Expert Systems with Applications 40 (2013) 7250–7257

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

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

Semantic orientation for polarity classification in Spanish reviews

M. Dolores Molina-González [a], Eugenio Martínez-Cámara [a,*], María-Teresa Martín-Valdivia [a], José M. Perea-Ortega [b]

[a] SINAI Research Group, University of Jaén, Campus Las Lagunillas, 23071 Jaén, Spain

[b] European Commission, Joint Research Centre (JRC), Institute for the Protection and Security of the Citizen (IPSC), Via Fermi, 2749 21027 Ispra (VA), Italy

ARTICLE INFO

Keywords:
Sentiment polarity detection
Multilingual opinion mining
Spanish resources for sentiment analysis

ABSTRACT

Until now most of the published methods for polarity classification have been applied to English texts, but other languages are becoming increasingly important. This paper presents a new resource for the Spanish sentiment analysis research community. We have generated a new lexicon by translating into Spanish the Bin Liu English Lexicon. In order to assess the validity of the proposed lexicon a set of experiments on a Spanish review corpus are presented. In addition, the resource presented is compared with another existing Spanish lexicon. The results show that our resource outperforms the currently available Spanish lexicon for sentiment analysis.

1. Introduction

Recently, interest in Opinion Mining (OM) has grown significantly due to different factors. On the one hand, the rapid evolution of the World Wide Web has changed our view of the Internet. It has turned into a collaborative framework where technological and social trends come together, resulting in the over exploited term Web 2.0. On the other hand, the tremendous use of e-commerce services has been accompanied by an increase in freely available online reviews, comments and opinions about products and services. Web sites such as Amazon,[1] Epinions[2] or IMDb,[3] are queried everyday by customers who want to buy a product and are interested in other buyer's opinions. However, the huge amount of information makes it necessary to develop new methods and strategies to tackle the problem.

Sentiment analysis (SA) systems can be both helpful and influential not only for individual customers but also for any company or institution. These systems automatically accumulate feedback and comments originating from multiple sources, effectively aggregate this information, and present the results in an appropriate way to the user. Thus, SA is becoming one of the main research areas that combines Natural Language Processing (NLP) and Text Mining (TM) to automatically identify and analyze opinions and emotions in documents (Tsytsarau & Palpanas, 2012).

Several subtasks related to SA have been studied such as subjectivity detection (Wiebe, Wilson, & Bell, 2001), review summarization (Somprasertsri & Lalitrojwong, 2010), humor detection (Mihalcea & Strapparava, 2006) and emotion classification (Strapparava & Mihalcea, 2008). One of the most widely studied tasks is sentiment classification, which focuses on determining the polarity of a document, sentence or feature (positive or negative) and on measuring the degree of the polarity expressed in the document (Pang & Lee, 2008). Polarity classification aims to classify a subjective text as positive or negative, according to the overall sentiment expressed by the author. Thus, given a subjective text a sentiment classifier must determine whether the opinion is positive or negative. Although different approaches have been applied to the field of polarity classification, the mainstream basically consists of two major methodologies. On the one hand, the Machine Learning (ML) 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 computing the semantic orientation (SO) of the words in the texts does not need prior training, but takes into account the orientation of words, positive or negative (Turney, 2002). Both methodologies have their advantages and drawbacks. For example, the ML approach requires training data, which in many cases are difficult or impossible to obtain, partially due to the novelty of the task. On the contrary, the SO approach requires a large amount of linguistic resources which generally depend on the language. In order to take advantage of both methods, some studies apply a hybrid approach (Prabowo & Thelwall, 2009) (Martín-Valdivia, Martínez-Cámara, Perea-Ortega, & Ureña-López, 2013). Usually, the ML approaches obtain better results and currently we can find very good systems working over different domains (Rushdi-Saleh, Martín-Valdivia, Ureña-López, &

* Corresponding author. Tel.: +34 953 21 19 56.
E-mail addresses: mdmolina@ujaen.es (M.D. Molina-González), emcamara@ujaen.es (E. Martínez-Cámara), maite@ujaen.es (M.-T. Martín-Valdivia), jose-manuel.perea-ortega@jrc.ec.europa.eu (J.M. Perea-Ortega).

[1] http://www.amazon.com.
[2] http://www.epinions.com.
[3] http://www.imdb.com.

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.06.076

Perea-Ortega, 2011c). However, the SO method needs more research in order to obtain similar ML results. This is one of the reasons why this paper is focused on semantic orientation for polarity classification.

Another reason is concerned with language. Although opinions and comments in the Internet are expressed in any language, most of the research in OM, and specifically in polarity classification, only deals with English documents. However, languages such as Chinese (Tan & Zhang, 2008), Spanish (Martín-Valdivia et al., 2013) or Arabic (Rushdi-Saleh, Martín-Valdivia, Ureña-López, & Perea-Ortega, 2011a), are ever more present on the web. Therefore, it is important to develop resources to help researchers to work with these languages.

There are two main ways of addressing the problem of applying SA to non-English languages: on the one hand, we can generate resources for the target language, for example corpora, dictionaries, and lists of opinion words. These resources are then used in order to carry out the classification process. On the other hand, we can extract information in the target language, for example in Spanish or Arabic, and translate it into English. This information can then be managed using the available English resources like SentiWordNet (Esuli & Sebastiani, 2006) or WordNet Affect (Strapparava & Valitutti, 2004). This second approach has been successfully applied in several studies, for example translating into German (Denecke, 2008), Arabic (Rushdi-Saleh, Martín-Valdivia, Ureña-López, & Perea-Ortega, 2011b) or Spanish (Martín-Valdivia et al., 2013). However, the generation of resources for the target language is a more difficult and time-consuming task that requires deeper research. Some corpora have been created in other languages than English in order to apply them to a polarity classification system, for example in Arabic (Rushdi-Saleh et al., 2011a) and Chinese (Zhang, Zeng, Li, Wang, & Zuo, 2009). Although we can find some lexicons in several languages, it is noteworthy that there are very few resources for Spanish. Therefore, another motivation of this work is to investigate the effect of using a Spanish lexicon over a corpus of reviews.

In this paper we present a new Spanish resource for OM composed of a list of opinion words; SOL (Spanish Opinion Lexicon). Our main goal is to develop a Spanish lexicon based on one of the most widely-used English lexicons for polarity classification (we will call it BLEL: the Bing Liu English Lexicon). Specifically, we focus on the use of opinion words. In the research literature opinion words are also known as polar words, opinion-bearing words, and sentiment words. Positive opinion words are used to express desired states while negative opinion words are used to express undesired states. Apart from individual words, there are also opinion phrases and idioms. Collectively, they are called the opinion lexicon.

Thus, we have taken the BLEL[4] (Hu & Liu, 2004) and have automatically translated it into Spanish, obtaining the SOL resource. Then we have manually reviewed the lexicon in order to improve the final list of words obtaining iSOL (improved SOL). In order to demonstrate the validity of this resource we have carried out several experiments over a Spanish corpus of movie reviews called MuchoCine (Cruz, Troyano, Enríquez, & Ortega, 2008). The results obtained show that the use of an improved list of sentiment words from the same language can be considered a good strategy for unsupervised polarity classification. Moreover, we have generated another list by integrating the positive and negative words present in the MuchoCine corpus. In this way, we attempt to integrate domain knowledge in the lexicon. Experiments with this enriched eSOL (enriched SOL) show the advantages of integrating external knowledge. Furthermore, we provide a comparative study between our eSOL and other recently published Spanish lexicon, which is known as Spanish Emotion Lexicon (SEL), with the aim of showing the relevance for the research community of the lexicons introduced in this paper. SEL is a resource provided by Sidorov (Sidorov et al., 2012) and it has two implementation details that worth pointing out. Firstly, SEL is composed of 2036 words. Secondly, these SEL words are associated to the measure of Probability Factor of Affective use (PFA) with respect to at least one basic emotion: joy, anger, sadness, surprise and disgust. The higher the value of the PFA, the more probable the association of the word with the emotion is.

Due to our resource is focused on opinion words, our classification is binary and therefore in order to establish a feasible comparison, we have had to consider the joy and surprise categories as positive and the other as negative words. Thus, we notice that our polarity lexicon is significantly larger than SEL and the experiments show that eSOL has improved accuracy on a reviews polarity classification task opposed to SEL.

The remainder of the paper is organized as follows: Section 2 briefly describes previous related work on semantic orientation for polarity classification and papers that study the problem regarding non-English texts. In Section 3 we explain the methodology used to build the Spanish lexicon as well as different improvements achieved. Section 4 describes the different resources used in our experiments. Section 5 presents the experiments carried out and discusses the main results obtained. Finally, we outline conclusions and further work.

2. Related work

Two main approaches can be distinguished in the field of polarity classification. On the one hand, ML techniques are more extensively used for the classification of reviews. In this approach, the document is represented by different features that may include the use of *n*-grams or defined grammatical roles like, for instance, adjectives or other linguistic feature combinations. Then a machine learning algorithm is applied. Commonly used machine learning algorithms are Support Vector Machines (SVM), Maximum Entropy (ME) and Naïve Bayes (NB). A survey of studies using ML can be found in Pang and Lee (2008), Liu (2012) or Tsytsarau and Palpanas (2012).

On the other hand there is a lot of work based on the semantic orientation approach, which represents the document as a collection of words. Then the sentiment of each word can be determined by different methods, for example using a web search (Hatzivassiloglou & Wiebe, 2000) or consulting a lexical database like WordNet[5] (Kamps, Marx, Mokken, & de Rijke, 2004). Regarding methods that consider some linguistic features such as adjectives and adverbs, we can find many studies in the literature (Ding & Liu, 2007; Hatzivassiloglou & McKeown, 1997; Kamps et al., 2004; Turney, 2002; Wiebe, 2000). Specifically, our paper is based on the paper by Hu and Liu (2004).

Regarding polarity classification using non-English languages, we can find some interesting studies that apply a semantic orientation approach based on sentiment words. Kim and Hovy (2006) compared opinion expression between an aligned corpus of emails in German and English. One of their experiments translates English opinion-bearing words into German and then analyzes German emails using the German opinion-bearing words. Zhang et al. (2009) applied Chinese sentiment analysis to 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). They proposed a rule-based approach including two phases: firstly, determining each

[4] Available in http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html#lexicon.

[5] http://wordnet.princeton.edu.

sentence's sentiment based on word dependency, and secondly, aggregating sentences in order to predict the document sentiment. Wan (2009) studied how to reduce the need of using Chinese linguistic resources for SA in Chinese. The author followed a supervised approach and proposed a co-training system based on the use of an English corpus for polarity classification of Chinese products reviews and the usage of a machine translation system.

Finally, there are also some remarkable studies regarding polarity classification focused on Spanish using SO based on bearing words lists. For example, Baena et al. (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 work showed that automatic translation is a viable alternative for the construction of resources and tools for subjectivity analysis in a new target language. Cruz et al. (2008) gathered a corpus of Spanish movie reviews from the MuchoCine website[6]. The MuchoCine (MC) corpus was manually annotated and used to develop a polarity classifier based on the semantic orientation of the words. Brooke, Tofiloski, and Taboada (2009) presented several experiments dealing with Spanish and English resources. 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).

3. Sentiment word lists for Spanish

Three main approaches exist for the compilation of a set of polar words: the manual approach, dictionary-based approach and corpus-based approach. The manual approach is tedious and time consuming, so it is not usually used. However, the manual method is used combined with automated approaches as the final check, because automated ones may make mistakes.

The dictionary-based approach consists of taking manually a small set of sentiment words as seeds with known positive or negative orientations. The following step is enlarging the initial set of seeds by searching in a lexical knowledge base such as WordNet (Miller, Beckwith, Fellbaum, Gross, & Miller, 1990) for their synonyms and antonyms. The newly found words are included in the seed list. It is an iterative process which ends when no more new words can be found. An example of this method is the paper (Hu & Liu, 2004) where BLEL is presented.

The corpus-based approach is usually applied in two different situations:

1. Given a list of polar words, encounter other opinion words and their polarity from a domain sentiment label corpus.
2. To adapt a general-purpose sentiment lexicon to a new one using a domain corpus for SA applications in the domain.

A good representative of this method is the work of Kanayama and Nasukawa (2006).

Each method has its advantages and drawbacks. The dictionary-based approach is more suitable for the compilation of general-purpose lexicons, while the corpus-based method is better for the generation of domain-dependent sentiment lexicons.

Most of the studies on sentiment words only deal with English documents, perhaps due to the lack of resources in other languages. However, according to the Internet World Stats[7], the number of Internet users with Spanish as their source language is 8%, third after English and Chinese. For this reason, we consider the need to develop a resource as complete as possible composed of sentiment words for Spanish that would be useful for further research activities. This resource was developed in incremental versions which are described in detail below.

[6] http://www.muchocine.net

Table 1
Examples of English words with same meaning in Spanish.

Several English words	Spanish meaning
Bogus disingenuous dud false phony spurious untrue	*Falso*
Castigate chasten chastise penalize punish	*Castigar*
Crabby glum ill-tempered moody peevish sullen	*Malhumorado*
Absurd absurdness farcical ludicrous preposterous	*Absurdo*
Gaily jolly joyfully joyously merrily	*Alegremente*
Beautifully gloriously marvelously splendidly wonderfully	*Maravillosamente*
Bright lustrous shiny sparkling twinkly	*Brillante*
Affordable economical low-cost low-priced thrifty	*Económico*

3.1. Original list SOL (Spanish Opinion Lexicon)

In a first version we generated a parallel list of sentiment words in Spanish from the opinion lexicon in English provided by Liu[4] (BLEL). This resource was generated by applying automatic machine translation techniques and is composed of approximately 6800 positive and negative opinion words. Reverso[8] was used as an automatic machine translation system, taking into account the first translated word that the system returned for each original word from BLEL. In the process, 1068 negative and 364 positive opinion words were eliminated because their meanings were the same. Table 1 shows some examples of these English words that shared the same first translations.

During the process, we found words that do not have a translation in Spanish with Reverso. We noticed that some of these words were misspelled in the BLEL lists, but they should not be considered mistakes, because they appear frequently in social media content. The rest of words simply have not been recognized by Reverso. Due to both reasons, 435 negative and 159 positive words have been discarded in our resource. Table 2 shows some examples of misspelled and not recognized words.

Other conventions followed when we generated the list were related to writing criteria. For example, each word was written by using non capital letters, without special characters and without accented vowels.

Finally, our first lexicon is composed of approximately 4800 positive and negative opinion words. This resource was called SOL[9] (Spanish Opinion Lexicon).

3.2. Improved SOL (iSOL)

After generating the SOL list, we decided to improve it by addressing some issues that were raised. The resulting list was called iSOL[10] (improved SOL).

At first, we included misspelled Spanish words frequently used in sentiment opinion following the philosophy of the BLEL. Table 3 shows some examples of these misspelled Spanish words.

One of the problems we had to take into consideration was related to the fact that the translation of an English word returned two or more words. For these cases we had to assign manually

[7] . Estimation of the number of Internet users by language as of 31 May 2011.

[8] http://www.reverso.net

[9] http://sinai.ujaen.es/?p=1224.

[10] http://sinai.ujaen.es/?p=1202.

Table 2
Examples of discarded words.

Misspelled words	Not recognized words
Assult	Bonny
Goood	Fav
Prospros	Pettifog
Slooow	Bumpping
2-faces	Jollily
Danken	Brainiest
Jutter	Prik

Table 3
Examples of possible translations of misspelled Spanish words in English.

Misspelled Spanish word	Possible translated English word
Cool	*Cool*
Kaput	*Thumbs down*
Pillin	*Naughty*
Coñacete	*Pain in the neck*
Top	*Number one*

Table 4
Examples of some manually reviewed translations.

English	Automatic Spanish translation	Manual assignment
Brainless	*Sin cerebro*	*Descerebrado*
Aimless	*Sin rumbo*	*Desorientado*
Arrogantly	*Con arrogancia*	*Arrogantemente*
Deadlock	*Punto muerto*	*Estancado*
Worthless	*Sin valor*	*Devaluado*
Fashionable	*A la moda*	*Moderno*

the best synonym (composed of only one term) for the translated word. Table 4 shows some examples of these translations performed manually.

Another issue was related to the repetition of words in both lists. For these cases we decided to discard them. A total of 36 words were discarded. Some examples of these words were: *ansioso*, *presumido*, *aturdido*, *increible*, exaltado, *asombrar*, etc.

Finally, the last issue was related to the genre and number present in Spanish grammar. While an English adjective has neither genre nor number and is usually represented by a single term, a Spanish adjective can have four possible translated words, two for the genre (male or female) and two for the number (singular or plural). Table 5 shows some examples of possible translations of English adjectives in Spanish.

3.3. Enriched SOL (eSOL)

The two lexicons described are general-purpose sentiment lexicons. 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 and McKeown (1997) take some adjectives as seeds to find additional sentiment adjectives in the corpus. Their method took advantage of a set of conventions on connectives with the aim of identifying more polar words and their orientation from a sentiment label corpus.

Taking as baseline the lexicon iSOL, we generated a list of opinion words for the cinema domain. We followed the corpus-based approach. The key element of the corpus-based approach is the use of a sentiment labeled corpus. The Spanish corpus selected

Table 5
Examples of possible translations of English adjectives in Spanish.

English	Spanish
Good	*Bueno, buena, buenos, buenas*
Famous	*Famoso, famosa, famosos, famosas*
Pretty	*Guapo, guapa, guapos, guapas*
Ugly	*Feo, fea, feos, feas*
Aching	*Dolido, dolida, dolidos, dolidas*
Bad	*Malo, mala, malos, malas*

Table 6
Number of sentiment words in resources.

Resourses	Number of negative words	Number of positive words
BLEL	4783	2006
SOL	3280	1483
iSOL	5626	2509
eSOL	5639	2536

for the process was MuchoCine (MC), which is described in detail in Section 4.

We followed the same assumption as (Du, Tan, Cheng, Yun, 2010), i.e. a word should be positive (or negative) if it appears in many positive (or negative) documents. Thus, we calculated the word frequency in each class of documents (positive and negative). We found about 15 negative words and 25 positive words. Therefore, these 40 most frequent words that were not yet contained in the iSOL list were added to the final list. This new list integrating information from the corpus was called eSOL[11] (enriched SOL).

Finally these lists have been freely made available as a lexical resource of positive and negative opinion words[9] for use in sentiment analysis for Spanish.

As can be seen in Table 6, we have increased the size of the iSOL and eSOL lists for both negative and positive lists of words with regard to the original list provided by Liu (BLEL). Specifically, for the iSOL list 843 negative and 503 positive words were added, while for the eSOL list 856 negative and 530 positive words were added.

4. Experimental framework

This section presents the measures employed for evaluating the experiments carried out in this paper. Moreover, the main features of the MuchoCine corpus are also shown.

4.1. Evaluation measures

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

$$P = \frac{TP}{TP + FP}$$
$$R = \frac{TP}{TP + FN}$$
$$F1 = \frac{2PR}{P + R}$$
$$Acc = \frac{TP + TN}{TP + TN + FP + FN}$$

where TP (True Positives) are those assessments where the system and a human expert agree on a label, FP (False Positives) are those labels assigned by the system that do not agree with the expert assignment, and FN (False Negatives) are those labels that the sys-

[11] http://sinai.ujaen.es/?p=1188.

Table 7
Rating distribution.

Rating	#Reviews
1	351
2	923
3	1253
4	890
5	461
Total	3875

Table 8
Binary classification of the MC corpus.

Classes	#Reviews
Positive	1274
Negative	1351
Total	2625

tem failed to assign as they were given by the human expert. F1 is a measure that combines both precision and recall, calculating the proportion of true results (both true positives and true negatives) (Sebastiani, 2002). For ease of comparison, we summarize the F1 scores over the different categories (positive and negative) using the macro-averages of F1 scores:

$$\text{Macro-F1} = \text{Average of within-category F1 values}$$

In the same way, we can obtain the Macro-Recall and Macro-Precision as follows:

$$\text{Macro-Recall} = \text{Average of within-category Recall values}$$
$$\text{Macro-Precision} = \text{Average of within-category Precision values}$$

4.2. The MC corpus

In order to demonstrate the effectiveness of our approach we selected the MuchoCine corpus (MC), available for the SA research community in Spanish (Cruz et al., 2008). The corpus consists of 3878 movie reviews collected from the MuchoCine website. The reviews are written by web users instead of professional film critics. This increases the difficulty of the task because the sentences found in the documents may not always be grammatically correct, or they may include spelling mistakes or informal expressions. The corpus contains about 2 million words and an average of 546 words per review.

The opinions are rated on a scale from 1 to 5. One point means that the movie is very bad and 5 means very good. Films with a rating of 3 can be considered as "neutral", which means that the user considers the film is neither bad nor good. Table 7 shows the number of reviews per rating. This corpus has been widely used in different studies such as (del-Hoyo, Hupont, Lacueva, & Abadía, 2009), (Barreiro & Gonçalo, 2011), (Malvar-Fernández & Pichel-Campos, 2011), (Martínez-Cámara, Martín-Valdivia, & Ureña-López, 2011) and (Martín-Valdivia et al., 2013).

In our experiments we discarded the neutral examples. In this way, opinions rated with 3 were not considered, the opinions with ratings of 1 or 2 were considered as positive and those with ratings of 4 or 5 were considered as negative. Table 8 shows the class distribution of the binary classification of MC.

In MC corpus a movie review consists of four fields: the identifier of the review, the review rating, the summary and the description. Fig. 1 shows an excerpt of a review from MC.

id|rating|summary|body

1000|-1|*Silicona, esteroides, pactos demoniacos y otras basuras habituales son la base que sustentan esta aberración de vergüenza.| Una fiesta llena de excesos, rubias despampanantes, musculitos por doquier, algún que otro muerto. Nada nuevo. La alianza del mal es el nombre de este thriller sobrenatural que narra las peripecias de unos jóvenes…*

Fig. 1. Excerpt of a review from the MuchoCine corpus.

5. Experiments and results

Several experiments were carried out in order to verify the utility of the three lists of sentiment words generated for Spanish: SOL, iSOL and eSOL.

Before carrying out the experiments we performed a preprocessing step to the MC corpus in order to apply the same criteria followed during the generation of the lists. For example, for both summary and body we had to change capital letters to non-capital letters, accented letters to non-accented letters and special characters had to be deleted from the opinions. Moreover, the stop words and proper nouns were discarded. The named entity recognition was carried out using the Freeling tool[12].

In order to decide whether a review was positive or negative we followed a simple approach based on counting the number of words included in the lists of sentiment words for Spanish. Therefore, our system assesses the review as positive if the number of positive words is greater than or equal to the negative ones, and as negative in the opposite case. Using this approach, we carried out the binary classification of the MC corpus by using these three lists of sentiment words. Table 9 shows the comparison between the different results obtained.

As shown in Table 9, using the last versions of the lexicon (improved and enriched) we obtained the best results for all the measures employed. Employing the Macro-F1 as the evaluation measure and the approach applying SOL as base case, the system using the iSOL achieves an improvement of +10.29%, while using eSOL the improvement achieved is +12.95%. Therefore, we can conclude that the most completed list of sentiment words provided (eSOL) can be considered an interesting resource for use in sentiment analysis tasks related to Spanish language, particularly for unsupervised approaches.

5.1. Comparison with other related work

In the literature we can find an interesting resource called the Spanish Emotion Lexicon (SEL[13]) provided by Sidorov et al. (2012). This resource is freely available[11] for research purposes. SEL is composed of 2036 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*. It was marked manually by 19 annotators by using a scale with four values: null, low, medium and high.

In order to establish a feasible comparison by using the SEL resource for binary classification of MC, we considered the *joy* and *surprise* categories as positive and the others as negative. We carried out two different experiments: taking into account all the words provided by SEL and considering only those words whose PFA value was greater than or equal to 0.2. Table 10 shows the comparison between the results obtained by using the SEL resource and those obtained by using the eSOL resource provided in this paper.

Regarding the results in Table 10 the words included in SEL are

[12] http://nlp.lsi.upc.edu/freeling.

[13] http://www.cic.ipn.mx/~sidorov/#SEL.

Table 9
Results obtained for the binary classification of the MC corpus by using the three lists of sentiment words generated.

	Macro-precision	Macro-recall	Macro-F1	Accuracy
SOL	0.5615	0.5600	0.5607	0.5623
iSOL	0.6222	0.6147	0.6184	0.6183
eSOL	**0.6393**	**0.6274**	**0.6333**	**0.6316**

Table 10
Comparison for binary classification of MC by using the SEL and eSOL resources.

	Macro-precision	Macro-recall	Macro-F1	Accuracy
SEL (all words)	0.5240	0.5162	0.5200	0.5249
SEL (PFA > 0,2)	0.5256	0.5181	0.5218	0.5264
eSOL	**0.6393**	**0.6274**	**0.6333**	**0.6316**

Table 11
Some samples of MC corpus classified with eSOL and SEL resources.

Id	Original rating	#positive words in eSOL	#negative words in eSOL	eSOL rating	#positive words in SEL	#negative words in SEL	SEL rating
1002	−1	8	9	−1	4	0	1
1008	1	23	13	1	5	6	−1
1011	−1	8	13	−1	7	5	−1
1023	1	13	11	1	4	5	−1
1036	1	49	26	1	1	16	−1
1045	1	20	13	1	4	10	−1

Table 12
Negative and positive words classified with eSOL and SEL.

Id	Negative words found with eSOL	Negative words found with SEL	Positive words found with eSOL	Positive words found with SEL
1008	Falta Prision Impotencia Pequeña Ansiedad Intenso Nerviosa Dolor Muerte Soledad Irregular Sobria Falta	Impotencia Dolor Muerte Soledad Decaer Solo	Maestra Mayor Ventaja Prometedor Emociones Intimas Unica Profunda Mejor Estabilidad Amor Emotiva Afecto Amor Impecable Sabio Humano Maestra Mayor Ventaja Prometedor Emociones Intimas	Amor Afecto Amor Expresion Suave

not as discriminative as eSOL. Due to the low performance of SEL, we revised SEL and noticed that words with a PFA value closed to zero should not be taken into account, because those terms could introduce noise. Thus we only considered terms with a PFA value over 0.2. This subset of SEL achieved slightly better results, +0.3456%, +0.2854% regarding Macro-F1 and Accuracy respectively. However, the results achieved by SEL over 0.2 are unsatisfactory. The difference between SEL over 0.2 is noteworthy, +19.3057% and +18.1693% considering Macro-F1 and Accuracy, respectively.

Table 11 shows some samples of MC corpus classified with our resource eSOL and the resource SEL with PFA > 0.2 for the reason that we have explained before. The results show that the resource eSOL is more suitable for polarity classification of Spanish texts than the SEL resource.

In next table we show the negative and positive words that eSOL and SEL found in the review with identifier 1008.

As we can see in Table 12, negative words such as *ansiedad* (*anxiety*), *nerviosa* (*nervous*), *irregular* (*irregular*), and positive words such as *sabio* (*wise*), *prometedor* (*promising*), *maestra* (*masterpiece*) are not included in the resource SEL. This is the main reason why this review was classified correctly by using our resource eSOL and incorrectly by using SEL.

On the other hand, it is noteworthy that the MC corpus has been widely used by the SA Spanish research community. Some authors assessed different methods over MC corpus, so in Table 13 we show a comparison of our proposed method with other works.

Table 13
Polarity classification results over MC corpus.

	Approach	Precision	Recall	F1	Accuracy
Cruz et al. (2008)	Unsupervised	N/A	N/A	N/A	0.6950
	Supervised	N/A	N/A	N/A	0.7750
del-Hoyo et bal. (2009)	Hybrid	N/A	N/A	N/A	0.8086
Malvar-Fernández & Pichel-Campos (2011)	Supervised	0.77	0.77	N/A	N/A
Martínez-Cámara et al. (2011)	Supervised	0.8684	0.8667	0.8675	0.8674
Martín-Valdivia et al. (2013)	Hybrid	0.8858	0.8857	0.88575	0.8857
eSOL	Unsupervised	**0.6393**	**0.6274**	**0.6333**	**0.6316**

As usual in data mining the approaches based on supervised methods achieve better results than those based on unsupervised methods. The authors of the MC corpus followed the unsupervised method proposed by Turney (2002), which takes advantage of the search engine AltaVista[14]. That study is the only which describes an unsupervised approach over the MC Corpus. As the authors indicate in their paper, they did not use the whole corpus (3878 reviews), neither the 2625 reviews resultant of getting rid of the opinions tagged with a polarity value of 3, as we do in our experiments. They only used 400 reviews (200 positive and 200 negative) for the experimentation which had been randomly selected from the subset of 2625 reviews, i.e. the reviews labeled with a value of 1–2 (negative) or a value of 4–5 (positive). With this subset of 400 reviews the authors achieved 0.6950 of Accuracy. On the other hand, we used a subset of the original MC corpus to assess the lexicons which concerns 2625 reviews. With a larger set of data our unsupervised method achieved 0.6316 of Accuracy, only 0.0634 lower than the method proposed by the authors of the corpus. Taking into account the simplicity of our lexicon-based method, the results achieved by eSOL can be considered as good. Also, we highlighted the fact that, as far as we know, this is the first work that has used all the positive and negative reviews of the MC corpus for a polarity classification experimentation following an unsupervised method.

[14] http://www.altavista.com/

6. Conclusions and further work

In this paper we have presented an experimental study of polarity classification over a corpus of film reviews written in Spanish, the MuchoCine corpus (MC). Firstly, we translated the lexicon of BLEL in order to generate sentiment word lexicons for Spanish. Several improvements were carried out in order to build an improved sentiment words list, and finally, an enriched lexicon for Spanish.

The results show the validity of the two lexicons presented in this paper, iSOL and eSOL, for polarity classification of Spanish reviews. In addition, the results show that a lexicon-based method is suitable for solving the task of polarity classification of Spanish texts. The experiments carried out in this paper encourage us to continue working along this line. A lexicon such as iSOL and eSOL can be used as the sole semantic resource or can be used as another element within the workflow of a polarity classification system. Therefore, we consider that the lexicons developed, which are freely available, are valuable resources for the Spanish SA research community.

Currently we are working on the development of several Spanish lexicons for domain-dependent SA following the method proposed here, i.e. selecting the words with a higher frequency in a corpus. In addition, we are interested in another novelty method combining a random walk algorithm for building domain-oriented sentiment lexicons (Tan & Wu, 2011). Finally, we are studying the treatment of negation in SA, which we think is essential for the resolution of the polarity classification task.

Acknowledgements

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

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.

Barreiro, A., & Gonçalo, O. (2011). Cross-language semantic relations between English and Portuguese. In *Proceedings of the workshop on Iberian cross-language natural language processing tasks (ICL 2011)* (pp. 49–58).

Brooke, J., Tofiloski, M., & Taboada, M. (2009). Cross-linguistic sentiment analysis: From english to spanish. In *Proceedings of the RANLP Conference* (pp. 50–54).

Cruz, F., Troyano, J. A., Enríquez, F., & Ortega, J. (2008). Experiments in sentiment classification of movie reviews in Spanish. *Procesamiento de Lenguaje Natural, 41*, 73–80.

del-Hoyo, R., Hupont, I., Lacueva, F. J. & Abadía, D. (2009). Hybrid text affect sensing system for emotional language analysis. In *Proceedings of the international workshop on affective-aware virtual agents and social robots*. Boston, Massachusetts.

Denecke, K. (2008). Using SentiWordNet for multilingual sentiment analysis. In *Data Engineering Workshop, 2008. ICDEW 2008* (pp. 507–512). IEEE 24th International Conference on IEEE.

Ding, X., & Liu, B. (2007). The utility of linguistic rules in opinion mining. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 811–812). ACM.

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

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'06)* (pp. 417–422).

Hatzivassiloglou, V. & McKeown, K. (1997). Predicting the semantic orientation of adjectives. In *Proceedings of the eighth conference on European chapter of the association for computational linguistics* (pp. 174–181).

Hatzivassiloglou, V. & Wiebe, J. (2000). Effects of adjective orientation and gradability on sentence subjectivity. In *Proceedings of the international conference on computational linguistics (COLING)* (pp. 299–305).

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.

Kamps, J., Marx, M., Mokken, R. J. & de Rijke, M. (2004). Using WordNet to measure semantic orientations of adjectives. In *LREC, European Language Resources Association*.

Kanayam, H., Nasukawa, T. (2006). Fully automatic lexicon expansion for domain-oriented sentiment analysis. In *Proceedings of conference on empirical methods in natural language processing (EMNLP-2006)* (pp. 355–363).

Kim, S. M., & Hovy, E. (2006). Identifying and analyzing judgment opinions. In *Proceedings of the joint human language technology/North American chapter of the ACL conference (HLT-NAACL)* (pp. 200–207). Stroudsburg, PA, USA: Association for Computational Linguistics.

Liu, B. (2012). *Sentiment analysis and opinion mining. synthesis lectures on human language technologies*. Morgan & Claypool Publishers.

Malvar-Fernández, P., & Pichel-Campos, J. R. (2011). Generación semiautomática de recursos de Opinion Mining para el gallego a partir del portugués y el español. In *Proceedings of the workshop on Iberian cross-language natural language processing tasks (ICL 2011)* (pp. 59–63).

Martínez-Cámara, E., Martín-Valdivia, M. T., & Ureña-López, L. A. (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* (pp. 169–176). Springer-Verlag.

Martín-Valdivia, M. T., Martínez-Cámara, E., Perea-Ortega, J. M., & Ureña-López, L. A. (2013). Sentiment polarity detection in Spanish reviews combining supervised and unsupervised approaches. *Expert Systems with Applications, 40*(10), 3934–3942. http://dx.doi.org/10.1016/j.eswa.2012.12.084.

Mihalcea, R., & Strapparava, C. (2006). Learning to laugh (automatically): Computational models for humor recognition. *Journal of Computational Intelligence, 22*, 126–142.

Miller, G., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. (1990). *WordNet: An on-line lexical database*. Oxford University Press.

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 10*, (pp. 79–86).

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

Prabowo, R., & Thelwall, M. (2009). Sentiment analysis: A combined approach. *Journal of Informetrics, 3*(2), 143–157.

Rushdi Saleh, M., Martín-Valdivia, M. T., Montejo-Ráez, A., & Ureña-López, L. A. (2011c). Experiments with SVM to classify opinions in different domains. *Expert Systems with Applications, 38*(12), 14799–14804.

Rushdi-Saleh, M., Martín-Valdivia, M. T., Ureña-López, L. A. & Perea-Ortega, J. M. (2011b). Bilingual Experiments with an Arabic-English Corpus for Opinion Mining. In *Proceedings of the International Conference Recent Advances in Natural Language Processing (RANLP)* (pp. 740–745).

Rushdi-Saleh, M., Martín-Valdivia, M. T., Ureña-López, L. A., & Perea-Ortega, J. M. (2011a). OCA: Opinion corpus for Arabic. *Journal of the American Society for Information Science and Technology, 62*(10), 2045–2054.

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

Sidorov, G., Miranda-Jiménez, S., Viveros-Jiménez, F., Gelbukh, A., Castro-Sánchez, N., Velásquez, F., Díaz-Rangel, I., Suárez-Guerra, S., Treviño, A., & Gordon, J. (2012). Empirical study of opinion mining in Spanish tweets. *LNAI, 7629–7630*, 14.

Somprasertsri, G., & Lalitrojwong, P. (2010). Mining feature-opinion in online customer reviews for opinion summarization. *Journal of Universal Computer Science, 16*, 938–955.

Strapparava, C. & Mihalcea, R. (2008). Learning to identify emotions in text. In *Proceedings of the 2008 ACM symposium on applied computing (SAC '08)* (pp. 1556–1560). ACM.

Strapparava, C. & Valitutti, A. (2004). WordNet-Affect: An affective extension of WordNet. In *Proceedings of the 4th international conference on language resources and evaluation* (pp. 1083–1086).

Taboada, M., Brooke, J., Tofiloski, M., Voll, K. D., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational Linguistics, 37*(2), 267–307.

Tan, S., & Wu, Q. (2011). A random walk algorithm for automatic construction of domain-oriented sentiment lexicon. *Expert Systems with Applications, 38*(10), 12094–12100.

Tan, S., & Zhang, J. (2008). An empirical study of sentiment analysis for Chinese documents. *Expert Systems with Applications, 34*(4), 2622–2629.

Tsytsarau, M., & Palpanas, T. (2012). Survey on mining subjective data on the web. *Data Mining and Knowledge Discovery, 24*(3), 478–514. http://dx.doi.org/10.1007/s10618-011-0238-6.

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* (pp. 417–424).

Wan, X. (2009). Co-training for cross-lingual sentiment classification. In Keh-Yih Su, Jian Su, & Janyce Wiebe (Eds.), *'ACL/AFNLP', The Association for Computer, Linguistics* (pp. 235–243).

Wiebe, J. (2000). Learning subjective adjectives from corpora. In *Proceedings of AAAI*.

Wiebe, J., Wilson, T. & Bell, M. (2001). Identifying collocations for recognizing opinions. In *Proceedings of the ACL'01 association for computational linguistics workshop on collocation: computational extraction, analysis, and exploitation* (pp. 24–31).

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*(12), 2474–2487.