Building layered, multilingual sentiment lexicons at synset and lemma levels

Fermín L. Cruz , José A. Troyano, Beatriz Pontes, F. Javier Ortega

Department of Languages and Computer Systems, University of Seville, Av. Reina Mercedes s/n, 41012 Sevilla, Spain

ABSTRACT

Many tasks related to sentiment analysis rely on sentiment lexicons, lexical resources containing information about the emotional implications of words (e.g., sentiment orientation of words, positive or negative). In this work, we present an automatic method for building lemma-level sentiment lexicons, which has been applied to obtain lexicons for English, Spanish and other three official languages in Spain. Our lexicons are multi-layered, allowing applications to trade off between the amount of available words and the accuracy of the estimations. Our evaluations show high accuracy values in all cases. As a previous step to the lemma-level lexicons, we have built a synset-level lexicon for English similar to SentiWordNet 3.0, one of the most used sentiment lexicons nowadays. We have made several improvements in the original SentiWordNet 3.0 building method, reflecting significantly better estimations of positivity and negativity, according to our evaluations. The resource containing all the lexicons, ML-SentiCon, is publicly available.

1. Introduction

Sentiment analysis is a modern subdiscipline of Natural Language Processing which deals with subjectivity, affects and opinions in texts (a good survey on this subject can be found in Pang & Lee (2008) and Liu & Zhang (2012)). It is a very active research area, since opinions expressed on the Internet by users constitute a very valuable information for governments, companies and consumers, and its large volume and the high rate of appearance require automated analysis methods. Detection of subjectivity, text classification based on the overall sentiment expressed (positive vs. negative), or extraction of individual opinions and their participants, are three of many tasks addressed. Some of these tasks rely on *sentiment lexicons* as a component of the solutions.

A sentiment lexicon is a lexical resource containing information about the emotional implications of words. Commonly, this information refers to the prior polarity (positive vs. negative) of words, i.e. the positive or negative nature of words, regardless of context. For example, the word "good" has a positive prior polarity, although it may be used in a negative sentence ("His second album is not so good"). In this paper we present new sentiment lexicons for English, Spanish and other three official languages in Spain. The

lexicons are multi-layered, allowing applications to trade off between the amount of available words and the accuracy of the estimations of their prior polarities. As a previous step, we have reproduced the method proposed by Baccianella, Esuli, and Sebastiani (2010) to build SentiWordNet 3.0, one of the most used sentiment lexicons nowadays. We have introduced several improvements to the original method, affecting positively the accuracy of the resource obtained, according to our evaluations.

We believe that the resource containing all the lexicons, ML-SentiCon, can be useful in many sentiment applications for both English and Spanish. The automatic method proposed here could also be reproduced for new languages, whenever WordNet versions for those languages are available. This is advantageous in that it allows to quickly obtain sentiment lexicons for new languages that lack such resources. However, it should also be noted that any lexicon constructed by automatic or semi-automatic methods must be used with caution, as they will inevitably contain errors (words incorrectly labelled as positive or negative). In this sense, it is a good practice to have the lexicons reviewed by native speakers. In the case of ML-SentiCon, layers 1–4 have been completely reviewed. Although the remaining layers have not been reviewed, evaluations based on statistically representative random sample indicate a tolerable error rate up to layer 7 (see Section 4.3 for details).

The structure of the paper is as follows. In Section 2, we review some related works on sentiment lexicons, including a description

of the method used to build SᴇɴᴛɪWᴏʀᴅNᴇᴛ 3.0. Some references to works on non-English sentiment lexicons are also included. In Section 3, we describe our SentiWordNet-based method, and compare the lexicon obtained with the original SᴇɴᴛɪWᴏʀᴅNᴇᴛ 3.0. Section 4 explains the steps followed to obtain the layered, multilingual sentiment lexicons, and shows some results concerning the evaluation of the resource. Finally, in Section 5 we point out some conclusions and final remarks.

2. Related works

There exist many works that deal with the creation of sentiment lexicons with different approaches. General Inquirer (Stone, Dunphy, & Smith, 1966) can be considered, among other things, the first sentiment lexicon. It is a hand-made lexicon constituted by lemmas. Lemmas are semantic units that can appear in multiple lexicalized forms, e.g. the verb "approve" is a lemma that can be found in texts with different inflections, like "approved" or "approving". General Inquirer includes a great amount of information (syntactic, semantic and pragmatic) related to each lemma. Among all this information, there are 4206 lemmas which are tagged as positive or negative. In spite of its age, General Inquirer is still widely used in many works on Sentiment Analysis.

MPQA Subjectivity Lexicon (Wilson, Wiebe, & Hoffmann, 2005) is an example of a piece of work based on General Inquirer. In particular, it is a lexicon which comprises, in addition to the positive and negative words from General Inquirer, a set of automatically compiled subjective words (Riloff & Wiebe, 2003) and also other terms obtained from a dictionary and a thesaurus. It totals 8221 words, whose polarities (positive, negative or neutral) were manually annotated. The resulting list contains 7631 positive and negative elements, and it is very heterogeneous as it is comprised of both lemmas and inflections. As in General Inquirer, this list does not include multi-words, i.e. terms constituted by more than one word.

According to the number of cites, the two most used lexicons nowadays are Bing Liu's Opinion Lexicon (Hu & Liu, 2004; Liu, Hu, & Cheng, 2005) and SᴇɴᴛɪWᴏʀᴅNᴇᴛ (Baccianella et al., 2010; Esuli & Sebastiani, 2006). They are two very different approaches and, to some extent, opposed. Bing Liu's lexicon is formed by 6800 inflections, including mispellings and slangs (informal expressions frequently used on the Internet). On the other hand, SᴇɴᴛɪWᴏʀᴅNᴇᴛ is built based on WᴏʀᴅNᴇᴛ (Fellbaum, 1998), a lexical resource where the basic units, the so-called synsets, comprise a set of words which share the same meaning. Bing Liu's lexicon is built using an automatic method, but the lists of positive and negative words have been manually updated until the current version available on the web, which dates from 2011. On the contrary, SᴇɴᴛɪWᴏʀᴅNᴇᴛ assigns real values, between 0 and 1, representing positive or negative polarities to each of the +100 K synsets of WᴏʀᴅNᴇᴛ. These values have been automatically computed based on two sets of positive and negative seeds, respectively.

It is worthy to note the difference between word-level and lemma-level lexicons, like General Inquirer, MPQA Subjectivity Lexicon or Bing Liu's Opinion Lexicon, and the synset-level lexicons like SᴇɴᴛɪWᴏʀᴅNᴇᴛ. The first ones are formed by terms with semantic ambiguity due to the polysemy of many words. On the contrary, the synset-level lexicons do not have this problem because their basic units univocally represent one meaning. Nevertheless, the use of this kind of lexicons makes it necessary to pre-process the texts with a Word Sense Disambiguation tool, which has a relatively low accuracy nowadays. Most of the works using SᴇɴᴛɪWᴏʀᴅNᴇᴛ compute the polarity at the level of words or lemmas by aggregating the polarity values from all the respective synsets (Agrawal et al., 2009; Denecke, 2008; Desmet & Hoste, 2013;

Kang, Yoo, & Han, 2012; Martín-Valdivia, Martínez-Cámara, Perea-Ortega, & Alfonso Ureña-López, 2012; Saggion & Funk, 2010; Taboada, Brooke, Tofiloski, Voll, & Stede, 2011). In our work, we face the building of lexicons of both types: at a synset level, adding some improvements to the method implemented for SᴇɴᴛɪWᴏʀᴅNᴇᴛ 3.0; and also at a lemma level, using the values computed in the synset-level lexicon.

2.1. SᴇɴᴛɪWᴏʀᴅNᴇᴛ 3.0

The current version of SᴇɴᴛɪWᴏʀᴅNᴇᴛ (Baccianella et al., 2010) assigns positivity and negativity values between 0 and 1 to each synset in WᴏʀᴅNᴇᴛ 3.0. It uses an automatic method divided in two steps (Fig. 1).

In the first step, the polarities of synsets are estimated individually: various ternary classifiers are trained, which are able to classify each synset as positive, negative or neutral, depending on the words contained in the definition of the synset (WᴏʀᴅNᴇᴛ provides a definition, the so-called "gloss", for each synset). Starting from some positive and negative seeds, and after applying an expansion method, different training sets are obtained. Then, standard techniques for text classification are applied (tf-idf vectorial representation of the glosses, plus SVM and Rocchio algorithms). Finally, the resulting classifiers are applied to each synset in WᴏʀᴅNᴇᴛ and their positivity and negativity scores are computed from the outputs of each classifier.

In the second step, these scores are globally refined. A graph of synsets is built, where nodes n_i correspond to each synset s_i, and an edge from n_i to n_j is created if, and only if, the synset s_i appears in the definition of the synset s_j. Note that WᴏʀᴅNᴇᴛ glosses are non-disambiguated texts, so it is necessary the use of Princenton Word-Net Gloss Corpus,[1] a resource where WᴏʀᴅNᴇᴛ glosses are partially disambiguated. This graph is defined as a part of the inverse flow model (Esuli & Sebastiani, 2007). The intuition behind is the assumption that those synsets whose definition contains positive synsets are likely to be positive, and analogously for the negative ones. In the inverse flow model, a variation of the random-walk algorithm PageRank (Page, Brin, Motwani, & Winograd, 1998) is applied to the graph. This algorithm propagates the positivity scores computed in the previous step through the edges of the graph, in order to obtain the positivity values for each synset. The same process is applied to the negativity scores in a second computation of the algorithm.

2.2. Non-English sentiment lexicons

Although there are not many sentiment lexicons for other languages than English, the number is growing slowly. There exist works focused on the creation of sentiment lexicons for very diverse languages, such as Hindu and French (Rao & Ravichandran, 2009), Arabian (Abdul-Mageed, Diab, & Korayem, 2011), German (Clematide & Klenner, 2010), Japanese (Kaji & Kitsuregawa, 2007), Chinese (Lu, Song, Zhang, & Tsou, 2010; Xu, Meng, & Wang, 2010), Romanian (Banea, Mihalcea, & Wiebe, 2008) and Spanish. Two Spanish lexicons are automatically built in Brooke, Tofiloski, and Taboada (2009) from an English sentiment lexicon by using two resources: a bilingual dictionary[2] and Google Translator.[3] The authors do not show any evaluation of the resulting lexicons, but they provide the results obtained by a sentiment classification tool based on them. A similar technique is used in Molina-González, Martínez-Cámara, Martín-Valdivia, and Perea-Ortega (2013), where an automatic translation process (from English

---

[1] http://wordnet.princeton.edu/glosstag.shtml.
[2] http://www.spanishdict.com.
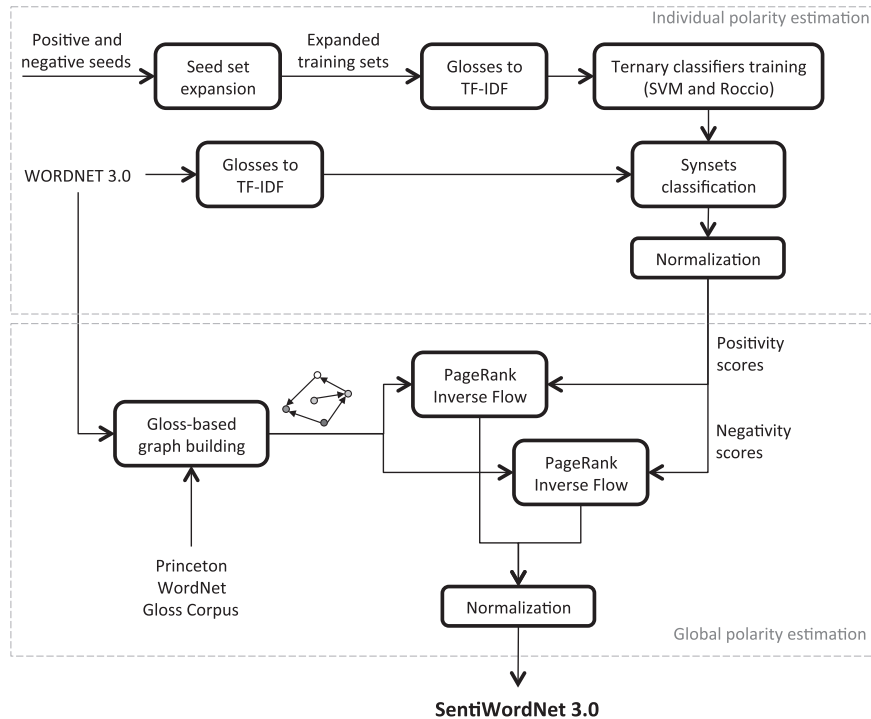[3] http://translate.google.com.

Fig. 1. Schematic representation of the SentiWordNet 3.0 building method.

to Spanish) is applied to Bing Liu's Opinion Lexicon. Some inherent ambiguities of the automatic translation were manually solved. Furthermore, they manually added translations for some of the informal words and slangs from the original lexicon. There are no direct evaluation results in this work either. Instead, the authors show the performance of an extrinsic evaluation based on a binary sentiment classification of texts. In Pérez-Rosas, Banea, and Mihalcea (2012), two English lexicons are the basis of the approach: MPQA Subjectivity Lexicon (automatically mapped over WORDNET) and SENTIWORDNET. They also use a Spanish version of WORDNET, included as a part of the EuroWordNet project (Vossen, 1998), in order to obtain lists of positive and negative terms in Spanish. Two variations of this method are proposed, resulting in two Spanish lexicons with 1347 and 2496 terms, respectively. The evaluations are performed through the manual annotation of a test set of 100 terms from each lexicon, and they show an accuracy of 90% and 74%, respectively.

2.3. Our approach

In our approach we face the task of building lexicons at both synset level, as SentiWordNet (Baccianella et al., 2010), and lemma level, as General Inquirer (Stone et al., 1966) or MPQA Subjectivity Lexicon (Wilson et al., 2005) – though we include multiwords, and our lemmas are morphologically disambiguated. We propose both approaches because most of the works using synset-level lexicons actually compute polarity at the level of words or lemmas, by some kind of aggregation of the polarity values for all the respective synsets (Agrawal et al., 2009; Denecke, 2008; Desmet & Hoste, 2013; Kang et al., 2012; Martín-Valdivia et al., 2012; Saggion & Funk, 2010; Taboada et al., 2011).

In the case of the lexicon at the synset level, we have adapted the technique used for the creation of SENTIWORDNET 3.0, improving each step of the method. Compared with the original method in Baccianella et al. (2010), we have added WORDNET-AFFECT (Strapparava, Valitutti, & Stock, 2006) as a new source of information, applied a new meta-learning scheme, and proposed new

kinds of WordNet-based graphs and a different random-walk algorithm based on POLARITYRANK (Cruz, Vallejo, Enríquez, & Troyano, 2012) (see Section 3 for details). As we will discuss in Section 3.3, these contributions lead to estimations of the positivity and negativity of synsets 24.2% and 7.36% better than those in Baccianella et al. (2010).

Once this resource was built, we have induced lexicons at lemma level. Compared with all other lemma-level lexicons in the literature, we propose a new layered structure, which allows to choose different compromises between the amount of available lemmas and the accuracy of the estimations of their polarities. As we are interested in the availability of sentiment lexicons for other languages than English, we have also induced lemma-level lexicons for Spanish and three other official languages in Spain. We have employed the various languages versions of WordNet from EuroWordNet project (Vossen, 1998), as they did in Pérez-Rosas et al. (2012), but we have also relied on the more recent Multilingual Central Repository 3.0 (Gonzalez-Agirre, Laparra, & Rigau, 2012). The Spanish version of our lemma-level lexicon achieves sensibly better accuracy and volume than those reported by Pérez-Rosas et al. (2012) (the only work on Spanish sentiment lexicons that reports accuracy).

3. Building a synset-level sentiment lexicon

Based on the method used by Baccianella et al. (2010) to generate SENTIWORDNET 3.0, and incorporating various modifications, we have calculated values of positivity, negativity and neutrality for each of the synsets in WORDNET 3.0. In the same way as the method on which we rely, our approach is divided into two distinct parts: firstly, we have combined various classifiers (trained using various sources of information) in order to obtain positivity, negativity and neutrality values of the synsets from their glosses. This step is named *individual polarity estimation*. Secondly, the obtained values have been refined by applying a random-walk algorithm called POLARITYRANK (Cruz et al., 2012), capable of working with graphs

having both positive and negative weighted edges. This algorithm has been applied to two different graphs created from WORDNET: a gloss-based graph and a semantic relations based graph. We named this step *global polarity estimation*. The obtained lists of values of positivity and negativity were evaluated and compared with those of SENTIWORDNET 3.0, obtaining results 24.2% and 7.36% better than those of the cited work.

In the following, we explain in detail each of the steps involved in the process (see Fig. 2 for an schematic summary).

3.1. First step: an ensemble of text classifiers of glosses

In this step several ternary classifiers are created, capable of deciding whether a synset is positive, negative or neutral from the text of its gloss. In order to train these classifiers, different examples of positive, negative and neutral synsets are used. These examples are extracted from two different sources of information, such as WORDNET-AFFECT (Strapparava et al., 2006) and General Inquirer (Stone et al., 1966), among others. The individual classifiers, which have been trained starting from the different sources and using different classification algorithms, are afterwards combined using a meta-learning scheme. We now proceed to explain the creation of the training sets, the training of the individual classifiers and, finally, the assembling of the final meta-classifiers.

3.1.1. Collecting the training sets

To build the training sets of positive and negative synsets we have used two different data sources. On one hand, we used the paradigmatically positive and negative terms proposed by Turney and Littman (2003).[4] Since a term may be associated with multiple synsets in WORDNET, in SENTIWORDNET 3.0 all synsets associated to the positive and negative terms were used as initial sets. In our case, we have manually reviewed the list of synsets to filter out those senses that are not clearly positive or negative. Thus, we selected 38 positive synsets (out of 75) and 43 negative (out of 65). We call these sets P_{Turney} and N_{Turney}, respectively.

On the other hand, we obtained two sets of positive and negative synsets from WORDNET-AFFECT 1.1 (Strapparava et al., 2006), a resource in which affective tags are associated to a subset of synsets in WORDNET 1.6. Specifically, we conserved those synsets labelled as POSITIVE-EMOTION or NEGATIVE-EMOTION, or other affective tags corresponding to specializations of the above (for example, POSITIVE-EXPECTATION and LOVE are specializations of POSITIVE-EMOTION, and GENERAL-DISLIKE and NEGATIVE-FEAR are specializations of NEGATIVE-EMOTION). The synsets obtained from these labels correspond to version 1.6 of WORDNET. To obtain equivalent sets formed by synsets from version 3.0 of WORDNET, we used WN-Map,[5] a resource containing mappings among WORDNET versions. In some cases there are several synsets in version 3.0 corresponding to a synset in version 1.6, and vice versa. These cases were reviewed manually. The sets finally obtained are significantly larger than P_{Turney} and N_{Turney}, having 243 positive and 436 negative synsets. These sets are called P_{WNA} and N_{WNA}.

The initial sets are expanded using the same method in SENTIWORDNET 3.0. The procedure starts from a set of positive synsets P and a set of negative ones N. For each synset in P, we add to the same set P other synsets connected to the first one by semantic relations in WORDNET preserving the polarity. Similarly, we add to N those synsets connected to the first one by semantic relationships inverting the polarity. The same process is performed with synsets in N. Since in Baccianella et al. (2010) is not explained

which semantic relationships were used, we have examined each of those in WORDNET 3.0. We have chosen 7 relationships that preserve polarity and 1 that reverse it (see Table 1). The expansion process is applied 6 times over sets P_{Turney} and N_{Turney}, on one hand, and over sets P_{WNA} and N_{WNA} on the other, thus obtaining sets with increasing sizes (though presumably increasingly unreliable). Table 2 shows the sizes of the different sets, where P^k and N^k represent sets of positive and negative synsets obtained after k executions of the expansion method ($k = 0$ represents the initial sets).

In order to train the ternary classifiers we also need a set of neutral synsets, i.e., without positive or negative connotations. For this purpose, we use those synsets not marked either as positive or negative in the General Inquirer lexicon (Stone et al., 1966). The set of objective synsets O_{Turney} is obtained by discarding synsets included in either P_{Turney}^6 or N_{Turney}^6. Similarly, O_{WNA} is obtained by discarding synsets in P_{WNA}^6 or N_{WNA}^6. Sets O_{Turney} and O_{WNA} contains 50286 and 49590 synsets, respectively.

3.1.2. Training the ternary classifiers

Synsets need to be represented as vectors so that classifiers can work with them. For this purpose, the standard technique tf-idf is applied to their glosses. Before obtaining this representation, stop words are removed from the glosses, as well as those parts of the text that may be affected by negations, since these words might mislead the classifiers. Once these vectorial representations have been obtained, diverse ternary classifiers are trained using on one hand all the synsets from P_{Turney}^k, N_{Turney}^k and O_{Turney}, with $k \in [0, 6]$ (7 training sets), and on the other hand all the synsets from P_{WNA}^k, N_{WNA}^k and O_{WNA}, with $k \in [0, 6]$ (another 7 training sets). Two different classification algorithms are used: SVM and Rocchio. A total of 28 different classifiers are obtained, which are finally combined in a meta-learning phase to create the final classifiers.

3.1.3. Assembling the classifiers

When applying any of the previous classifiers to a synset in WORDNET, three output probabilities are obtained, corresponding to the three considered classes (positive, negative and neutral). In SENTIWORDNET, individual classifiers are combined by a simple voting scheme. However, in our proposal, we opted for a more powerful combination scheme based on meta-learning, making use of the information in Micro-WN (Op)-3.0.[6] This is a resource containing manually entered values of positivity, negativity and neutrality, for a total of 1054 synsets. Obtaining the final classifiers involves the following steps:

1. The 946 synsets that form Sections 2 and 3 of Micro-WN (Op)-3.0 are selected (the resource contains three sections in total). Vectorial representations of the synsets are computed from their glosses.
2. The 28 individual classifiers are run on each of the above vectors, obtaining as output the class membership probabilities of each synset for the positive, negative and neutral classes.
3. Three datasets are created, accumulating each one of the previous 84 values for each synset, along with one of the values listed in Micro-Wn (Op)-3.0 (according to the dataset: positivity, negativity or neutrality).
4. Three regressional classifiers are trained, one for each of the above datasets, using Partial Least Squares Regression algorithm (de Jong, 1993). Given a synset, each of these classifiers is able to predict a positivity, negativity or neutrality score, respectively.

Table 3 shows the correlation coefficients between the true values of positivity, negativity and neutrality, from Micro-WN
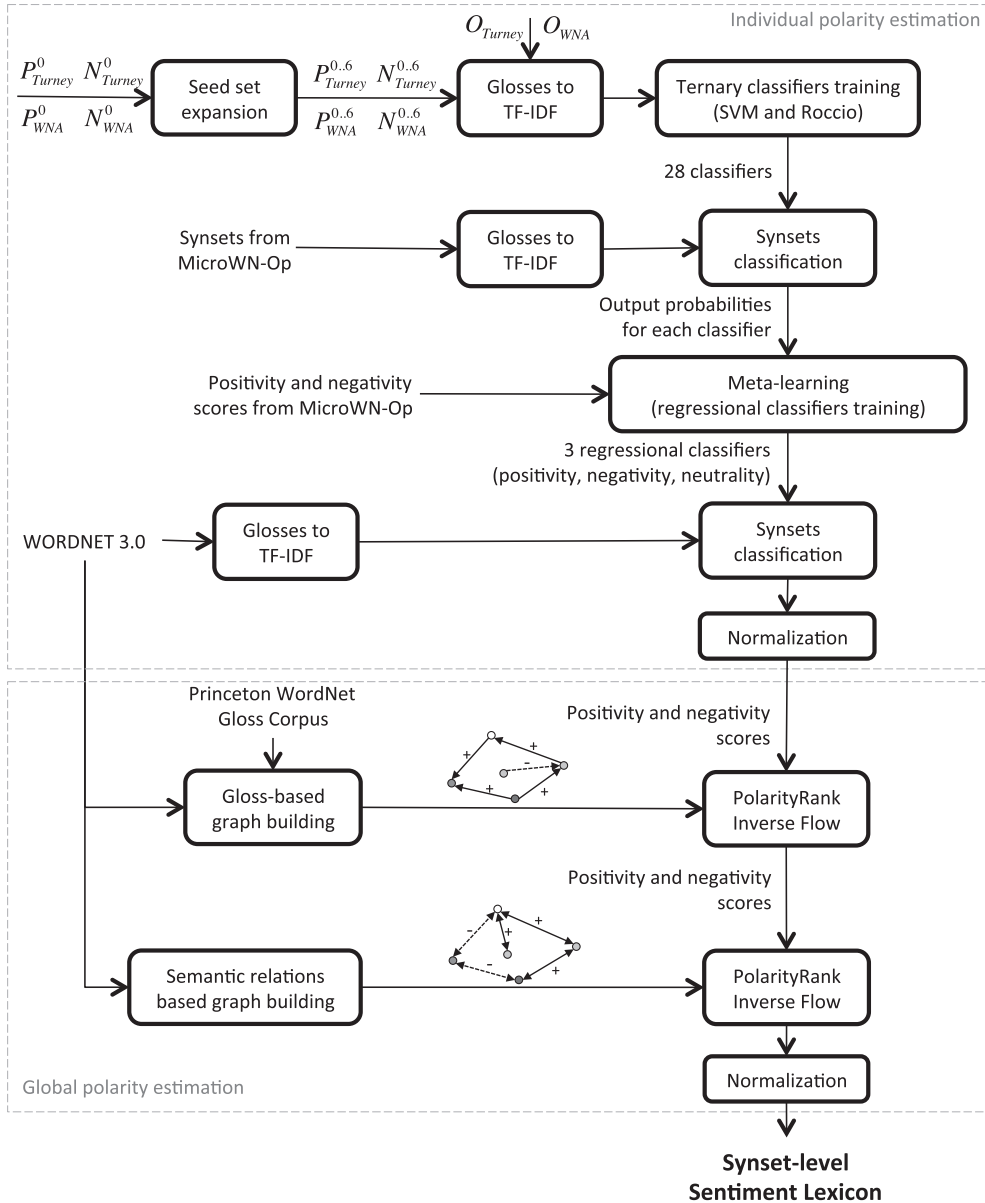

Fig. 2. Schematic representation of our method for building the synset-level sentiment lexicon.

<div style="display:flex">
<div>

Table 1
Semantic relations from WORDNET 3.0 used in the seed set expansion algorithm.

same polarity:	derivationally_related_form, entailment, cause, similar_to, derived_from_adjective, pertainym_(pertains_to_nouns), also_see
opposite polarity:	antonym

</div>
<div>

Table 2
Number of synsets of each training set.

k	P_{Turney}^k	N_{Turney}^k	P_{WNA}^k	N_{WNA}^k
0	38	43	243	436
1	209	222	1014	1121
2	747	774	2124	2336
3	1907	2025	4017	4221
4	3948	4115	6707	6921
5	7046	7016	10216	10535
6	10891	10802	14408	14298

</div>
</div>

(Op)3.0, and those estimated by the regressional classifiers. In order to measure the information provided to the classifiers by the two types of seed sets used, we show in this table the correlation coefficients using only the individual classifiers based on P_{Turney} and N_{Turney}, or on P_{WNA} and N_{WNA}. As it can be observed, both types of seeds provide the regressional classifiers with valuable information.

After obtaining the positivity, negativity and neutrality regressional classifiers, they are applied to other synsets in WORDNET. The following steps are performed for each of them:

1. Its vectorial representation is obtained from its gloss.
2. Former 28 individual classifiers are executed over the previous vector, obtaining as output the probabilities of belonging to positive, negative and neutral classes of the synset.
3. An instance is created from the prior probabilities, and the positivity, negativity and neutrality regressional classifiers are executed over this instance.

Table 3
Correlation coefficients of estimated and real values of positivity, negativity and neutrality.

Seeds	Positivity	Negativity	Neutrality
WNA	.5873	.4941	.6291
Turney	.6433	.5981	.6169
Both	.6891	.6290	.6680

4. The obtained values are normalized and discretized so that the sum of all of them is equal to 1, and their magnitudes are adapted to increments of 0.125.

3.2. Second step: a random-walk through synsets using PolarityRank

The aim of this second step is to refine the values of positivity and negativity assigned to each synset, using different types of relationships among the synsets. These relations are modeled by means of a graph, in which nodes represent synsets and directed edges indicate some kind of relationship between the positivity and negativity values of the synsets. Each node is assigned a numerical value (for example, positivity values), and then a random-walk algorithm is applied to the graph. These algorithms are able to iteratively compute the interactions occurring between the values assigned to nodes: the values initially assigned to the nodes "flow"[7] along the graph through edges. Once converged, the algorithm obtains final values for the nodes, which will depend both on their initial assigned values and the relationships between nodes at a global level.

Two main differences can be discerned of our approach regarding SENTIWORDNET 3.0. First, we build two different types of graphs, one from the glosses and another one from the semantic relations in WORDNET (SENTIWORDNET only uses a graph based on the glosses). In both cases, the graphs include edges with positive weights, representing a direct transfer between the positivity and negativity values of the connected synsets, and negative weighted edges, indicating a cross-transfer between the two types of values (in SENTIWORDNET only edges without weights are taken into account). Second, we apply a random-walk algorithm that allows to compute the final values of positivity and negativity in a single run. There also exists an interdependence between the final positive and negative values (in SENTIWORDNET two independent executions of the random-walk algorithm were carried out, one for positivity values and another one for the negativity ones).

Next, we describe the two types of graphs we have built, present the employed random-walk algorithm formulation, and, finally, we list the steps to obtain the final lexicon at synset level.

3.2.1. The gloss-based graph

This graph is similar to the one used in the random-walk step of SENTIWORDNET 3.0 (see Section 2.1), in which positivity and negativity values flow from synsets comprising a gloss to the synset being defined. Based on this idea, we have added weights to the edges of the graph (in some cases negative values), as explained below.

To build the graph we rely on the Princeton WordNet Gloss Corpus resource, in which the glosses of WORDNET are partially disambiguated, i.e. some of the words in the glosses are tagged with the synset to which they correspond, according to the context where they are being used. However, only 56% of the words are disambiguated in the resource. In those cases in which the words are disambiguated, we assign a weight of 1.0 to the corresponding

---

[7] The term "flow" is often used evoking the metaphor in which nodes would correspond to fluid reservoirs, the initial values would indicate a volume of liquid in each of these tanks, and the edges would be pipes connecting the tanks, permitting liquid to flow between nodes.

edges. For those words that are not disambiguated, edges are created for each of the possible synsets, whenever there is a maximum of 5 (words with high polysemy are discarded this way). A weight of $1/n$ is assigned in this case to each of the n resultant edges.

For those words affected by a negation in the glosses, we reverse the weight of their corresponding edges. The idea behind is that the participation in a gloss of a positive word that appears negated indicates that the synset being defined may be negative, and vice versa. The random-walk algorithm that will be applied should be aware of these cross-dependencies between the positivity and negativity values of the nodes.

3.2.2. The semantic relations based graph

In this second type of graph, nodes also represent the synsets, but in this case the edges represent the semantic relations of WORDNET *similar_to* and *antonym*. Since both relations are reflexive, two edges are added between the synsets involved, one in each direction. The edges corresponding to the *similar_to* relation are assigned a weight of 1.0, since similar synsets should obtain similar values of positivity (or negativity). The edges on the relation *antonym* are assigned a weight of −1.0, since antonym synsets should get crossed values of positivity and negativity.

The number of edges is sensibly lower than in the previous graph, although the proportion of positive and negative ones is more balanced (see Table 4).

3.2.3. PolarityRank inverse flow

In order to compute how the values of positivity and negativity flow through these graphs, we use a variation of the POLARITYRANK algorithm (Cruz et al., 2012). POLARITYRANK is a variation of PAGERANK (Page et al., 1998) capable of handling graphs with edges of positive and negative weights. It calculates two values for each node in a single run, PR^+ and PR^-. The following equation corresponds to the formula used to calculate these values:

$$PR^+(n_i) = (1-d)e_i^+ + d\left(\sum_{j \in In^+(n_i)} \frac{w_{ji}}{outW_j}PR^+(n_j) + \sum_{j \in In^-(n_i)} \frac{-w_{ji}}{outW_j}PR^-(n_j)\right)$$

$$PR^-(n_i) = (1-d)e_i^- + d\left(\sum_{j \in In^+(n_i)} \frac{w_{ji}}{outW_j}PR^-(n_j) + \sum_{j \in In^-(n_i)} \frac{-w_{ji}}{outW_j}PR^+(n_j)\right)$$

(1)

where $outW_j$ refers to the sum of the absolute values of the weights of the outgoing edges from node n_j, w_{ji} represents the weight of the edge from node n_j to node n_i, and $In^+(n_i)$ and $In^-(n_i)$ are the sets formed by those j having edges from n_j to n_i with a positive or negative weight, respectively. Constants e_i^+ and e_i^- correspond to the initial values of positivity and negativity of node n_i. Constant d is a real value between 0 and 1 that allows measuring the influence of these initial values on the final values: if $d = 0$, the final values will match up the initial ones, while, at the other end, if $d = 1$, the final values will only depend on the graph topology, and never on the initial values. Initially assigning a constant value to PR^+ and PR^- for every node, the values are iteratively recalculated using the formula, until the differences between two iterations are less than a threshold.

Table 4
Number of nodes and links in graphs.

Graph type	Nodes	Links	Positive links	Negative links
Glosses-based	117659	711063	697149	13914
Semantic relations based	117659	28990	21386	7604

The intuition behind the formula is that the value of PR^+ (or PR^-) of a node is directly proportional to the values of PR^+ (or PR^-) of those other nodes having incoming edges to the first, provided that the weights of these edges are positive. Besides, the value of PR^+ (or PR^-) of the node is also directly proportional to the values of PR^- (o de PR^+) of those nodes having incoming edges to the first, provided that the weights of these edges are negative. Here we can see the cross-dependency between the values PR^+ and PR^-, which fits with the semantics of the two types of graphs that have been proposed in the previous sections.

Moreover, and similarly to PageRank, $outW_j$ denominators penalize the contribution of those nodes having many outgoing edges. It makes sense in some specific contexts in which outgoings edges represent votes, where it is intended that the votes of those nodes emitting fewer total votes have more impact than those votes from nodes voting massively. However, this penalty does not make sense in the graphs presented above: if we focus on the gloss-based graph, the fact that a synset is used in many or few glosses should not affect its ability to convey positivity or negativity. Moreover, it should be desirable that the positivity or negativity value of a synset should be more influenced by a specific synset of its gloss when the size of the gloss is smaller. This size is related to the total incoming edges (not outcoming). The same considerations can be made to the semantic relations based graph.

Taking into account all former issues, we use a modification of PolarityRank, which has been named PolarityRank Inverse Flow[8]:

$$PR_{IF}^+(n_i) = (1-d)e_i^+ + \frac{d}{inW_i}\left(\sum_{j \in In^+(n_i)} w_{ji}PR^+(n_j) + \sum_{j \in In^-(n_i)} -w_{ji}PR^-(n_j)\right)$$

$$PR_{IF}^-(n_i) = (1-d)e_i^- + \frac{d}{inW_i}\left(\sum_{j \in In^+(n_i)} w_{ji}PR^-(n_j) + \sum_{j \in In^-(n_i)} -w_{ji}PR^+(n_j)\right)$$

$$(2)$$

where inW_i represents the sum of the absolute values of the weights of the incoming edges to node n_i, and the other terms have the same meaning as in Eq. (1).

3.2.4. Method

Once the participants at this stage have been presented, we describe below the steps followed to obtain the final refined values of positivity and negativity at synset level:

1. The gloss-based graph is created. Constants e_i^+ and e_i^- are obtained from the positivity and negativity values calculated in the individual polarity estimation step. These values are normalized so that the size of vectors e^+ y e^- is equal to the total number of nodes.
2. Values for constant d are adjusted, using for this purpose Section 1 of MicroWN (Op)-3.0: those values of d leading to better final results of positivity and negativity for the nodes are chosen.
3. PR^+ and PR^- values obtained for the selected values of constant d are fit to a power law distribution, whose parameters are computed from the values of positivity and negativity of the individual polarity estimation step (this calculation is more detailed in Baccianella et al. (2010)). These values correspond to the new refined positivity and negativity values of the nodes. If the sum of the two is greater than 1, they are normalized to ensure the sum is 1.

4. Former steps are repited, this time for creating the semantic relations based graph, and using the positivity and negativity values from step 3 as initial values. This implies a second refinement of the positivity and negativity values.

3.3. Evaluation

We have used the same gold-standard and metric used in Baccianella et al. (2010) to measure the accuracy of the final values of positivity and negativity, as well as the contributions of the modifications introduced regarding the original SentiWordNet 3.0 method. The evaluation consists of comparing the rankings (i.e., ordered lists) of positivity and negativity scores from manually annotated synsets in MicroWN (Op)-3.0 to those obtained by our method. Kendall distance τ_p (Fagin, Kumar, Mahdian, Sivakumar, & Vee, 2004) is used to compare both rankings, measuring the similarity between a gold standard raking and a tested ranking. The distance is computed with the following formula:

$$\tau_p = \frac{n_d + p \cdot n_u}{Z},\qquad(3)$$

where n_d is the number of discordant pairs (reversely ordered in one ranking and the other), n_u is the number of not tied pairs in the gold standard ranking that are tied in the tested ranking, and p represents a penalty factor assigned to each of these pairs (we used $1/2$). Z is the total number of ordered pairs in the gold standard ranking, and ensures the range of the measure to be in the interval $[0, 1]$. The closer to zero is τ_p, the more similar the tested and the gold standard *rankings* are.

As in Baccianella et al. (2010), we used Sections 2 and 3 of MicroWN (Op)-3.0 for the evaluation. Since this resource is also needed to train the regressional classifiers, we have used cross-validation, preventing the same instances being used for both training
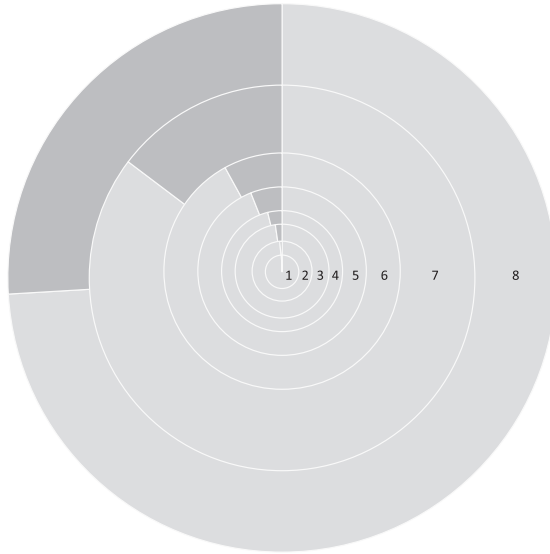
Table 5
Values of τ_p for each experiment.

Step	Experiment	Positivity	Negativity
Individual Estimation	[SentiWordNet 3.0]	0.339	0.286
	Meta-Learning	0.238	0.284
Global Estimation	[SentiWordNet 3.0]	0.281	0.231
	GB + PageRank	0.226	0.283
	GB + PolarityRank	0.218	0.216
	GB + SRB + PolarityRank	0.213	0.214

Table 6
Total available lemmas for each language.

Language	Nouns	Adjectives	Verbs	Adverbs	Total
English	119052	21538	11531	4481	156602
Spanish	59436	9539	6819	913	76707
Catalan	38751	4237	4648	2	47638
Basque	22879	57	3456	0	26392
Galician	17535	4910	965	0	23410

Table 7
Distribution of lemmas by languages and layers.

Layer	English	Spanish	Catalan	Basque	Galician
1	157	353	512	138	49
2	982	642	530	278	223
3	1600	891	699	329	370
4	2258	1138	860	404	534
5	3595	1779	1247	538	883
6	6177	2849	1878	888	1429
7	13517	6625	4075	2171	2778
8	25690	11918	7377	4323	4930

---

[8] We have used this name because the changes from PolarityRank are similar to the changes in the random-walk algorithm of the inverse flow model with respect to PageRank.

Layer	Accuracy	Size
1	99.36%	157
2	98.88%	982
3	97.75%	1600
4	96.24%	2258
5	93.95%	3595
6	91.99%	6177
7	85.29%	13517
8	74.06%	25690

Fig. 3. Accuracy and number of lemmas by layers for English lemma-level lexicons. The area of each pie chart is proportional to the number of lemmas in each layer. The dark areas represent the incorrect entries in each layer.

and testing at the same time. Table 5 contains the results of various experiments, grouped into two sets. First, the evaluation results for the positivity and negativity scores from the individual polarity estimation step are shown (META-LEARNING), together with the results reported in Baccianella et al. (2010) corresponding to the same step. Second, the results of applying the random-walk phase are shown, compared with the evaluation results of SENTIWORDNET 3.0. We have performed diverse intermediate experiments, in order to measure the contribution of each of the two types of graphs, the gloss-based one (GB) and the semantic relations based one (SRB). We have also quantified the contribution of POLARITYRANK with regard to PAGERANK (we used the modified versions for the inverse flow model in both cases). In the case of PAGERANK, negative edges were removed from the graph, since this algorithm is not capable of working with them.

Results confirm the benefits of each of our contributions, individually and globally, with regard to the original method used for the construction of SENTIWORDNET 3.0. In the individual polarity estimation step, using WORDNET-AFFECT for building the initial sets of positive and negative synsets, together with the combination of individual classifiers in a meta-learning scheme, lead to a better estimation of the positivity scores (τ_p is reduced in almost a 30%). Interestingly, there is no significant improvements in the estimation of the negativity scores. As for the overall polarity calculation, the final estimations of positivity and negativity scores using our complete method are fairly more accurate than those of SENTIWORDNET 3.0 (τ_p is reduced a 24.2% and 7.4%, respectively). The mayor contribution to this improvement is due to the use of POLARITYRANK instead of PAGERANK, especially in relation to negativity values estimation. Nevertheless, the use of the semantic relations graph provides a more moderate improvement.

4. ML-SENTICON: multilingual, layered sentiment lexicons at lemma level

A review of the bibliography reveals that many works on sentiment analysis based on SENTIWORDNET do not use directly the polarity values of the synsets (Agrawal et al., 2009; Denecke, 2008; Desmet & Hoste, 2013; Kang et al., 2012; Martín-Valdivia et al., 2012; Saggion & Funk, 2010; Taboada et al., 2011). Instead,

these works focus on the computation of polarity scores at a lemma level from the polarity values of the synsets which those lemmas are related to. The main reason for implementing a method like this is to avoid the use of a Word Sense Disambiguation component due to its low accuracy, or maybe because they need to analyze a huge amount of texts in a real-time environment where the use of a time-consuming linguistic pre-process is not an option. In our case, we have built a lemma-level lexicon from the synset-level lexicon, in order to facilitate the use of the resource to those researchers who do not want to use a Word Sense Disambiguation tool. Furthermore, we have generated different versions of the lexicon, for Spanish and three other official languages in Spain, by using resources that allow us to interconnect each synset with lemmas in other languages. The resulting resource, ML-SENTICON, is publicly available.[9]

The lemma-level sentiment lexicons are composed by 8 layers, each one containing both positive and negative lemmas. The layers are sorted in such a way that a given layer contains all the lemmas from the previous layers, in addition to some new lemmas. The lemmas of each layer are obtained by gradually relaxing a set of restrictions. In this way, the number of lemmas that satisfy the restrictions increases in each layer, at the same time as the reliability of those lemmas as indicators of positivity and negativity decreases. In the next section we formally define the process intended to build the layers.

4.1. Defining the layers

Each synset s_i in WORDNET is related to a set of lemmas $L_i = \{l_1, l_2, \ldots, l_n\}$ (so-called variants) with the same morphosintactic category (noun, adjective, verb or adverb). Furthermore, each synset s_i has a positivity value, p_i, and a negativity value, n_i, in our synset-level sentiment lexicon. Thus, we define the polarity of a synset s_i as $pol_i = p_i - n_i$. In order to define our resource, let us consider a synset s_i as the tuple composed of the set of lemmas in s_i and its polarity, that is $s_i = (L_i, pol_i)$. From the other point of view, we can say that each lemma, l, is related to a set of synsets $S_l = \{s_i : l \in L_i\}$. Let us note as $\overline{pol_l}$ the mean value of the polarities pol_i of the synsets s_i in S_l. A high value of $\overline{pol_l}$ could mean a higher

---

[9] http://www.lsi.us.es/~fermin/index.php/Datasets.
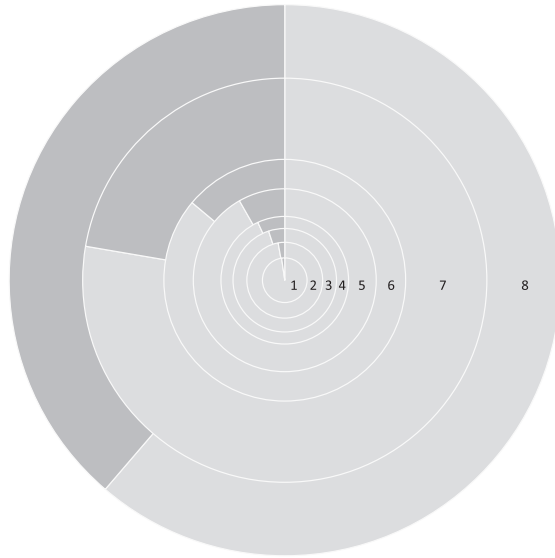
Layer	Accuracy	Size
1	97.73%	353
2	97.20%	642
3	94.95%	891
4	93.06%	1138
5	91.75%	1779
6	86.09%	2849
7	77.69%	6625
8	61.29%	11918

Fig. 4. Accuracy and number of lemmas by layers for Spanish lemma-level lexicons. The area of each pie chart is proportional to the number of lemmas in each layer. The dark areas represent the incorrect entries in each layer.

```xml
<?xml version="1.0" encoding="UTF-8" ?>
<layers lang="en">
  <layer level="0">
    <positive>
      <lemma pos="a" pol="1.0" std="0.0"> admirable </lemma>
      <lemma pos="a" pol="0.438" std="0.088"> amorous </lemma>
      <lemma pos="n" pol="0.594" std="0.12"> approval </lemma>
      <!-- it continues... -->
    </positive>
    <negative>
      <lemma pos="a" pol="-0.437" std="0.157"> afraid </lemma>
      <lemma pos="n" pol="-0.275" std="0.105"> aggression </lemma>
      <lemma pos="v" pol="-0.25" std="0.0"> alarm </lemma>
      <!-- it continues... -->
    </negative>
  </layer>
  <layer level="1">
    <positive>
      <lemma pos="a" pol="1.0" std="0.0"> adept </lemma>
      <lemma pos="r" pol="0.25" std="0.0"> admirably </lemma>
      <lemma pos="n" pol="0.667" std="0.315"> admiration </lemma>
      <!-- it continues... -->
    </positive>
    <negative>
      <lemma pos="v" pol="-0.875" std="0.0"> abase </lemma>
      <lemma pos="a" pol="-0.25" std="0.0"> abashed </lemma>
      <lemma pos="v" pol="-0.25" std="0.0"> abhor </lemma>
      <!-- it continues... -->
    </negative>
  </layer>
  <layer level="2">
      <!-- it continues... -->
  </layer>
</layers>
```

Fig. 5. An extract of the English lexicon of ML-SentiCon.

intensity of the positive or negative polarity of lemma l, but also a lower ambiguity of the term, or even a bigger reliability of the estimations computed by the automatic method used in the creation of the synset-level sentiment lexicon.

Each of the 8 layers is constituted by a set of positive lemmas l, with $\overline{pol}_l > 0$, and a set of negative lemmas l, with $\overline{pol}_l < 0$. The first two layers only contains lemmas $l \in L_i$ from synsets s_i that appear in some of the training sets used in the individual polarity

computation step (P^0_{Turney}, N^0_{Turney}, P^0_{WNA} and N^0_{WNA}, see Section 3.1.1). In particular, if at least one half of the synsets in S_l are contained in the previously mentioned seed sets, the lemma l belongs to the first layer. On the other hand, if at least one of the synsets in S_l is contained in those seed sets, the lemma l belongs to the second layer. The rest of layers are constituted by lemmas $l \in L_i$ from any synsets s_i in WORDNET. Furthermore, each layer requires different minimum values over the absolute value of $\overline{pol}_l$ of its lemmas. These values have been chosen in such a way that the number of lemmas in each layer follows a geometric progression (see Table 7).

4.2. Mapping synsets to multilingual lemmas

We use the Multilingual Central Repository 3.0 (MCR 3.0) (Gonzalez-Agirre et al., 2012) in order to obtain the mapping between synsets and lemmas in other languages. MCR 3.0 integrates incomplete WordNets of four different languages: Spanish, Catalan, Basque and Galician. The synsets of these WordNets are connected to those in WORDNET 3.0, which allows us to carry out the same process for the creation of the layers explained in Section 4.1. In the case of Spanish and Catalan, we use as well the information provided by the EuroWordNet project (Vossen, 1998) up to November 2006, increasing the number of lemmas for these languages. EuroWordNet includes incomplete WordNets linked to WORDNET 1.6, so we have mapped it to WORDNET 3.0 through WN-Map, the same resource used in Section 3.1.1. We show in Table 6 the total number of lemmas available for each language from these resources. On the other hand, in Table 7 we show the distribution of the lemmas in layers and languages for the resulting lemma-level sentiment lexicons.

4.3. Evaluation of the English and Spanish lexicons

In order to evaluate the quality of the lemma-level lexicons, we have manually reviewed the lists of positive and negative lemmas in each layer, annotating each element as correct or incorrect. We follow this method for the evaluation of the English and Spanish lexicons. For the first four layers (layers 1–4), we have reviewed the complete lists of lemmas. For the rest of layers (5–8), we have reviewed a statistically representative random sample of each layer. We have estimated a sample size that warranties an error lower than $\pm 5\%$ in the estimation of correct elements, assuming a binomial distribution of the random variable with $p = q = 0.5$ (the worst case scenario), and a confidence interval of 95% ($\alpha = 0.05$). With these parameters we obtain sample sizes between 300 and 400 elements, depending on the level.

In Figs. 3 and 4 we show the accuracy (percentage of correct elements over the total) for each layer of Spanish and English lexicons, respectively. The results prove a high reliability of the generated lists of positive and negative lemmas, with accuracies above 90% in layers 1–6 of the English lexicon and layers 1–5 in the Spanish one. The accuracy is higher in the English lexicon than in the Spanish lexicon, which is reasonable because the Spanish lexicon has been built from resources generated using semi-supervised methods, with possible errors. The difference between the accuracy in both lexicons increases through the layers, ranging from 1.63% in the first layer to 12.27% in the last one. Despite this, we believe that the accuracy of the Spanish lexicon is really good, if we compare layers 5 and 6 (91.75% in a lexicon with 1779 terms and 86.09% in a lexicon with 2849 terms) to the Spanish lexicons by Pérez-Rosas et al. (2012) (90% in a lexicon with 1347 lemmas and 74% in a lexicon with 2496 lemmas). Moreover, the next layer of our resource still has a better accuracy (77.69%) than the biggest lexicon by Pérez-Rosas et al. (2012), with a much higher number of lemmas (6625).

4.3.1. ML-SENTICON structure

The lexicons are publicly available.[10] For layers 1–4 of the English and Spanish lexicons, the lemmas tagged as incorrect in the evaluation process have been deleted. A sample of the English lexicon is shown in Fig. 5. For each positive or negative lemma, we also include the POS tag (*pos*), the average of the polarities from synsets related to the lemma (*pol*) and the standard deviation of these values (*std*).

5. Conclusions

In this paper we have presented new lemma-level sentiment lexicons for English, Spanish and other three official languages in Spain. The lexicons are multilayered, allowing applications to choose different compromises between the amount of available words and the accuracy of the estimations of their prior polarities. For each lemma in the resource, we provide a real value representing the prior polarity, between −1 and 1, and a standard deviation reflecting the ambiguity of that value. According to our evaluations, the lexicons for English and Spanish have both high accuracies, over 90% for layers 1–6 and 1–5, respectively. In the case of the Spanish lexicon, the accuracy is sensibly better than the accuracy reported in other recent work (Pérez-Rosas et al., 2012).

As a previous step to the lemma-level lexicons, we have built a synset-level lexicon for English, based on the method used by Baccianella et al. (2010) to build SENTIWORDNET 3.0, one of the most used sentiment lexicons nowadays. This method comprises two steps, one involving the classification of individual synsets from WordNet as positive, negative or neutral, and another one involving a global, graph-based refination of the positivity and negativity scores of the synsets. We have introduced several improvements in both steps. In the first one, we have added a new source of information for training the classifiers, using WORDNET-AFFECT 1.1 (Strapparava et al., 2006). We have also applied a meta-learning scheme for combining multiple classifiers. In the second step of the method, we have proposed new kinds of WordNet-based graphs, and a different random-walk algorithm called POLARITYRANK (Cruz et al., 2012). We have evaluated the positivity and negativity scores obtained in each step, achieving significant improvements in both cases with respect to the original method.

Relating the practical implications on opinion mining, ML-SentiCon can be integrated in a wide variety of systems in which the sentiment polarity is a key element: polarity classification, product review mining, social media monitoring, etc. These systems can benefit of the specific characteristics of our lexicons: the ability to choose different compromises between volume and accuracy, the availability of languages with few sentiment resources (Spanish, Catalan, Basque and Galician), the inclusion of morphologically disambiguated lemmas and multi-words, and the low error rate (especially in the first layers). It is important to notice that the information contained in sentiment lexicons refers to the prior polarity of words, i.e. the positive or negative nature of words, regardless of context. In order to achieve better results in concrete applications, the prior polarity estimations may be adapted to the context, taking into account other nearby words (e.g. negations and other modifiers), the sentence as a whole or even document-level features as the topic concerned or the genre of the text.

A clear future research direction is the application of the method described in this paper to other languages. The method is fully applicable subject to the availability of a WordNet version in the target language. In this regard, a good resource is Open Multilingual WordNet (Bond & Foster, 2013), which provides access to WordNet versions in a variety of languages. Each version has been made by many different projects and varies greatly in size and

---

[10] http://www.lsi.us.es/~fermin/index.php/Datasets.

accuracy. Therefore, it would be desirable to conduct accuracy assessments of the sentiment lexicons obtained from them. Another interesting research direction is the integration of ML-SentiCon in sentiment analysis systems and the evaluation of the contributions to the task addressed. In this respect, it is easy to replicate results of those works based on SentiWordNet, but using ML-SentiCon instead. All works relying on word-level or lemma-level sentiment lexicons can also reproduced their experiments based on ML-SentiCon.

Acknowledgement

References

Abdul-Mageed, M., Diab, M. T., & Korayem, M. (2011). Subjectivity and sentiment analysis of modern standard arabic. In *ACL (Short Papers)* (pp. 587–591).

Agrawal, S. et al. (2009). Using syntactic and contextual information for sentiment polarity analysis. In *Proceedings of the second international conference on interaction sciences: Information technology, culture and human* (pp. 620–623). ACM.

Baccianella, S., Esuli, A., & Sebastiani, F. (2010). Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the seventh conference on international language resources and evaluation (Vol.10, pp. 2200–2204). ELRA.

Banea, C., Mihalcea, R., & Wiebe, J. (2008). A bootstrapping method for building subjectivity lexicons for languages with scarce resources. In *LREC* (pp. 2764–2767).

Bond, F., & Foster, R. (2013). Linking and extending an open multilingual wordnet. In *51st Annual meeting of the association for computational linguistics: ACL-2013* (pp. 1352–1362).

Brooke, J., Tofiloski, M., & Taboada, M. (2009). Cross-linguistic sentiment analysis: From english to spanish. In *Proceedings of the seventh international conference on recent advances in natural language processing* (pp. 50–54). Borovets, Bulgaria.

Clematide, S., & Klenner, M. (2010). Evaluation and extension of a polarity lexicon for german. In *Proceedings of the first workshop on computational approaches to subjectivity and sentiment analysis* (p. 7).

Cruz, F. L., Vallejo, C. G., Enríquez, F., & Troyano, J. A. (2012). Polarityrank: Finding an equilibrium between followers and contraries in a network. *Information Processing & Management, 48*(2), 271–282.

de Jong, S. (1993). Simpls: An alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems, 18,* 251–263.

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

Desmet, B., & Hoste, V. (2013). Emotion detection in suicide notes. *Expert Systems with Applications, 40*(16), 6351–6358.

Esuli, A., & Sebastiani, F. (2006). SentiWordNet: A publicly available lexical resource for opinion mining. In *Proceedings of language resources and evaluation (LREC)* (Vol. 6, pp. 417–422).

Esuli, A., & Sebastiani, F. (2007). Random-walk models of term semantics: An application to opinion-related properties. *Proceedings of LTC 2007, 221–225.

Fagin, R., Kumar, R., Mahdian, M., Sivakumar, D., & Vee, E. (2004). Comparing and aggregating rankings with ties. In *PODS '04: Proceedings of the twenty-third ACM SIGMOD-SIGACT-SIGART symposium on principles of database systems* (pp. 47–58). New York, NY, USA: ACM.

Fellbaum, C. (Ed.). (1998). *WordNet: An electronic lexical database.* MIT Press.

Gonzalez-Agirre, A., Laparra, E., & Rigau, G. (2012). Multilingual central repository version 3.0. In *LREC* (pp. 2525–2529).

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

Kaji, N., & Kitsuregawa, M. (2007). Building lexicon for sentiment analysis from massive collection of html documents. In *EMNLP-CoNLL* (pp. 1075–1083).

Kang, H., Yoo, S. J., & Han, D. (2012). Senti-lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews. *Expert Systems with Applications, 39*(5), 6000–6010.

Liu, B., Hu, M., & Cheng, J. (2005). Opinion observer: Analyzing and comparing opinions on the web. In *Proceedings of WWW* (pp. 342–351).

Liu, B., & Zhang, L. (2012). A survey of opinion mining and sentiment analysis. In C. C. Aggarwal & C. Zhai (Eds.), *Mining text data* (pp. 415–463). US: Springer.

Lu, B., Song, Y., Zhang, X., & Tsou, B. K. (2010). Learning chinese polarity lexicons by integration of graph models and morphological features. In *Information retrieval technology* (pp. 466–477). Springer.

Martín-Valdivia, M.-T., Martínez-Cámara, E., Perea-Ortega, J.-M., & Alfonso Ureña-López, L. (2012). Sentiment polarity detection in spanish reviews combining supervised and unsupervised approaches. *Expert Systems with Applications, 40*(10), 3934–3942.

Molina-González, M. D., Martínez-Cámara, E., Martín-Valdivia, M.-T., & Perea-Ortega, J. M. (2013). Semantic orientation for polarity classification in spanish reviews. *Expert Systems with Applications, 40*(18), 7250–7257.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1998). The pagerank citation ranking: Bringing order to the web. Tech. rep., Stanford Digital Library Technologies Project. <http://citeseer.ist.psu.edu/page98pagerank.html>.

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

Pérez-Rosas, V., Banea, C., & Mihalcea, R. (2012). Learning sentiment lexicons in spanish. In *LREC* (pp. 3077–3081).

Rao, D., & Ravichandran, D. (2009). Semi-supervised polarity lexicon induction. In *Proceedings of the 12th conference of the european chapter of the association for computational linguistics* (pp. 675–682). Association for Computational Linguistics.

Riloff, E., & Wiebe, J. (2003). Learning extraction patterns for subjective expressions. In *Proceedings of the conference on empirical methods in natural language processing (EMNLP)* (pp. 105–112).

Saggion, H., & Funk, A. (2010). Interpreting sentiwordnet for opinion classification. In *Proceedings of the seventh conference on international language resources and evaluation LREC10* (pp. 1129–1133).

Stone, P. J., Dunphy, D. C., & Smith, M. S. (1966). The general inquirer: A computer approach to content analysis. MIT Press.

Strapparava, C., Valitutti, A., & Stock, O. (2006). The affective weight of lexicon. In *Proceedings of the fifth international conference on language resources and evaluation* (pp. 423–426).

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

Turney, P. D., & Littman, M. L. (2003). Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems, 21,* 315–346.

Vossen, P. (1998). *EuroWordNet: A multilingual database with lexical semantic networks.* Boston: Kluwer Academic.

Wilson, T., Wiebe, J., & Hoffmann, P. (2005). 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 (HLT/EMNLP)* (pp. 347–354).

Xu, G., Meng, X., & Wang, H. (2010). In *Proceedings of the 23rd international conference on computational linguistics* (pp. 1209–1217). Association for Computational Linguistics.