serie* (series) | Books | 2 | 9 |
| *Ritmo* (rhythm) | Books | 0 | 5 |
| *Covertura* (coverage) | Phones | 0 | 8 |
| *Carga* (charge) | Phones | 0 | 5 |
| *Remix* (remix) | Music | 1 | 6 |
| *Versiones* (versions) | Music | 0 | 4 |
| *Pantalla* (screen) | Computer | 0 | 4 |
| *Computadora* (computer) | Computer | 0 | 8 |
| *Trailer* (trailer) | Movies | 1 | 6 |
| *Saga* (saga) | Movies | 0 | 9 |

On the one hand, we generated the new eSOLLocal resource taking into account only the most frequent sentiment words. Then, we generated the new eSOLLocal* taking into account the sentiment words and also the most frequent domain words, which were obtained following the Eq. (1).

On the other hand, in the compilation of the latter two lists the difference is how to find the most used sentiment and domain-dependent words. If one word is used one or more times in one positive or negative review we have considered that its frequency is one. Therefore, in these lists the highest frequency is 120, and this only happens if the word is in all the reviews. The new resource, with only sentiment words added to iSOL, is called eSOLGlobal, and the resource with not only the sentiment words but also including the most frequent domain words is called eSOLGlobal*.

Regarding the original lexicon iSOL, we increased the size of the generated eSOLdomainLocal and eSOLdomainLocal* lists for both negative and positive lists of words. Tables 3 and 4 show the number of words added to iSOL in each resource respectively.

**Table 3**
Number of words included in the new eSOLdomainLOCAL lexicon and final size of the lists.

| eSOLdomainLocal | #positive words | #negative words |
|---|---|---|
| Cars | 18 (2527) | 23 (5649) |
| Hotels | 9 (2518) | 10 (5636) |
| Washing machines | 11 (2520) | 13 (5639) |
| Books | 19 (2528) | 26 (5652) |
| Cell phones | 20 (2529) | 33 (5659) |
| Music | 27 (2536) | 19 (5645) |
| Computers | 17 (2526) | 19 (5645) |
| Movies | 32 (2541) | 22 (5648) |

**Table 4**
Number of words included in the new eSOLdomainLocal* lexicon and final size of the lists.

| eSOLdomainLocal* | #positive words | #negative words |
|---|---|---|
| Cars | 28 (2537) | 36 (5662) |
| Hotels | 24 (2533) | 15 (5641) |
| Washing machines | 18 (2527) | 22 (5648) |
| Books | 29 (2538) | 36 (5662) |
| Cell phones | 42 (2551) | 36 (5662) |
| Music | 43 (2552) | 26 (5652) |
| Computers | 51 (2560) | 25 (5651) |
| Movies | 58 (2567) | 29 (5655) |

Concerning the eSOLdomainGlobal and eSOLdomainGlobal*, we also increased the size of the original iSOL lexicon. Tables 5 and 6 show the number of words added to iSOL and also the final size of each new list.

Regarding the eSOLLocal, eSOLLocal*, eSOLGlobal and eSOLGlobal*, the size also increased compared to the original iSOL lexicon, and the number of positive and negative words integrated in the new lists is shown in Table 7.

## 4. Experiments and results

In order to evaluate the different approaches, we used the traditional measures employed in text classification: precision (P), recall (R), F1 and Accuracy:

$$P = \frac{TP}{TP + FP} \tag{2}$$

$$R = \frac{TP}{TP + FN} \tag{3}$$

$$F1 = \frac{2PR}{P + R} \tag{4}$$

$$Acc. = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

where TP (True Positives) and TN (True Negatives) are those assessments where the system and a human expert agree on a label (in this case, TP and TN are those positive or negative reviews rightly classified), FP (False Positives) and FN (False Negatives) are those labels assigned by the system that do not agree with the expert assignment, in plain English, the positive and negatives reviews misclassified. F1 is a measure that combines both precision and recall, calculating the proportion of true results (both true positives and true negatives) Sebastiani (2002). Due to the fact that the system classifies two classes, the P, R and F1 of each class have been calculated. Then, the overall P, R and F1 of the system have been obtained following the macro-averaged evaluation measures. The macro-avergaed evaluation measures formulae for P, R and F1 are the following:

$$\text{Macro-F1} = \frac{2 * \text{Macro-Precision} * \text{Macro-Recall}}{\text{Macro-Precision} + \text{Macro-Recall}} \tag{6}$$

Where Macro-Recall and Macro-Precision are obtained as follows:

$$\text{Macro-Recall} = \frac{\sum_{i=1}^{c} R_i}{c} \tag{7}$$

$$\text{Macro-Precision} = \frac{\sum_{i=1}^{c} P_i}{c} \tag{8}$$

where $c$ is the number of classes ($c = 2$).

**Table 5**
Number of words included in the new eSOL*domain*Global lexicon and final size of the lists.

| eSOLdomainGlobal | #positive words | #negative words |
|---|---|---|
| Cars | 19 (2528) | 22 (5648) |
| Hotels | 9 (2518) | 10 (5636) |
| Washing machines | 11 (2520) | 13 (5639) |
| Books | 20 (2529) | 25 (5651) |
| Cell phones | 20 (2529) | 31 (5657) |
| Music | 29 (2538) | 19 (5645) |
| Computers | 18 (2527) | 19 (5645) |
| Movies | 26 (2535) | 22 (5648) |

**Table 6**
Number of words included in the new eSOL*domain*Global* lexicon and final size of the lists.

| eSOLdomainGlobal* | #positive words | #negative words |
|---|---|---|
| Cars | 28 (2537) | 34 (5660) |
| Hotels | 21 (2530) | 15 (5641) |
| Washing machines | 18 (2527) | 17 (5643) |
| Books | 27 (2536) | 29 (5655) |
| Cell phones | 35 (2544) | 33 (5659) |
| Music | 37 (2548) | 21 (5647) |
| Computers | 30 (2539) | 21 (5647) |
| Movies | 39 (2548) | 25 (5651) |

**Table 7**
Number of words included in the new resources considering all the domains and final size of the lists.

| All domains | #positive words | #negative words |
|---|---|---|
| eSOLLocal | 113 (2622) | 399 (6025) |
| eSOLLocal* | 118 (2627) | 150 (5776) |
| eSOLGlobal | 113 (2622) | 278 (5904) |
| eSOLGlobal* | 118 (2627) | 141 (5767) |

Several experiments were carried out in order to verify the utility of the new resources generated for Spanish from iSOL: eSOLdomainLocal(*), eSOLdomainGlobal(*), eSOLLocal(*) and eSOLGlobal(*) where domain = cars, hotels, washing machine, books, phones, music, computers, movies. The general method consists of finding the presence in the reviews of opinion words which are included in a lexicon of opinion words. If a review has more positive words than negative ones, the document polarity is positive, otherwise negative (Eq. (9)).

$$p(r) = \begin{cases} 1 & \text{if } |positive| > |negative| \\ -1 & \text{if } |positive| \leqslant |negative| \end{cases} \tag{9}$$

where $p(r)$ is the polarity of the review, |positive| is the number of positive words and |negative| is the number of negative words.

Before carrying out the experiments we performed a pre-processing step on the SFC corpus in order to apply the same criteria followed in the generation of the enriched iSOL lists. For example, we changed capital letters to non-capital ones, accented letters to non-accented ones and special characters were separated from words in the reviews. Moreover, the stop words were discarded.

For the baseline experiment we took the same 20 reviews used to test each domain separately and applied the iSOL lexicon. The results are presented in Table 8.

The next experiments were carried out over the 160 reviews used for testing purposes (10 positive reviews and 10 negative reviews per domain chosen randomly and not used to generate the lexicons). Thus the results obtained with the eSOLdomainLocal are shown in Table 9. This resource was built by adding the most used sentiment words in positive or negative reviews of each domain to iSOL. Table 8 also includes the percentage of improvement over the baseline experiment (Table 8) using the following equation:

$$\text{Improvement} = \frac{\text{Macro-F1eSOLdomain[Local|Global]} - \text{MacroF1iSOLdomain}}{\text{MacroF1iSOLdomain}} * 100 \tag{10}$$

The second eSOLdomainLocal* resource enriched the iSOL lexicon with sentiment words, but also the most frequent domain-dependent words (Eq. (1)). Table 10 shows the results obtained with this lexicon.

**Table 8**
Polarity classification over the SFU corpus using iSOL.

| | Macro-P | Macro-R | Macro-F1 | Accuracy |
|---|---|---|---|---|
| Cars | 0.8125 | 0.7000 | 0.7521 | 0.7000 |
| Hotels | 0.8571 | 0.8000 | 0.8276 | 0.8000 |
| Washing machines | 0.5667 | 0.5500 | 0.5582 | 0.5500 |
| Books | 0.7083 | 0.7000 | 0.7041 | 0.7000 |
| Cell phones | 0.7778 | 0.6000 | 0.6774 | 0.6000 |
| Music | 0.4333 | 0.4500 | 0.4415 | 0.4500 |
| Computers | 0.5667 | 0.5500 | 0.5582 | 0.5500 |
| Movies | 0.5549 | 0.5500 | 0.5525 | 0.5500 |

**Table 9**
Polarity classification over the SFU corpus using eSOLdomainLocal.

| | Macro-P | Macro-R | Macro-F1 | Accuracy | Improvement (%) |
|---|---|---|---|---|---|
| Cars | 0.8571 | 0.8000 | 0.8276 | 0.8000 | 10.04 |
| Hotels | 0.8125 | 0.8000 | 0.8062 | 0.8000 | −2.58 |
| Washing machines | 0.8000 | 0.8000 | 0.8000 | 0.8000 | 43.31 |
| Books | 0.7083 | 0.7000 | 0.7041 | 0.7000 | 0.00 |
| Cell phones | 0.7941 | 0.6500 | 0.7149 | 0.6500 | 5.53 |
| Music | 0.5000 | 0.5000 | 0.5000 | 0.5000 | 13.24 |
| Computers | 0.5000 | 0.5000 | 0.5000 | 0.5000 | −14.42 |
| Movies | 0.5000 | 0.5000 | 0.5000 | 0.5000 | −9.49 |

**Table 10**
Polarity classification over the SFU corpus using eSOLdomainLocal*.

| | Macro-P | Macro-R | Macro-F1 | Accuracy | Improvement (%) |
|---|---|---|---|---|---|
| Cars | 0.8571 | 0.8000 | 0.8276 | 0.8000 | 10.04 |
| Hotels | 0.8571 | 0.8000 | 0.8276 | 0.8000 | 0.00 |
| Washing machines | 0.8846 | 0.8500 | 0.8670 | 0.8500 | 55.31 |
| Books | 0.7083 | 0.7000 | 0.7041 | 0.7000 | 0.00 |
| Cell phones | 0.7778 | 0.6000 | 0.6774 | 0.6000 | 0.00 |
| Music | 0.5980 | 0.5500 | 0.5730 | 0.5500 | 29.78 |
| Computers | 0.2368 | 0.4500 | 0.3103 | 0.4500 | −44.41 |
| Movies | 0.5000 | 0.5000 | 0.5000 | 0.5000 | −9.49 |

The next two experiments are similar to the two previous ones but using eSOLdomainGlobal and eSOLdomainGlobal*. The difference between Local and Global is how to find the most used sentiment and domain-dependent words. If one word is used one or more times in a positive or negative review we considered that its frequency is one. Therefore, in this experiment the highest possible frequency of a word is 15. Tables 11 and 12 show the results obtained using the eSOLdomainGlobal and eSOLdomainGlobal* resources respectively.

Taking as an example the cars domain to simplify, Table 13 shows how many new words were found in the reviews when we use the new lists generated. As we can see, in reviews "coches_no_2_12" and "coches_no_2_20" whose rank is −1, with iSOL the review is classified as Positive (FP), and with the new lists as Negative, which means an improvement of the system.

For the last experiments we did not take into account the different domains individually, and so we did not separate the domains of the SFU Reviews Corpus. Therefore, we have two new groups of reviews, one for generating the lists and another

**Table 11**
Polarity classification over the SFU corpus using eSOLdomainGlobal.

| | Macro-P | Macro-R | Macro-F1 | Accuracy | Improvement (%) |
|---|---|---|---|---|---|
| Cars | 0.8571 | 0.8000 | 0.8276 | 0.8000 | 10.04 |
| Hotels | 0.8125 | 0.8000 | 0.8062 | 0.8000 | −2.58 |
| Washing machines | 0.8000 | 0.8000 | 0.8000 | 0.8000 | 43.31 |
| Books | 0.7083 | 0.7000 | 0.7041 | 0.7000 | 0.00 |
| Cell phones | 0.7941 | 0.6500 | 0.7149 | 0.6500 | 5.53 |
| Music | 0.5000 | 0.5000 | 0.5000 | 0.5000 | 13.24 |
| Computers | 0.5000 | 0.5000 | 0.5000 | 0.5000 | −10.42 |
| Movies | 0.5549 | 0.5500 | 0.5525 | 0.5500 | 0.00 |

**Table 12**
Polarity classification over the SFU corpus using eSOLdomainGlobal*.

| | Macro-P | Macro-R | Macro-F1 | Accuracy | Improvement (%) |
|---|---|---|---|---|---|
| Cars | 0.8571 | 0.8000 | 0.8276 | 0.8000 | 10.04 |
| Hotels | 0.8333 | 0.7500 | 0.7895 | 0.7500 | −4.6 |
| Washing machines | 0.8846 | 0.8500 | 0.8670 | 0.8500 | 55.31 |
| Books | 0.7083 | 0.7000 | 0.7041 | 0.7000 | 0.00 |
| Cell phones | 0.7778 | 0.6000 | 0.6774 | 0.6000 | 0.00 |
| Music | 0.5980 | 0.5500 | 0.5730 | 0.5500 | 29.78 |
| Computers | 0.5980 | 0.5500 | 0.5730 | 0.5500 | 2.64 |
| Movies | 0.5000 | 0.5000 | 0.5000 | 0.5000 | −9.49 |

one for testing the new resources. The group for generating new resources eSOLLocal(*) and eSOLGlobal(*) includes a total of 240 documents: 120 positive reviews (15 positive reviews per domain multiplied by 8 domains) and 120 negative reviews. The group of reviews used for evaluating the generated lists is composed of 160 documents (10 positive and 10 negative reviews for each of the 8 domains). Thus, for the eSOLLocal we added only the most frequent sentiment words to iSOL lists and for the eSOLLocal* resource we also included the most frequent domain words if their frequency in positive (or negative) reviews was three or more times as much as negative (or positive) reviews. A similar process was followed to generate the eSOLGlobal and eSOLGlobal* resources. The difference is how to find the most used sentiment and domain words. If one word is used one or more times in one positive or negative review we considered that its frequency is one. Therefore, in this experiment the highest possible frequency of a word is 120. Table 14 shows the results obtained with these new resources, including the baseline experiment using the original iSOL list.

## 5. Analysis of results

In Table 14 we can see that the results obtained are improved for all the cases when we integrate domain information without taking into account the domains individually but considering all them together.

As we can see by comparing the different tables of results, the baseline experiment is improved upon for five domains (Cars, Washing machines, Music and Mobile phones) when we apply domain sentiment lexicons. Nevertheless, the results obtained with the new resources over the domains Hotels, Books, Computers and Movies are worse or equal than the original iSOL list.

On the other hand, taking the means of the Macro-F1 and accuracy of results, we can see that the eSOLdomainGlobal and eSOLdomainGlobal* lists obtained results a little better than eSOLdomainLocal and eSOLdomainLocal* lexicons, respectively. This means that in order to generate domain adapted opinion bearing word lists it is advisable to measure the frequency of the words as the number of documents of the corpus where the word appears. So, if a word is in most of the documents of the corpus, it is more representative than the word which is repeated a lot of times in a single document but does not appear in the others. In our case, if a word is in most of the positive documents it is very likely that the word expresses a positive opinion or sentiment, but if that word is only repeated several times in a positive document it does not mean that it expresses a positive meaning.

**Table 13**
Number of words of the different lists in the reviews.

| Id. Review | Rank | iSOL | | eSOLcarsLocal | | eSOLcarsLocal* | | eSOLcarsGlobal | | eSOLcarsGlobal* | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Pos. | Neg | Pos. | Neg | Pos. | Neg | Pos. | Neg | Pos. | Neg |
| coches_yes_5_10 | 1 | 4 | 0 | 6 | 0 | 8 | 0 | 6 | 0 | 8 | 0 |
| coches_yes_5_12 | 1 | 28 | 14 | 31 | 17 | 38 | 18 | 31 | 16 | 38 | 18 |
| coches_yes_5_15 | 1 | 17 | 3 | 17 | 4 | 24 | 4 | 17 | 4 | 24 | 4 |
| coches_yes_5_17 | 1 | 7 | 3 | 10 | 3 | 10 | 3 | 10 | 3 | 10 | 3 |
| coches_yes_5_21 | 1 | 16 | 12 | 20 | 12 | 24 | 15 | 22 | 12 | 24 | 15 |
| coches_yes_5_25 | 1 | 8 | 7 | 8 | 7 | 10 | 7 | 8 | 7 | 10 | 7 |
| coches_yes_5_4 | 1 | 16 | 3 | 16 | 3 | 22 | 3 | 16 | 3 | 22 | 3 |
| coches_yes_5_5 | 1 | 3 | 2 | 8 | 3 | 9 | 5 | 8 | 3 | 9 | 5 |
| coches_yes_5_7 | 1 | 10 | 5 | 10 | 5 | 15 | 5 | 10 | 5 | 15 | 5 |
| coches_yes_5_8 | 1 | 20 | 3 | 22 | 4 | 29 | 6 | 23 | 4 | 29 | 6 |
| coches_no_2_10 | −1 | 8 | 16 | 9 | 17 | 15 | 17 | 12 | 17 | 15 | 17 |
| coches_no_2_12 | −1 | 3 | 2 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| coches_no_2_14 | −1 | 3 | 4 | 3 | 6 | 3 | 6 | 3 | 6 | 3 | 6 |
| coches_no_2_16 | −1 | 7 | 3 | 7 | 6 | 8 | 3 | 7 | 3 | 8 | 3 |
| coches_no_2_17 | −1 | 14 | 5 | 18 | 8 | 20 | 8 | 18 | 8 | 20 | 8 |
| coches_no_2_20 | −1 | 7 | 2 | 8 | 8 | 9 | 9 | 8 | 8 | 9 | 9 |
| coches_no_2_22 | −1 | 2 | 6 | 3 | 6 | 3 | 8 | 3 | 6 | 3 | 8 |
| coches_no_2_24 | −1 | 13 | 10 | 13 | 12 | 17 | 12 | 13 | 11 | 17 | 12 |
| coches_no_2_7 | −1 | 8 | 7 | 11 | 7 | 13 | 8 | 11 | 7 | 13 | 8 |
| coches_no_2_9 | −1 | 3 | 6 | 3 | 11 | 6 | 11 | 3 | 11 | 6 | 11 |

**Table 14**
Polarity classification over the SFU corpus using lexicons without taking into account the domain (eSOLLocal, eSOLLocal*, eSOLGlobal and eSOLGlobal*).

| | Macro-P | Macro-R | Macro-F1 | Accuracy | Improvement (%) |
|---|---|---|---|---|---|
| iSOL | 0.6452 | 0.6125 | 0.6284 | 0.6125 | – |
| eSOLLocal | 0.6950 | 0.6438 | 0.6684 | 0.6438 | 6.36 |
| eSOLLocal* | 0.7067 | 0.6125 | 0.6562 | 0.6125 | 4.42 |
| eSOLGlobal | 0.6950 | 0.6438 | 0.6684 | 0.6438 | 6.36 |
| eSOLGlobal* | 0.6953 | 0.6250 | 0.6583 | 0.6250 | 4.75 |

However, the differences between the eSOLdomainGlobal lexicons and eSOLdomainGlobal* are not significant because they achieved very similar results, so we consider that the eSOLdomainGlobal resource has the most suitable list of opinion bearing words, because it adds less words than eSOLdomainGlobal*. Although we should emphasise that the performance with the domains computers and movies is not good.

As we have said previously the domain adaptation process has not worked as we expected for some domains. One of the problematic domains is "Computers". After reading some of the reviews we have noticed that in the some reviews the author expresses his disagreement and also advice the purchase of distinct computer. Thus in the same review there are positive and negative expressions that could have driven the domain adaptation method to introduce unsuitable words to the lists.

### 5.1. Negation and irony

A deep analysis of the results shows that the number of FP is quite high. After reading some of the test reviews we can see that some possible causes of the misclassification are associated with the poor treatment of some issues in opinion mining: negation and irony. For example, some reviews that belong to the negative class use negative expressions with positive words to state a negative opinion. An example can be read in file "no_2_4.txt"(Fig. 1) of the Movies domain: *no me ha gustado* (I did not like it). The word *gustado* (liked) is in the positive lists of eSOLLocal and eSOLGlobal, so the system considers the word as positive, but it must be negative because the word *no* (not) changes the polarity of *gustado* (liked). Another example can be found in file "no_2_12.txt" (Fig. 2) of the Music domain. This file includes the sentence *sin ideas geniales* (without brilliant ideas). The word *geniales* is also in the positive lists of all resources, so the system considers it as positive. As in the previous example the word *sin* (without) changes the polarity of the word *geniales* (brilliant).

However, this kind of error could not be associated to the lexicon because the lexicon only includes bearing words. On the contrary, it is necessary to perform a deeper analysis of the content and develop strategies for dealing with negation.

On the other hand, one of the features of irony is the use of positive words to express a negative point of view about something or somebody. After reading some reviews of the corpus the use of irony is very common in some domains. The expression *!una maravilla!* (it is wonderful!) in the review "no_1_20.txt" (Fig. 3) of the Washing machines domain is a clear example.

These are the main reasons for the low performance in some domains, so the errors are not caused by a low quality of the lexicons. Thus the main problem is that the classifier built for domain lexicons assessment only takes into account the words that are on the lists and does not consider other issues of OM. The classifier does not consider negation or irony because the main goal of the paper is the description of the new domain specific sentiment lexicons.

### 5.2. Evaluating the lexicons over other corpus

To finish our analysis of results, we would like to evaluate the validity of the generated lexicons by comparing the system with other corpora. However, the availability of Spanish corpora is very sparse, so this evaluation is very difficult to carry out.

...Espero que algún día alguien en Hollywood lea "Casa de muñecas", de Ibsen, y sepan que hay mujeres que no se rinden al mandato de la vida cómoda y standar, que quizás alguna sepa dar portazo a tanta tonteria. **No me ha gustado**. Se nota, ¿no?

...I hope one day somebody in Hollywood reads ''House of dolls'' by Ibsen, and they will know that there are women who do not surrender to the command of comfortable living standard, and also who will know slamming to such foolishness. **I did not like it**. You notice, don't you?

**Fig. 1.** Excerpt of the review "no_2_4.txt" from the movies domain.

... Tras disolver Smashing Pumpkins e intentar volar por si sólo Billy ha tenido que volver al grupo que le dió el éxito en el pasado (eso sí, sólo se mantienen 2 miembros originales) pero ha vuelto **sin ideas geniales** y se ha dedicado a recrear un Revival (bastante mediocre) de lo que el grupo...

... After dissolving Smashing Pumpkins and trying to go in his own business Billy had to return to the group that gave him success in the past (only two original members remain) but he has returned **without brilliant** ideas and has dedicated to recreate a Revival (pretty average) of the group...

**Fig. 2.** Excerpt of the review "no_2_12.txt" from the music domain.

...pero ni acero ni nada, baquelita o malquita o no sé qué historias, que se parte por la mitad. Y eso sí, hasta que acabó de romperse, me pasé un par de meses recogiendo del suelo el agua que se salía del aparato **¡una maravilla!**

...but not steel or anything, bakelite or malquita or what stuff to be split in half isn't known . And yes, until finally they break, I spent a couple of months collecting water out of the appliance, **it is wonderful!**

**Fig. 3.** Excerpt of the review "no_1_20.txt" from the washing machines domain.

**Table 15**
Polarity classification over the MC corpus using iSOL and eSOLMovieGlobal lexicons.

| | Macro-P | Macro-R | Macro-F1 | Accuracy | Improvement (%) |
|---|---|---|---|---|---|
| iSOL | 0.6222 | 0.6147 | 0.6184 | 0.6183 | – |
| eSOLMovieGlobal | 0.6253 | 0.6151 | 0.6206 | 0.6198 | 0.35 |

Also, for a complete evaluation of all the lists we need eight different corpora, one per each domain. The only Spanish corpus available is a corpus of movie reviews. The corpus is called MuchoCine (Cruz et al., 2008). Thus, only the list that achieved better results in the Movie domain (eSOLMovieGlobal) has been evaluated with the corpus MuchoCine. The results achieved by iSOL and eSOLMovieGlobal are shown in Table 15. The evaluation of eSOLMovieGlobal with the corpus MuchoCine has shown that the domain adaptation method presented in this paper is also valid for other corpus.

## 6. Conclusions and further work

In this paper we study the integration of domain information for a Spanish polarity classification system. We have carried out several experiments in order to test the different resources generated from the original Spanish lexicon iSOL: a polarity classification of each domain using iSOL; a polarity classification with the sentiment lexicons eSOLdomainLocal and eSOLdomainLocal*; the same ex- periment but with the lexicons eSOLdomainGlobal and eSOLdomainGlobal*; and the last experiments with the lists iSOL, eSOLLocal, eSOLLocal*, eSOLGlobal and eSOLGlobal*. All these resources are freely available for research purposes.[7]

The results obtained in the polarity classification of the entire corpus indepen- dently for the domain, the lexicons eSOLLocal, eSOLLocal*, eSOLGlobal and eSOLGlobal* are very similar, although we highlight that eSOLLocal and eSOLGlobal achieve better results than eSOLLocal* and eSOLGlobal*. The four lists surpass the results achieved by iSOL, but the differences between them are not significant.

However, according to the domain polarity classification the results over four domains surpass the baseline, while the other four domains seem to be harder to classify. An analysis of the errors shows that the possible cause of the misclassification could be the use of irony and negation in these reviews. Thus, our future work would be focused on the development of techniques for the treatment of negation in OM with the goal of improving the polarity classification systems. Another research line for the future is the analysis whether the application of a homogeneous factor for all the domains is a good strategy, because the analysis of the results shows up that it is very likely that each domain needs its own factor in Eq. (1).

## Acknowledgements

This work has been partially supported by a grant from the Fondo Europeo de Desarrollo Regional (FEDER), ATTOS project (TIN2012-38536-C03-0) from the Spanish Government. The project AORESCU (P11-TIC-7684 MO) from the regional government of Junta de Andalucía partially supports this manuscript, and the project CEATIC-2013-01 from the University of Jaén.

## References

Agić, Ž., Ljubešić, N., & Tadić, M. (2010). Towards sentiment analysis of financial texts in croatian. In N.C.C. (Chair), K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, & D. Tapias (Eds.), *Proceedings of the Seventh conference on international language resources and evaluation (LREC'10)* (pp. 19–21). Valletta, Malta; ELRA: European Language Resources Association.

Banea, C., Mihalcea, R., Wiebe, J., & Hassan, S. (2008). Multilingual subjectivity analysis using machine translation. In *Proceedings of the conference on empirical methods in natural language processing EMNLP '08* (pp. 127–135). Stroudsburg, PA, USA: Association for Computational Linguistics.

Blitzer, J., Dredze, M., & Pereira, F. (2007). Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In *Proceedings of the 45th annual meeting of the association of computational linguistics* (pp. 440–447). Prague, Czech Republic: Association for Computational Linguistics.

[7] urlhttp://sinai.ujaen.es/?p=1264.

Brooke, J., Tofiloski, M., & Taboada, M. (2009). Cross-linguistic sentiment analysis: From english to spanish. In *Proceedings of the international conference RANLP-2009* (pp. 50–54). Borovets, Bulgaria: Association for Computational Linguistics.

Cruz, F., Troyano, J. A., Enriquez, F., & Ortega, J. (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.

Denecke, K. (2008). Using sentiwordnet for multilingual sentiment analysis. In *IEEE 24th international conference on data engineering workshop, 2008 (ICDEW 2008)* (pp. 507–512).

Du, W., Tan, S., Cheng, X., & Yun, X. (2010). Adapting information bottleneck method for automatic construction of domain-oriented sentiment lexicon. In *Proceedings of the third ACM international conference on Web search and data mining WSDM '10* (pp. 111–120). New York, NY, USA: ACM.

Esuli, A., & Sebastiani, F. 2006. Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proceedings of the 5th conference on language resources and evaluation (LREC 2006)* (pp. 417–422).

Ghorbel, H., & Jacot, D. (2011). Sentiment analysis of French movie reviews. In V. Pallotta, A. Soro, & E. Vargiu (Eds.), *Advances in distributed agent-based retrieval tools. Studies in computational intelligence* (Vol. 361, pp. 97–108). Berlin, Heidelberg: Springer.

Hatzivassiloglou, V., & McKeown, K. R. (1997). Predicting the semantic orientation of adjectives. In *Proceedings of the eighth conference on European chapter of the association for computational linguistics EACL '97* (pp. 174–181). Stroudsburg, PA, USA: Association for Computational Linguistics.

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on knowledge discovery and data mining KDD '04* (pp. 168–177). New York, NY, USA: ACM.

Jiang, J., & Zhai, C. (2007). Instance weighting for domain adaptation in nlp. In *Proceedings of the 45th annual meeting of the association of computational linguistics* (pp. 264–271). Prague, Czech Republic: Association for Computational Linguistics.

Martínez-Cámara, E., Martín-Valdivia, M., & Ureña-López, L. (2011). In *Opinion classification techniques applied to a spanish corpus. Natural language processing and information systems* (Vol. 6716, pp. 169–176). Berlin/ Heidelberg: Springer. 10.1007/978-3-642-22327-3_17.

Martínez-Cámara, E., Martín-Valdivia, M. T., Molina-González, M. D., & Ure na López, L. A. (2013). Bilingual experiments on an opinion comparable corpus. In *Proceedings of the 4th workshop on computational approaches to subjectivity, sentiment and social media analysis* (pp. 87–93). Atlanta, Georgia: Association for Computational Linguistics.

Molina-González, M. D., Martínez-Cámara, E., Martín-Valdivia, M. T., & Perea-Ortega, J. M. (2013). Semantic orientation for polarity classification in spanish reviews. *Expert Systems with Applications, 40*, 7250–7257.

Pan, S. J., Ni, X., Sun, J.-T., Yang, Q., & Chen, Z. (2010). Cross-domain sentiment classification via spectral feature alignment. In *Proceedings of the 19th international conference on World Wide Web WWW '10* (pp. 751–760). New York, NY, USA: ACM.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval, 2*, 1–135.

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up?: Sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on empirical methods in natural language processing* (pp. 79–86). Stroudsburg, PA, USA: Association for Computational Linguistics volume 10 of EMNLP '02.

Ponomareva, N., & Thelwall, M. (2012). Do neighbours help?: An exploration of graph-based algorithms for cross-domain sentiment classification. In *Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning EMNLP-CoNLL '12* (pp. 655–665). Stroudsburg, PA, USA: Association for Computational Linguistics.

Rushdi-Saleh, M., Martín-Valdivia, M. T., Ureña López, L. A., & Perea-Ortega, J. M. (2011). Oca: Opinion corpus for arabic. *Journal of the American Society for Information Science and Technology, 62*, 2045–2054.

Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Computing Surveys, 34*, 1–47.

Sidorov, G., Miranda-Jiménez, S., Viveros-Jiménez, F., Gelbukh, A., Castro-Sánchez, N., Veláquez, F., et al (2013). Empirical study of machine learning based approach for opinion mining in tweets. In I. Batyrshin & M. González Mendoza (Eds.), *Advances in artificial intelligence. Lecture notes in computer science* (Vol. 7629, pp. 1–14). Berlin Heidelberg: Springer.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational Linguistics, 37*, 267–307.

Taboada, M., & Grieve, J. 2004. Analyzing appraisal automatically. Technical Report Stanford University.

Turney, P. D. (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* (pp. 417–424). Stroudsburg, PA, USA: Association for Computational Linguistics.

Xia, R., Zong, C., Hu, X., & Cambria, E. (2013). Feature ensemble plus sample selection: Domain adaptation for sentiment classification. *IEEE Intelligent Systems, 28*, 10–18.

Zhang, C., Zeng, D., Li, J., Wang, F.-Y., & Zuo, W. (2009). Sentiment analysis of chinese documents: From sentence to document level. *Journal of the American Society for Information Science and Technology, 60*, 2474–2487.