Article

Journal of Information Science
1–19

Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/0165551510000000
jis.sagepub.com
SAGE

Integrating Spanish Lexical Resources by Meta-classifiers for polarity classification

Eugenio Martínez-Cámara
Computer Science Department, University of Jaén, Spain

M. Teresa Martín-Valdivia
Computer Science Department, University of Jaén, Spain

M. Dolores Molina-González
Computer Science Department, University of Jaén, Spain

José M. Perea-Ortega
European Commission, Joint Research Centre (JRC), Institute for the Protection and Security of the Citizen (IPSC), Ispra, Italy.

Abstract
In this paper we focus on unsupervised Sentiment Analysis in Spanish. The lack of resources for languages other than English, as for example Spanish, adds more complexity to the task. However, we should take advantage of some good already existing lexical resources. We have carried out several experiments using different unsupervised approaches in order to compare the different methodologies for solving the problem of the Spanish polarity classification in a corpus of movie reviews. Among all these approaches, perhaps the newest one integrates SentiWordNet with the Multilingual Central Repository to tackle the polarity detection directly over the Spanish corpus. However, the results obtained are not as promising as we expected, and so we have carried out another group of experiments combining all the methods by using meta-classifiers. The results obtained with stacking outperform the individual experiments and encourage us to continue in this way.

Keywords
Unsupervised polarity detection, SentiWordNet (SWN), Multilingual Central Repository (MCR), Stacking algorithm, Meta-classifiers, Spanish lexical resources, Lexical resources for opinion mining

1. Introduction

Sentiment Analysis (SA), also known as Opinion Mining (OM), is an area of Natural Language Processing (NLP) that refers to the treatment of the subjective information in texts, mainly product reviews, comments on blogs or personal opinions. One of the basic tasks in sentiment analysis is classifying the polarity of a given text at the document, sentence, or feature level, i.e., whether the opinion expressed in a document, a sentence or an entity feature is positive, negative, or neutral. Many studies have investigated the polarity classification problem but most only consider documents written in English. However, nowadays more and more people express their comments, opinions or points of view in their own language, making languages like Spanish, Chinese or Arabic increasingly important in OM. For this reason it is necessary to develop systems than can extract and analyse all this information in different languages. In this work we focus on polarity detection for Spanish reviews. We are mainly concerned with linguistic resources for Spanish sentiment analysis because, in addition to the lack of resources for this language in this area, it is currently the third most used language in the Web according to the Internet World Stats[1].

Corresponding author:
Eugenio Martínez Cámara, Computer Science Department. University of Jaén, Campus Las Lagunillas, 23071 Jaén, Spain.
Email: emcamara@ujaen.es

On the other hand, polarity classification has usually been tackled following two main approaches. The first one applies Machine Learning (ML) algorithms in order to train a polarity classifier using a labelled corpus [1]. This approach is also known as the supervised approach. The second one is known as Semantic Orientation (SO), or the unsupervised approach, and it integrates linguistic resources in a model in order to detect the polarity [2]. Both approaches have advantages and drawbacks. For example, ML methods require annotated corpora to train the model that are normally difficult to achieve. However, this strategy usually obtains better performances. On the contrary, the SO methodology requires a large amount of linguistic resources which generally depend on the language, although the model does not require labelled corpora for learning. Until now the results obtained with unsupervised models do not outperform the ML classifiers. However, there are several semantic resources that we believe must be analysed and integrated in order to improve these systems.

In this study we use one of these interesting semantic resources: SentiWordNet (SWN) [3]. Specifically, our proposal focuses on adapting this resource to the Spanish language in order to be applied directly over a Spanish movie review corpus. As a main novelty we make use of the Multilingual Central Repository (MCR) [4] [5] by linking each synset of SWN to their equivalent Spanish semantic words. The MCR integrates wordnets from five different languages (English, Spanish, Catalan, Basque and Galician), allowing connections from words in one language to equivalent translations in any of the other languages thanks to the automatically generated mappings among WordNet versions. To our knowledge this is the first time that MCR has been integrated with SWN in order to classify the opinion polarity in a Spanish review corpus.

According to [6] there are two main approaches in the context of multilingual SA: The first one is the corpus-based approach, where a subjectivity-annotated corpus for the target language is built through projection, and then a statistical classifier is trained on the resulting corpus (Figure 1). The second one is the lexicon-based approach, where a target-language subjectivity classifier is generated by translating an existing lexicon into another language (Figure 2).

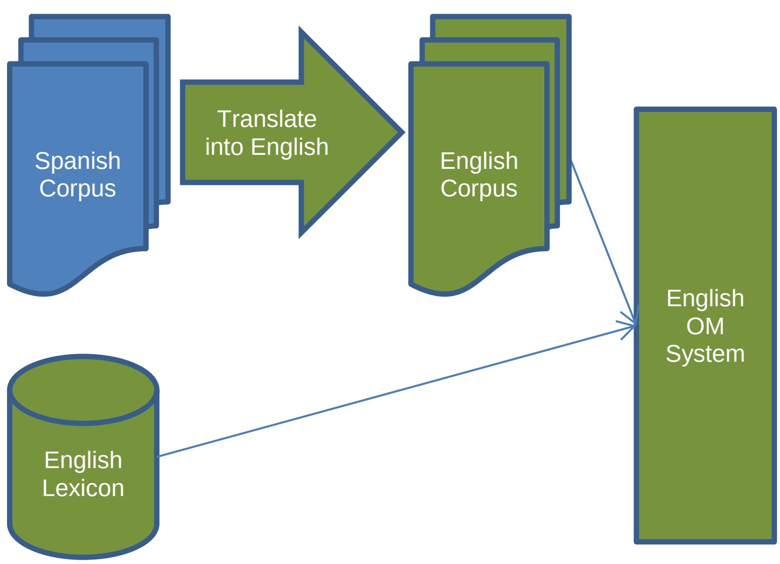

Figure 1. Corpus-based approach

In this paper we combine both approaches, using meta-classifiers in order to improve the final system. For the corpus-based approach, we translate a Spanish corpus of movie reviews called MuchoCine (MC) into English and then we apply different English resources (SWN and opinionated lists of words). For the lexicon-based approach, we use the MC corpus directly in Spanish. Therefore we have used two different semantic resources. First, we use a list of opinionated words translated into Spanish, and secondly we apply the MCR in Spanish linked with SWN in order to integrate a Spanish lexicon over the MC corpus.

Finally, we propose to take advantage of the combination of different linguistic resources and the certainty that subjectivity tends to be preserved between languages. Several combinations of classifiers were studied with the goal of improving the performance of the Spanish polarity classification. The results show that the combination of different linguistic resources and also the use of meta-classifiers enhance the performance of a polarity classification system for Spanish texts.

The rest of the article is organized as follows: The next section presents related studies that apply a semantic orientation approach focusing mainly on the use of SWN and other lexical resources. We also comment on some research which deals with languages other than English and multilingual OM. In addition, some interesting papers about meta-classifiers are also referred. Section 3 outlines the method proposed. Section 4 introduces the main resources used in our experiments. Then the individual and combined systems are described and the results obtained analysed. Finally, the conclusions and future work are presented.

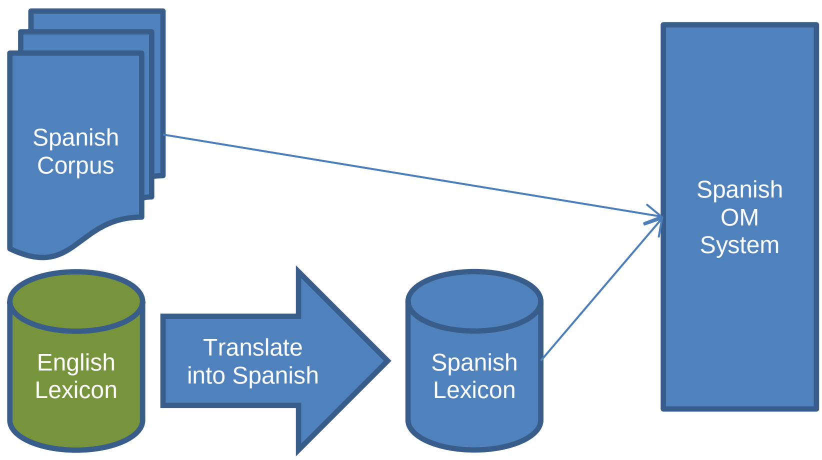

Figure 2. Lexicon-based approach

2. Related work

Research into opinion mining has experimented an exponential growing in recent years. Some recent surveys can be found in [7] and [8]. Concerning Semantic Orientation (SO) there are also several studies. In the SO approach the document is represented as a collection of words and manual rules and lexicons are applied. The sentiment of each word can be determined by different methods, for example using a list of opinionated words [9], applying web search [10], making use of annotated terms in dictionaries [11], or lexical resources such as General Inquirer [12] and WordNet [13]. Moreover, there are other studies that apply specific sentiment analysis resources like SentiSense [14], WordNet-Affect [15] or SentiWordNet [16].

2.1. Non-English Sentiment Analysis

Regarding opinion mining focused on languages other than English some studies can be highlighted. For example, Zhang et al. [17] applied Chinese sentiment analysis 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). Agić et al. [18] presented a manually annotated corpus with news on the financial market in Croatia. In [19] a corpus of movie 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 et al. [20] 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. Brooke et al. [21] 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 [11]. Cruz et al. [22] manually recollected the MuchoCine (MC) corpus to develop a sentiment polarity classifier based on semantic orientation. The corpus contains annotated Spanish movie reviews from the MuchoCine website. This MC corpus is used in this paper for our experimental study.

On the other hand, although SentiWordNet has been used in several studies most of them only deal with English documents. A few studies try to apply SWN to languages other than English. For example, Denecke [23] 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: LingPipe, SentiWordNet with classification rule, and SWN with machine learning. Ghorbel and Jacot [24] used a corpus with movie reviews in French. They applied a supervised classification combined with SentiWordNet in order to determine the polarity of the reviews. Martín-Valdivia et al. [25] presented an experimental study of supervised and unsupervised approaches over a Spanish-English parallel corpus, by integrating SWN in different ways over the translated English corpus. Perea-Ortega et al. [26] carried out several experiments by combining both machine learning and semantic orientation approaches over the Opinion Corpus for Arabic (OCA) [19] and its parallel English version named EVOCA. They applied a voting system based on majority rule showing a slight improvement when both approaches were combined. In all of these examples the original opinion corpus is translated into English and then SWN is applied over the translated English text.

In this paper we study the use of SWN directly over the original corpus. In order to apply SWN over a non-English corpus it is necessary to use another resource to link the synset in English to its corresponding synset in Spanish. For this case we have integrated the Multilingual Central Repository. MCR has been applied in several studies, but for sentiment analysis we can only find one [27]. The idea emerges after several studies related to Spanish polarity detection over the MC corpus. In our first paper [25] we followed the corpus-based approach and we generated an English parallel corpus, called MCE (MuchoCine English version)[2]. The MCE corpus was built by applying automatic machine translation techniques to the Spanish MC corpus. Then we combined supervised and unsupervised approaches using meta-classifiers. First we generated two individual models using these two corpora (MC Spanish and English corpus) and applying Support Vector Machines (SVM) algorithms. Then we integrated SentiWordNet into the English corpus, generating a new unsupervised model. Finally, the three systems were combined using a meta-classifier that allows us to apply several combination algorithms such as voting system or Stacking [28].

On the other hand, our second study [29] was oriented to a lexicon-based approach dealing with the Spanish MC corpus and using the semantic orientation strategy. The paper presented a new resource for the Spanish sentiment analysis research community (iSOL, improved Spanish Opinion Lexicon). We generated the new lexicon iSOL by translating into Spanish the Bing Liu English Lexicon (BLEL) [9], and then the resource was manually revised and improved.

2.2. Using Meta-classifiers for Sentiment Analysis

As mentioned before, there are several methods for generating classifiers for polarity detection. Each one has advantages and drawbacks. Thus some researchers have tried to take advantage of the ensemble methods or meta-classifiers theory. The main idea of ensemble methodology is to combine a set of classifiers in order to obtain a composite of combine learners, with more accurate estimations that can be achieved by using a single classifier [30]. Broadly speaking, the ensemble methodology tries to learn from the errors of the base classifiers with the aim of achieving a more accurate final classifier. A wide range of methodologies of combining classifiers are described in the literature due to their potential usefulness. Several factors differentiate the various ensemble or combined methods. The main factors are:

(1) Inter-classifier relationship: Depending on whether each classifier is affected by the other ones the ensemble methods can be divided into two types: sequential and concurrent. The sequential ensemble methods are those where the final model is built in an iterative process of model generation, in which the model of *i*-th iteration depends on the previous model. An example of a sequential combined classifier is AdaBoost [31]. On the other hand, the concurrent ensemble methods are those where dependency between the models that concern the

JIS Accepted for Publication
By the Journal of Information Science: http://jis.sagepub.co.uk

classification process is minimal, and they are built concurrently. The most representative concurrent ensemble method is Stacking [28].

(2) The combined method. Each ensemble method has to choose between several ways to combine the output of different classifiers. Voting schemes, Bayesian combinations, distribution summation, likelihood combination or statistical methods can be used.
(3) The diversity generator. Some ensemble methods require that the classifiers concerned in the process generate a diversity output. Loosely speaking, the combined classification will be more successful when the outputs of the classifiers are more different.
(4) Ensemble size. The number of classifiers involve in the classification process is another important factor.

In [32] the authors carried out a broad experimentation. The authors built two systems; one employing a lexicon-based method and the other one based on the use of the machine learning algorithm Support Vector Machines (SVM). Although the machine learning system achieved good results, the authors wanted to enhance the overall performance. They tried to combine the two systems. Firstly the authors developed a weighted voting method, and then they built a meta-set with the output of the two base learners. The meta-classifier based on SVM outperformed the weighted voting scheme and also all the variations of the two base learners. They concluded that the combination of several classification models helps to enhance the results of a polarity classification system.

Wan [33] worked with a corpus of Chinese product reviews. The author designed a framework focused on the combination of two unsupervised classifiers. One of them was used for classifying the original Chinese products reviews and the other one for its English translated version. The two base learners consisted of counting the number of positive and negative terms. For the combination, the author assessed three ensemble methods: average, weighted average and voting scheme. As in the previous study the results demonstrated that the combination of sentiment classification models enhance the final performance of the system.

In [6] it is asserted that subjectivity tends to be preserved across languages, but in [20] it is hypothesized that subjectivity is expressed differently in various languages due to lexicalization, formal versus informal markers, etc. Thus, in [34] the authors tried to demonstrate that several parallel corpora in different languages can complement each other in polarity classification. The authors took the MPQA corpus [35] and translated it into Spanish, Arabic, French, German and Romanian. Then several individual polarity classification experiments were carried out using Naïve Bayes, and they also combined the individual classifiers with a majority vote meta-classifier. The authors concluded that more languages are better for multilingual sentiment classification as they are able to complement each other, and together they provide better classification results.

Balahur and Tuchi [36] studied the manner in which Sentiment Analysis can be performed for languages other than English using machine translation. The authors studied the issue in three different languages (Spanish, French and German), taking an English sentiment corpus as the original source. As baseline they classified the English corpus with the SVM SMO machine learning algorithm and then the concurrent ensemble method Bagging [37] was applied in order to study the way in which noise in the training data could be removed. Finally they translated the English corpus into Spanish, French and German, and repeated the same experimentation. The results obtained were very similar, although the German classification with Bagging methodology outperformed the baseline. The authors concluded that the use of machine translation systems is a good strategy for tackling the problem of multilingual sentiment classification, mainly due to the noteworthy current performance of the machine translation systems. In addition, they mentioned that meta-classifiers can be used to reduce the noise that the translation process may include.

3. Proposed approach

Our latest studies have followed the hypothesis proposed by [6] which states that subjectivity tends to be preserved across languages. This affirmation is very valuable for research in those languages whose available sentiment linguistic resources are scarce. Spanish is one of these languages with limited resources for opinion mining. For this reason we are currently focusing on adapting English linguistic resources for polarity classification to Spanish, and on generating Spanish sentiment resources.

In our first study in this area [25] ensemble methods and meta-classifiers were explored with the aim of including English sentiment resources in a Spanish polarity classification system. The combined system enhanced the results obtained by the supervised system that only took into account the lexical information of the Spanish sentiment corpus, and encouraged us to continue researching into the application of ensemble methods in sentiment analysis. As a result, due to the lack of lexicons of opinion bearing words in Spanish, and following the lexicon-based approach proposed by [6], a new Spanish sentiment lexicon called iSOL was generated [29].

After our previous experiments we focused on using a greater number of linguistic resources in Spanish. Since SentiWordNet (SWN) is a widely used sentiment resource for opinion mining tasks, but is only available for English, we tried to apply some Spanish linguistic resource that would be able to link WordNet synsets (and therefore SentiWordNet words) with Spanish words. In this sense the Multilingual Central Repository (MCR) provides a reduced version of WordNet in Spanish and allows users to make connections between equivalent translations in different languages such as English and Spanish. A brief description of MCR is shown below in Section 4.3.

Therefore, focusing on an unsupervised approach, our proposal in this study is to combine different semantic resources available for different languages (English and Spanish) in order to address the polarity classification task with Spanish documents and try to improve the individual performance of these resources for the task. On the one hand, we combine the semantic resources (iSOL and SWN_SP) for the original Spanish corpus, where SWN_SP is the adapted version of SentiWordNet for Spanish by using the MCR resource. On the other hand, we combine the semantic resources (BLEL and SWN) for the English translated version of the corpus. Finally, the results of both classifications are combined again by following a stacking strategy. Figures 3, 4 and 5 show the overview of our proposed approach.

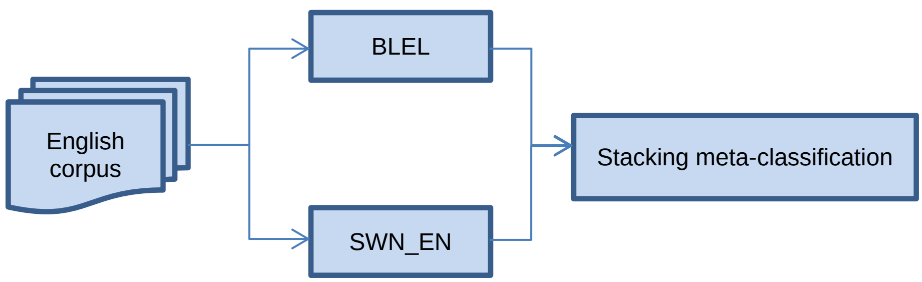

Figure 3. English Stacking Scheme

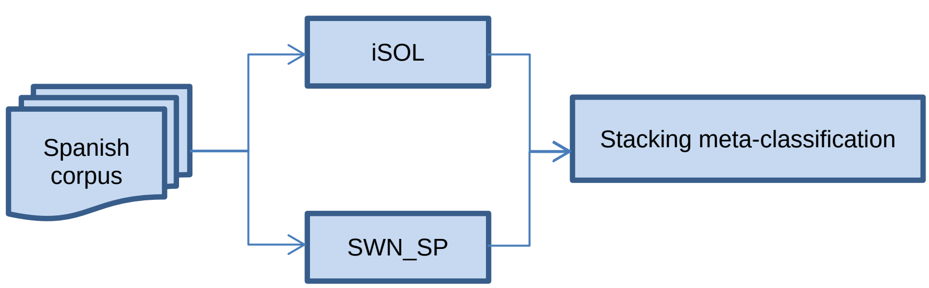

Figure 4. Spanish Stacking Scheme

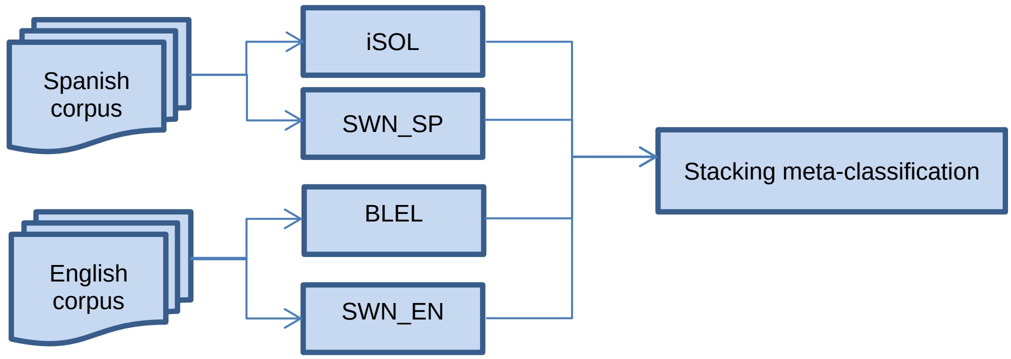

Figure 5. Bilingual Stacking Scheme

In order to carry out the stacking approach different machine learning algorithms have been applied: Support Vector Machines (SVM), Naïve Bayes (NB) and Bayesian Logistic Regression (BLR). SVM and NB are broadly known by the research community in Natural Language Processing, but BLR not so much. BLR [38] is a Bayesian implementation of the logistic regression that avoids over fitting the training data. The algorithm is based on the calculation of the following conditional likelihood:

$$P(y|\beta, x_i) = \omega(\beta^T, x_i) = \omega\left(\sum_i \beta_j x_{i,j}\right) \tag{1}$$

Where $y \in \{+1, -1\}$ are the classes. Each document is represented by a vector (xi) of values, βj are the predictors variables, and ω is a logistic link function.

$$\omega(r) = \frac{e^r}{1+e^r} \tag{2}$$

The use of a Bayesian approach to avoid over fitting involves a prior distribution on β specifying that each β_j is likely to be near 0. The prior distribution selected was a Gaussian distribution.

When the regression model is built an iterative optimization process starts. It begins by setting all variables to some initial value. It then sets the first variable to a value than minimizes the objective function holding all other variables constant. When all variables have been traversed, the algorithm begins again. Multiple passes are made over the variables until some convergence criterion is met. BLR has achieved good results in text classification problems [38] [39] and sentiment analysis [40].

4. Semantic resources

This section describes the semantic resources used for the experiments carried out. Firstly, two main lists of opinionated words were used: the improved Spanish Opinion Lexicon (iSOL) for the experiments using the Spanish corpus, and the Bing Liu English Lexicon (BLEL) for the English parallel corpus. Secondly, SentiWordNet was also applied for both corpora, but making use previously of the MCR resource when the Spanish corpus was processed.

4.1. Lists of opinionated words

The improved Spanish Opinion Lexicon[3] (iSOL) [29] was generated from the Bing Liu English Lexicon[4] (BLEL) [9] by automatically translating it into Spanish and obtaining the SOL (Spanish Opinion Lexicon) resource. Finally, the iSOL resource was obtained after carrying out a manual revision over SOL in order to improve the final list of opinion words. iSOL is composed of 2,509 positive and 5,626 negative words. Therefore this Spanish lexicon contains 8,135 opinion words.

On the other hand, the BLEL resource is composed of 2,006 positive and 4,783 negative words, resulting in a total of 6,789 opinion words. Both resources contain a higher proportion of adjectives, adverbs, nouns and verbs. Moreover, some misspelled words are included in both lists because they appear frequently in social media content.

The difference in the number of words in these two lexicons is due to the Spanish grammar. For instance, 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 1 shows some examples of possible translations of English adjectives in Spanish.

Table 1. Examples of possible translations of English adjectives in Spanish

English	Spanish
good	*bueno, buena, buenos, buenas*
famous	*famoso, famosa, famosos, famosas*
attractive	*guapo, guapa, guapos, guapas*
ugly	*feo, fea, feos, feas*
aching	*dolido, dolida, dolidos, dolidas*
bad	*malo,mala,malos,malas*

4.2. SentiWordNet

SentiWordNet [3] is a publicly available lexical resource for opinion mining which assigns three sentiment scores to each synset of WordNet[5]: positivity (how positive the word is), negativity (how negative the word is) and objectivity (how objective the word is). In other words, the sentiment scores of SentiWordNet mean the probability of a synset of being positive, negative and neutral. Each of the scores ranges from 0 to 1, and their sum equals 1. SentiWordNet scores have been semi-automatically computed based on the use of weakly supervised classification algorithms.

In SentiWordNet (SWN), each entry contains the pair Part Of Speech (POS) category and ID, which uniquely identifies a WordNet (3.0) synset, the PosScore and NegScore, which are the positivity and negativity scores assigned by SentiWordNet to the synset, and the terms with sense number belonging to the synset . The objectivity score can be calculated as 1 - (PosScore + NegScore).

Table 2 shows an excerpt of the subjectivity scores found in SWN for some synsets related to the words “good” and “bad”. There are 4 senses of the noun (POS ‘n’) “good”, 21 senses of the adjective (POS ‘a’) “good”, and 2 senses of the adverb (POS ‘r’) “good” in WordNet. There is one sense of the noun “bad”, 14 senses of the adjective “bad”, and 2 senses of the adverb “bad” in WordNet (3.0) synset.

Table 2. Fragment of SentiWordNet

POS	ID	PosScore	NegScore	SynsetTerms Gloss
a	00064787	0.625	0	good#5 beneficial#1
n	03076708	0	0	trade_good#1 good#4 commodity#1
r	00011093	0.375	0	well#1 good#1
a	01174222	0	1	unsound#5 unfit#3 bad#10
n	05144079	0	0.875	badness#1 bad#1
r	00016240	0.125	0.25	badly#6 bad#2

4.3. The Multilingual Central Repository

The Multilingual Central Repository[6] (MCR) [4] [5] constitutes a large-scale natural multilingual linguistic resource that can be used for semantic processes that need large amounts of linguistic knowledge. The MCR integrates into the same EuroWordNet framework wordnets from five different languages (including Spanish) together with four English WordNet versions. The final version of the MCR contains 1,642,389 semantic relations between synsets, most of them acquired by automatic means.

Describing more deeply how MCR works, it can be seen as a sense inventory for nouns, verbs, adjectives and adverbs for the languages involved (Basque, Catalan, English, Italian and Spanish). The wordnets in MCR have been constructed following the model proposed by EuroWordNet[7], i.e. the wordnets are linked to an Inter-Lingual-Index (ILI). Via this index the languages are interconnected, making it possible to go from the words in one language to similar words in any other language connected. The ILI is a set of meanings, mainly taken from WordNet[8]. The only purpose of the ILI is to mediate between the synsets of the local wordnets. Each synset in the local wordnets has at least one equivalent relation with a record in this ILI, either directly or indirectly via other related synsets. Language-specific synsets linked to the same ILI-record should thus be equivalent across the languages.

For the experiments carried out in this study we have used the ILI version that corresponds to WordNet 3.0. This resource allows us to obtain the synset id for each Spanish word and then look for the positive and negative scores in SWN.

5. Experimental Framework

5.1. Corpora

In this section, the two corpora used for the experiments carried out in this study are described. Firstly, the main features of the MuchoCine (MC) corpus are described. This corpus is composed of film reviews in Spanish. Then we explain briefly how the parallel English version of MC (MCE) was generated by applying machine translation techniques.

5.1.1. The MC corpus

MuchoCine [22] is a corpus of movie reviews in Spanish available for the SA research community[9]. The corpus consists of 3,878 movie reviews collected from the MuchoCine website[10]. 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 two million words and an average of 546 words per review.

In the MC corpus a movie review consists of four fields: the identifier of the review, the review rating, the summary and the description. Figure 6 shows an excerpt of a review from MC.

id|rank|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…*

Figure 6. Excerpt of a review from the MuchoCine corpus

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 neither bad nor good. Table 3 shows the number of reviews per rating. This corpus has been widely used in different studies such as [41], [42], [43], [40] and [25].

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 negatives and those with ratings of 4 or 5 were considered as positives. Table 4 shows the class distribution of the binary classification of MC.

Table 3. Rating distribution

Rating	#Reviews
1	351
2	923
3	1,253
4	890
5	461
Total	3,875

Table 4. Binary classification of the MC corpus

Classes	#Reviews
Positive	1,274
Negative	1,351
Total	2,625

5.1.2. The MCE corpus

The MuchoCine English corpus (MCE) [25] is the version of MC translated into English and it is also available for the research community[11]. It was generated by applying a machine translation process in which different automatic translation tools were tested. According to the authors some difficulties were encountered during this process, but finally the Microsoft Translator Java API was selected as the automatic translation tool. As for the MC corpus, in MCE a movie review consists of four fields: the identifier of the review, the review rating, the summary and the description. Figure 7[12] shows an excerpt of a review from MCE.

id|rank|summary|body
1000|-1|*Silicone, steroids, demonic pacts and other usual garbage are the basis underpinning this aberration.*| *A party filled with excesses, stunning blondes, musculitos everywhere, some other dead. Nothing new. The Alliance of evil is the name of this supernatural thriller which tells the adventures of a few young people ………*

Figure 7. Excerpt of a review from the MuchoCine English corpus

For the MCE corpus we followed the same criteria that we used for the MC corpus, i.e. we discarded the neutral examples. In this way opinions rated with 3 are not considered, the opinions with ratings of 1 or 2 are considered as negative and those with ratings of 4 or 5 are considered as positive.

5.2. Evaluation framework

In order to assess the proposal, a *k-fold* cross-validation process was carried out. *K-fold* cross-validation consists of dividing the dataset in *k* bins or folders. The algorithm is run *k* times with *k* different training and test sets. In each

JIS Accepted for Publication
By the Journal of Information Science: http://jis.sagepub.co.uk

iteration, *k-1* bins are considered to build the training set, and the other one is employed to test the classification model. For the experiments of this study we applied *10-fold* cross-validation (*k=10*).

For evaluating the classification accuracy, we employed the traditional measures used in text classification tasks: precision (P), recall (R), F1 and accuracy (Acc). For a feasible comparison, we summarize the F1 scores over the different categories (positive and negative) using the macro-averages of F1 scores:

$$Macro-F1 = \frac{2*Macro-Precision*Macro-Recall}{Macro-Precision+Macro-Recall} \quad (7)$$

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

$$Macro-Recall = \frac{\sum_{i=1}^{c} r_i}{c} \quad (8)$$

$$Macro-Precision = \frac{\sum_{i=1}^{c} p_i}{c} \quad (9)$$

Where r is the recall value, p is the precision value, and c is the number of classes.

6. Experiments and results

This section describes the experiments carried out and shows the results obtained. Firstly, we present the individual experiments that only make use of the two semantic resources explained in Section 4. Secondly, the experiments related to the proposed approach by combining the individual classifiers are described.

6.1. Individual experiments

6.1.1. Using lists of opinionated words

Before carrying out the experiments we performed a pre-processing step to the MC corpus in order to apply the same criteria followed during the generation of the iSOL list. For example, for both summary and body fields we had to change capital letters for non-capital letters, accented letters for non-accented letters, and all special characters had to be deleted from the opinions. Moreover, stop words and proper nouns were discarded.

For the MCE corpus we performed a simpler pre-processing step. We only had to change capital letters for non-capital letters and commas, semicolons, question marks and periods characters were deleted from both the summary and body fields of each opinion.

In order to calculate the polarity (p) of a review (r), we take into account the total number of positive words (#positive) and the total number of negative words (#negative) within the review, according to the following strategy:

$$p(r) = 1 \leftrightarrow \#positive \geq \#negative \quad (10)$$

$$p(r) = -1 \leftrightarrow \#positive < \#negative \quad (11)$$

Table 5 shows the results obtained by using the two lists of opinionated words over the corpora.

Table 5. Results obtained by using the two lists of opinionated words

	Macro-P	Macro-R	Macro-F1	Acc
iSOL over MC	62.22%	61.47%	61.84%	61.83%
BLEL over MCE	61.92%	56.58%	59.13%	57.56%

6.1.2. Using SentiWordNet

The semantic orientation approach using SentiWordNet has been also applied to both corpora, MC in Spanish and MCE in English. For both corpora we have followed the same procedure:

(1) **Part Of Speech tagging (POS tagging)**. The documents were processed by applying a POS tagger like TreeTagger[13] [44]. The aim of this process was to obtain all the nouns, adjectives, verbs and adverbs of each review.

(2) **Linguistic feature extraction**. This process extracts linguistic features detected in the previous step in order to generate different sub-corpora. A total of 15 sub-corpora from MC and MCE were provided by making a combination among the four linguistic features (nouns, adjectives, verbs and adverbs): *only-noun, only-adj, only-verb, only-adv, adj+noun, adj+verb, adj+adv, noun+verb, noun+adv, verb+adv, adj+noun+verb, adj+noun+adv, noun+verb+adv, adj+verb+adv, adj+noun+verb+adv*.

(3) **SWN score calculation**. The SWN score for each document was calculated in order to classify them as positive or negative. The SWN score or polarity score of a document was obtained by following the procedure described by Denecke [23] based on the calculation of a triplet of positivity, negativity and objectivity scores:

(3.1) For each token *A* with *n* synsets found in SWN, we calculate the average of its positivity score ($score_{pos}$) and the average of its negativity score ($score_{neg}$):

$$score_{pos}(A) = \frac{1}{n}\sum_{i=1}^{n} score_{pos}(i) \quad (12)$$

$$score_{neg}(A) = \frac{1}{n}\sum_{i=1}^{n} score_{neg}(i) \quad (13)$$

(3.2) The objectivity score ($score_{obj}$) is obtained for each token:

$$score_{obj}(A) = 1 - (score_{pos}(A) + score_{neg}(A)) \quad (14)$$

(3.3) The score-triplet for each document is determined by summing up the score-triplet of each term and dividing each score by the number of terms considered in such document.

In order to classify a review as "positive" or "negative" using SWN, we followed a similar strategy to that applied for the lists of opinionated words, i.e., we considered a review as "positive" if its positivity score ($score_{pos}$) is larger than or equal to the negativity score ($score_{neg}$) and as "negative" otherwise.

As described above, SentiWordNet is a semantic lexical resource available only in English. In order to deploy our SWN approach for the Spanish corpus (MC), we made use of the Multilingual Central Repository (see Section 4.3). The results obtained by using SWN over the corpora are shown in Table 6 for the MC corpus and Table 7 for the MCE corpus.

From the results shown in Table 6 and Table 7 we can observe that the highest accuracy results for both MC and MCE are obtained from those sub-corpora that include adjectives as linguistic features, as expected. Nevertheless, it is noteworthy the difference obtained (+8.87%) between the best results of both corpora (*adj+nouns+verb+adv* from MCE versus *adj+adv* from MC). We think that the main reason for this behavior is the significant difference between the number of synsets included in SentiWordNet (around 117,000 synsets) and those covered by the Multilingual Central Repository (around 38,000 synsets) used for applying SentiWordNet to the Spanish corpus.

Table 6. Results obtained by using SentiWordNet over the MC corpus (SWN_SP)

	Macro-P	Macro-R	Macro-F1	Acc
Only-nouns	51.27%	51.06%	51.16%	51.66%
Only-verb	51.48%	50.51%	50.99%	51.70%
Only-adj	62.06%	58.74%	60.36%	59.50%
Only-adv	55.20%	54.11%	54.65%	54.78%
Nouns+adj	60.41%	56.60%	58.45%	57.49%
Nouns+verb	48.56%	49.37%	48.96%	50.48%

Nouns+adv	53.87%	52.91%	53.39%	53.64%
Adj+verb	59.46%	54.18%	56.70%	55.28%
Adj+adv	**63.68%**	**58.79%**	**61.14%**	**59.66%**
Verb+adv	55.43%	51.79%	53.55%	52.99%
Nouns+adj+verb	58.93%	53.87%	56.29%	54.97%
Nouns+adj+adv	61.81%	56.93%	59.27%	57.87%
Nouns+verb+adv	50.13%	50.05%	50.09%	51.20%
Adj+verb+adv	61.16%	54.41%	57.59%	55.54%
Adj+nouns+verb+adv	59.19%	53.47%	56.19%	54.63%

Table 7. Results obtained by using SentiWordNet over the MCE corpus (SWN_EN)

	Macro-P	Macro-R	Macro-F1	Acc
Only-nouns	55.58%	54.31%	54.94%	55.01%
Only-verb	57.59%	54.95%	56.24%	55.81%
Only-adj	61.69%	60.78%	61.23%	61.18%
Only-adv	58.26%	55.09%	56.63%	54.17%
Nouns+adj	62.90%	60.48%	61.66%	61.10%
Nouns+verb	56.11%	53.76%	54.91%	54.67%
Nouns+adv	58.79%	58.41%	58.60%	58.10%
Adj+verb	63.42%	60.85%	62.11%	61.49%
Adj+adv	63.20%	62.82%	63.01%	62.55%
Verb+adv	58.14%	57.13%	57.63%	56.61%
Nouns+adj+verb	62.45%	58.96%	60.66%	59.73%
Nouns+adj+adv	64.27%	64.25%	64.26%	64.30%
Nouns+verb+adv	60.39%	60.39%	60.39%	60.34%
Adj+verb+adv	64.32%	64.33%	64.33%	64.30%
Adj+nouns+verb+adv	**65.13%**	**64.72%**	**64.92%**	**64.95%**

6.2. Combined experiments: stacking

We applied a stacking strategy in order to improve the results obtained with the individual experiments. The main purpose of the stacking method is to achieve the highest generalization accuracy by creating a meta-dataset which contains one tuple for each input example. The dimensions or number of features of those tuples are the outputs of the individual classifiers. Loosely speaking, the stacking technique consists of building a new classifier whose inputs are the outputs of the individual classifiers.

First we carried out the combination of the classifiers for each language. Three different machine learning algorithms were evaluated as stacking classifiers: Support Vector Machines (SVM), Naïve Bayes (NB) and Bayesian Logistic Regression (BLR).

As explained in the previous section, the formula used to calculate the polarity of a document for the SentiWordNet experiments was the one proposed by Denecke. This formula returns three scores (positivity, negativity and objectivity) for each document. The predicted class or output class is calculated taking into account those polarity scores. Thus the system based on the use of SentiWordNet returns four values: the predicted class and three polarity scores (positivity, negativity and objectivity). Taking into account these values, six meta-datasets for each language were built:

(1) MC_iSOL_SWN_SP_pred: It is composed of the predicted class of the base learners (SWN_SP and iSOL).
(2) MCE_BLEL_SWN_EN_pred: It is composed of the predicted class of the base learners (SWN_EN and BLEL).
(3) MC_iSOL_SWN_SP_polar: The dataset includes the predicted class of the classifier based on iSOL and the three polarity scores returned by the classifier based on SWN_SP.

(4) MCE_BLEL_SWN_EN_polar: The dataset includes the predicted class of the classifier based on BLEL and the three polarity scores returned by the classifier based on SWN_EN.
(5) MC_iSOL_SWN_SP_pred_polar: It is formed by the predicted classes returned by the two classifiers (iSOL and SWN_SP) and the polarity scores of SWN_SP.
(6) MCE_BLEL_SWN_EN_pred_polar: It is formed by the predicted classes returned by the two classifiers (BLEL and SWN_EN) and the polarity scores of SWN_EN.

The results achieved with these combinations of monolingual learners are shown in Table 8.

Table 8. Results obtained for the monolingual stacking experiments

Meta-dataset	Stacking learner	Macro-P	Macro-R	Macro-F1	Accuracy
	SVM	62.26%	61.47%	61.86%	61.83%
MC_iSOL_SWN_SP_pred	NB	63.94%	63.67%	63.80%	63.85%
	BLR	63.94%	69.67%	63.80%	63.85%
	SVM	62.26%	61.47%	61.86%	61.83%
MC_iSOL_SWN_SP_polar	NB	63.33%	62.64%	62.93%	62.93%
	BLR	62.75%	62.08%	62.41%	62.40%
	SVM	62.26%	61.47%	61.86%	61.83%
MC_iSOL_SWN_SP_pred_polar	NB	63.84%	61.90%	62.86%	62.44%
	BLR	63.77%	63.46%	63.61%	63.65%
	SVM	63.60%	60.85%	62.19%	61.48%
MCE_BLEL_SWN_EN_pred	NB	63.68%	62.26%	62.96%	62.70%
	BLR	63.68%	62.26%	62.96%	62.70%
	SVM	61.96%	60.07%	61.03%	60.65%
MCE_BLEL_SWN_EN_polar	NB	60.02%	60.07%	61.03%	60.65%
	BLR	62.76%	57.80%	60.18%	58.70%
	SVM	63.60%	60.85%	62.19%	61.48%
MCE_BLEL_SWN_EN_pred_polar	NB	63.45%	62.08%	62.76%	62.51%
	BLR	63.68%	62.26%	62.96%	62.70%

The results achieved with the two corpora follow the same pattern. The meta-datasets composed by the predicted classes of the base learners are those that achieved higher results. The performance of the system is usually worse when the polarity scores of the base learners are included in the meta-dataset. According to the results shown in Table 8, it is evident that the polarity scores do not offer valuable information for the stacking classification with Spanish texts (MC corpus). The same behaviour is repeated when the corpus is MCE. Therefore, we can conclude that the most suitable combination to enhance the performance of the monolingual polarity classification systems is the one that only takes into account the predicted classes of the base learners.

Taking into account that several classifiers have been developed for monolingual polarity classification and with the aim of improving the Spanish polarity classification, we studied the combination of the Spanish and English polarity classifiers. The meta-datasets built were:

(1) MC_iSOL_SWN_SP_pred_MCE_BLEL: Combination of the outputs of iSOL, SWN_SP and BLEL.
(2) MC_iSOL_SWN_SP_polar_MCE_BLEL: Combination of the output of iSOL, the probabilities calculated by SWN_SP, and the output of BLEL.
(3) MC_iSOL_SWN_SP_pred_polar_MCE_BLEL: Combination of the output of iSOL, the class and the probabilities calculated by SWN_SP and the output of BLEL.
(4) MC_iSOL_SWN_SP_pred_MCE_SWN_EN_pred: Combination of the outputs of iSOL, the class calculated by SWN_SP and the classes returned by SWN_EN.
(5) MC_iSOL_SWN_SP_polar_MCE_SWN_EN_polar: Combination of the output of iSOL, the probabilities calculated by SWN_SP, and the likelihoods returned by SWN_EN.

(6) MC_iSOL_SWN_SP_pred_polar_MCE_SWN_EN_pred_polar: Combination of the output of iSOL, the class and the probabilities calculated by SWN_SP and the class and the likelihoods returned by SWN_EN.
(7) MC_MCE_pred: The combination iSOL, BLEL, and the classes returned by SWN_SP and SWN_EN.
(8) MC_MCE_polar: The combination iSOL, BLEL, and the probabilities returned by SWN_SP and SWN_EN.
(9) MC_MCE_pred_polar: The combination iSOL, BLEL, and the classes and probabilities returned by SWN_SP and SWN_EN.

The results of the experiments above are shown in Table 9. This table shows different behaviours which depend on the combined classifiers and the stacking algorithm used. When the BLEL lexicon is used the results are very similar, but when the BLR is the stacking classifier the results usually increase slightly. Therefore, at least in this case, the BRL algorithm learns from the diversity of the base classifiers with better performance than SVM and NB. When SentiWordNet in English is used, the results are higher than those obtained when BLEL is combined to the Spanish base learners. Focusing only on the combinations between iSOL, SWN_SP and SWN_EN, it is interesting to highlight the fact that NB always enhances the results when the polar scores are added to the input of the stacking classifier. In the case of the monolingual stacking experiments, the polarity scores usually worsen the overall performance of the system. In this group of experiments, the polarity scores are the features which provided more information. The meta-dataset composed by the predicted class of the system based on iSOL, the predicted classes and the polarity scores of the classifiers based on the use of SWN_EN and SWN_SP is the one that achieves the highest results. NB is the algorithm with the best performance with that meta-dataset. At the beginning we expected that when more information was combined the results would be better, but as Table 9 shows, we were wrong. In this case, one of the classifiers introduces noise in the meta-classification, and taking into account the previous results we can conclude that BLEL does not provide valuable information to the classification process.

7. Analysis of Results

The main goal of this article is the improvement of Spanish polarity classification. To reach that purpose we propose a method which consists of the combination of two sentiment resources and the use of meta-classifiers. Our hypothesis has two main keys:

- The integration of semantic resources always helps the process of polarity classification.
- As Mihalcea et al. (2007) propose in their work, we also think that subjectivity tends to be preserved across languages.

In the previous sections we have described a number of experiments that assess the method proposed. As baseline, the MC corpus was classified by a system based on the use of a bag of opinion bearing words (iSOL). Then the Spanish corpus was also classified using a Spanish projection of the sentiment base-knowledge SentiWordNet. The next step was the evaluation of whether a meta-classifier method like staking can enhance the results. Following the second hypothesis the Spanish corpus was translated into English, with the aim of taking advantage of some English semantic resources. The same systems were built for English texts but using English semantic resources. The last step to evaluate our hypothesis was the combination of all the systems developed. The results obtained for Spanish sentiment classification are summed up in Table 10.

The results shown in Table 10 demonstrate that our hypothesis is correct. Firstly, the individual systems (1) and (2) achieved lower performance than the combined systems (3) and (4). Also, we highlight the fact that the system which uses iSOL obtained better results than that based on the use of the Spanish projection of SentiWordNet using MCR. Specifically, individual Spanish experiments using the lexicon based-approach and the corpus based-approach achieved an accuracy of 61.83% and 59.66% respectively (Table 5 and Table 6). However, if we compare the individual experiments for the English corpus MCE (Table 5 and Table 7), the use of the semantic resource SWN obtains better results than those applying the list of opinionated word BLEL (64.95 % and 57.56% of accuracy, respectively). We think that the main reason for this behavior is the low number of synsets managed in the MCR compared to those covered by the original SWN resource for English.

Regarding the combined systems, we can affirm that our hypothesis is valid because when we increase the number of semantic resources combined the overall performance is higher. System (4) achieves better results than (3) because system (4) takes advantage of Spanish and English semantic resources, while system (3) only uses Spanish semantic resources. On the other hand, we consider that subjectivity tends to be preserved across languages because the systems in Spanish and English obtain very similar results, and we think the loss of accuracy is only due to the reasonable noise included in the translation process.

Table 9. Bilingual Stacking results

Meta-dataset	Stacking learner	Macro-P	Macro-R	Macro-F1	Accuracy
MC_iSOL_SWN_SP_pred_MCE_BLEL	SVM	62.26%	61.47%	61.86%	61.83%
	NB	63.55%	61.06%	62.28%	61.68%
	BLR	63.94%	63.67%	63.80%	63.85%
MC_iSOL_SWN_SP_polar_MCE_BLEL	SVM	62.26%	61.47%	61.86%	61.83%
	NB	63.53%	62.08%	62.80%	62.55%
	BLR	63.09%	62.43%	62.76%	62.74%
MC_iSOL_SWN_SP_pred_polar_MCE_BLEL	SVM	62.26%	61.47%	61.86%	61.83%
	NB	63.99%	62.10%	63.03%	62.63%
	BLR	64.13%	63.74%	63.93%	63.96%
MC_iSOL_SWN_SP_pred_MCE_SWN_EN_pred	SVM	64.40%	63.98%	64.19%	64.07%
	NB	64.89%	63.63%	64.25%	64.04%
	BLR	64.70%	64.50%	64.60%	64.53%
MC_iSOL_SWN_SP_polar_MCE_SWN_EN_polar	SVM	62.26%	61.47%	61.86%	61.83%
	NB	64.30%	63.89%	64.09%	64.11%
	BLR	63.27%	62.63%	62.95%	62.93%
MC_iSOL_SWN_SP_pred_polar_**MCE_SWN_EN**_pred_polar	SVM	63.93%	62.88%	63.40%	63.23%
	NB	**65.25%**	**64.34%**	**64.79%**	**64.68%**
	BLR	64.12%	63.70%	63.91%	63.92%
MC_MCE_pred	SVM	63.55%	62.70%	63.12%	63.01%
	NB	64.97%	63.35%	64.15%	63.81%
	BLR	64.57%	63.42%	63.99%	63.77%
MC_MCE_polar	SVM	62.26%	61.47%	61.86%	61.83%
	NB	64.55%	63.56%	64.05%	63.92%
	BLR	63.47%	62.76%	63.11%	63.08%
MC_MCE_pred_polar	SVM	63.37%	62.55%	62.96%	62.89%
	NB	65.12%	64.34%	64.73%	64.65%
	BLR	64.39%	63.70%	64.04%	64.00%

Table 10. Summary of Spanish polarity classification results

	Macro-F1	Accuracy
(1) iSOL over MC	61.84%	61.83%
(2) SWN_SP (adj+verb)	61.14%	59.66%
(3) MC_iSOL_SWN_SP_pred (NB)	63.80%	63.85%
(4) MC_iSOL_SWN_SP_pred_polar_MCE_SWN_EN_pred_polar (NB)	**64.79%**	**64.68%**

8. Conclusions and future work

In this paper we have combined the corpus-based and lexicon-based approch using meta-classifiers in order to improve the final system. For the corpus-based approach, we translate a Spanish corpus of movie reviews called MuchoCine (MC) into English and then we apply different English resources (SWN and opinionated lists of words). For the lexicon-based approach, we use MC corpus directly in Spanish. Therefore we have used two different semantic resources. First, we use a list of opinionated words translated into Spanish, and secondly we apply the MCR in Spanish linked with SWN in order to integrate a Spanish lexicon over the MC corpus. Finally, several combinations of classifiers were studied with the goal of improving the performance of the Spanish polarity classification. The results show that the combination of different linguistic resources, and also the use of meta-classifiers enhance the performance of a polarity classification system for Spanish texts. These results encourage us to continue working along this line.

On the other hand, for sentiment analysis the study of the influence of contextual valence shifters is very interesting. In English there are several publications such as [45] that study the influence of this linguistic element for the polarity classification task. The Spanish sentiment analysis research community has studied the contextual valence shifters in Spanish, but not in great depth. So currently we are carrying out a study of these elements, because we think that it is essential to the polarity classification task. Another of our future steps in the improvement of polarity classification systems in Spanish is the study of the calculation of negation scope and the use of linguistic heuristics to calculate the sentiment orientation of a sentence.

Notes

1. http://www.internetworldstats.com
2. MCE is freely available in http://sinai.ujaen.es/wiki/index.php/MCE_Corpus_(English_version)
3. http://sinai.ujaen.es/?p=1202
4. http://www.cs.uic.edu/˜liub/FBS/sentiment-analysis.html
5. WordNet is a large lexical database of English. Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. It is available in http://wordnet.princeton.edu
6. http://adimen.si.ehu.es/web/MCR
7. http://www.illc.uva.nl/EuroWordNet
8. http://wordnet.princeton.edu
9. http://www.lsi.us.es/~fermin/corpusCine.zip
10. http://www.muchocine.net
11. http://sinai.ujaen.es/?p=1208
12. Figure 6 is the English translation of Figure 7
13. http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger

Funding

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. The project AORESCU (TIC - 07684) from the regional government of Junta de Andalucía partially supports this manuscript. Also, this paper is partially funded by the European Commission under the Seventh (FP7 - 2007-2013) Framework Program 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

[1] Pang B and Lee L. Thumbs up?: Sentiment classification using machine learning techniques. In: Proceedings of the ACL-02 conference on Empirical methods in natural language processing. Stroudsburg, PA, USA: Association for Computational Linguistics 2002; 10:79-86.

[2] Turney P D. 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. Stroudsburg, PA, USA: Association for Computational Linguistics. 2002, pp. 417-424.

[3] Baccianella S, Esuli A and Sebastiani F. Sentiwordnet 3.0: An en-hanced lexical resource for sentiment analysis and opinion mining. In: Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10). Valleta, Malta: European Language Resources Association (ELRA) 2010.

[4] Atserias J, et al. The MEANING Multilingual Central Repository. In: Proceedings of the Second International Global WordNet Conference (GWC'04). Brno, Czech Republic 2004.

[5] Gonzalez-Agirre A, Laparra E and Rigau G. Multilingual Central Repository version 3.0. In: Proceedings of the 8th international conference on Language Resources and Evaluation (LREC'12), Istambul, Turkey, 2012.

[6] Mihalcea R, Banea C, and Wiebe J. Learning multilingual subjective language via cross-lingual projections. In: Proceedings of Annual meeting-association for computational linguistics 2007; 45(1): 976-983

[7] Liu B. Sentiment analysis and opinion mining. In: Synthesis Lectures on Human Language Technologies series 2012; 5(1): 1-167.

[8] Tsytsarau M and Palpanas T. Survey on mining subjective data on the web. Data Mining and Knowledge Discovery 2012; 24(3): 478-514.

[9] Hu M and Liu B. (2004). Mining and summarizing customer reviews. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Seattle, Washington, USA, 2004; pp. 168-177.

[10] Hatzivassiloglou V and Wiebe J. Effects of Adjective Orientation and Gradability on Sentence Subjectivity. In: Proceedings of the International Conference on Computational Linguistics 2000; pp. 299-305.

[11] Taboada M, Brooke J, Tofiloski M, Voll K, and Stede M. Lexicon-based methods for sentiment analysis. Computational Linguistics, 37(2) (2011) 267-307.

[12] Wilson T, Wiebe J and Hoffmann P. Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. In: Proceedings of the Human Language Technology Conference and the Conference on Empirical Methods in Natural Language Processing 2005, pp. 347-354.

[13] Kamps J, Marx M, Mokken R J and de Rijke M. Using WordNet to measure semantic orientation of adjectives. In: The fourth international conference on Language Resources and Evaluation 2004; pp. 1115-1118.

[14] De Albornoz J C, Plaza L, and Gervás, P. SentiSense: An easily scalable concept-based affective lexicon for sentiment analysis. In: Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12) 2012; pp. 3562-3567.

[15] Chaumartin F R. UPAR7: a knowledge-based system for headline sentiment tagging. In: Proceedings of the 4th International Workshop on Semantic Evaluations, SemEval'07, Association for Computational Linguistics 2007, pp. 422-425.

[16] Thet T T, Na, J. C, and. Khoo C. S. G. Aspect-based sentiment analysis of movie reviews on discussion boards. Journal of Information Science 2010; 36(6): 823-848.

[17] Zhang C, Zeng D, Li J, Wang F-Y and Zuo W. Sentiment analysis of Chinese docu-ments: From sentence to document level. Journal of the American Society for Information Science and Technology 2009; 60(12): 2474–2487.

[18] Agić Z, Ljubešić N and Tadić M. Towards Sentiment Analysis of Financial Texts in Croatian. In: Proceedings of Language Resources and Evaluation, European Language Resources Association 2010; pp. 1164-1167.

[19] Rushdi-Saleh M, Martín-Valdivia M T, Ureña-López L A and Perea-Ortega J M. OCA: Opinion corpus for Arabic. Journal of the American Society for Information Science and Technology 2011; 62: 2045-2054.

[20] Banea C, Mihalcea R, Wiebe J and Hassan, S. Multilingual subjectivity analysis using machine translation. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, Stroudsburg, PA, USA, Association for Computational Linguistics 2008; pp. 127-135.

[21] Brooke J, Tofloski M and Taboada M. Cross-linguistic sentiment analysis: From English to Spanish. In: Proceedings of the International Conference RANLP-2009 2009; pp. 50-54.

[22] Cruz F L, Troyano J A, Enriquez F and Ortega J. Experiments in sentiment classification of movie reviews in Spanish. Procesamiento del Lenguaje Natural 2008; 41:79-80.

[23] Denecke K. Using SentiWordNet for multilingual sentiment analysis. In: Proceedings of Data engineering Workshop, ICDEW, IEEE 24th International Conference on, 2008; pp. 507-512.

[24] Ghorbel H and Jacot D. Sentiment analysis of French movie reviews. Advances in Distributed Agent-Based Retrieval Tools 2011; 361: 97-108.

[25] Martín-Valdivia M T, Martínez-Cámara E, Perea-Ortega J M and Ureña-López L A. Sentiment polarity detection in Spanish reviews combining supervised and unsupervised approaches. Expert Systems with Applications 2013; 40(10), 3934–3942.

[26] Perea-Ortega J M, Martín-Valdivia M T, Ureña-López L A and Martínez-Cámara E. Improving polarity classification of bilingual parallel corpora combining machine learning and semantic orientation approaches. Journal of the American Society for Information Science and Technology 2013; 64(9):1864-1877.

[27] Martínez-Cámara E Martín-Validivia M T, Molina-González M. D and Ureña-López, L. A. Bilingual Experiments on an Opinion Comparable Corpus. In: Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, Atlanta, Georgia. ACL 2013; pp 87-93.

[28] Wolpert d H. Stacked Generalization. Neural Networks 1992; 5: 241-259.

[29] Molina-González M D, Martínez-Cámara E, Martín-Valdivia M T and Perea-Ortega J M. Semantic Orientation for Polarity Classification in Spanish Reviews. Expert Systems with Applications 2013; 40(18):7250-7257.

[30] Rokach L. Ensemble Methods for Classifiers. In: Malmon O and Rokach L (EDS) Data Mining and Knowledge Discovery Handbook. Springer US, 2005, pp. 957-980.

[31] Freund Y and Schapire R E. Experiments with a new boosting algorithm. In: Machine Learning: Proceedings of the Thirteenth International Conference, 1996, pp. 325-332.

[32] Kennedy A and Inkpen D. Sentiment classification of movie reviews using contextual valence shifters. Computational Intelligence 2006, 22(2): 110-125.

[33] Wan X. Using bilingual knowledge and ensemble techniques for unsupervised Chinese sentiment analysis. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2008, pp. 553-561.

[34] Banea C, Mihalcea R. and Wiebe J. Multilingual subjectivity: are more languages better? In: Proceedings of the 23rd International Conference on Computational Linguistics, 2010, pp. 28-36.

[35] Wiebe J, Wilson T and Cardie C. Annotating expressions of opinions and emotions in language. Language Resources and evaluation 2005; 39(2-3):165-210.

[36] Balahur A and Turchi M. Comparative experiments using supervised learning and machine translation for multilingual sentiment analysis. Computer Speech & Language 2014; 28(1):56-75.

[37] Breiman L. Bagging predictors. Machine learning 1996; 24(2):123-140.

[38] Genkin H, Lewis D and Madigan D. Large-Scale Bayesian Logistic Regression for Text Categorization. Technometrics 2004. 49(3):291-384.

[39] Ortiz-Martos A, Martín-Valdivia M T, Ureña-López L A and Cumbreras-García, M. A. Detección automática de Spam utilizando Regresión Logística Bayesiana. Procesamiento del Lenguaje Natural 2005; 35:127-133.

[40] Martínez-Cámara E, Martín-Valdivia M T, Ureña López L A. Opinion classification techniques applied to a Spanish corpus. In: Proceedings of the 16th international conference on Natural language processing and information systems, 2011, pp. 169-176.

[41] Del-Hoyo R, Hupont I, Lacueva F J and Abadía, David. Hybrid Text Affect Sensing System for Emotional Language Analysis. In: Proceedings of the International Workshop on Affective-Aware Virtual Agents and Social Robots, 2009.

[42] Barreiro A and Gonçalo-Oliveira H. Cross-language semantic relations between English and Portuguese. In: Proceedings of the workshop on iberian cross-language natural language processing tasks, 2011, pp. 49-58.

[43] Malvar-Fernández P and Pichel-Campos J R. 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, 2011, pp. 59-63.

[44] Schmid H. Probabilistic Part-of-Speech Tagging Using Decision Trees. In: Proceedings of the International Conference on New Methods in Language Processing 1994; 12:44-49.

[45] Polanyi L. and Zaenen A. Contextual valence shifters. In Janyce Wiebe (eds) Computing Attitude and Affect in Text: Theory and Applications. Springer, Dordrecht, 2006, pp 1-10.