# LyS at TASS 2013: Analysing Spanish tweets by means of dependency parsing, semantic-oriented lexicons and psychometric word-properties

***LyS en TASS 2013: Analizando tuits en castellano a través de análisis de dependencias, lexicones de opiniones y propiedades psicométricas del lenguaje***

**David Vilares, Miguel A. Alonso** y **Carlos Gómez-Rodríguez**
Departamento de Computación, Universidade da Coruña
Campus de Elviña, 15071 A Coruña
{david.vilares, miguel.alonso, carlos.gomez}@udc.es

**Resumen:** Este artículo describe el enfoque utilizado por nuestro grupo para resolver las tareas de análisis global del sentimiento, identificación de tópicos y clasificación de la tendencia política sobre tuits en español; propuestas en el Taller de Análisis del Sentimiento en la SEPLN (TASS 2013). Como paso previo, se realiza un preprocesado *ad-hoc* para normalizar los tuits. A continuación, se lleva a cabo un análisis morfológico de los tuits para luego obtener su estructura sintáctica aplicando algoritmos de análisis de dependencias. Nuestra propuesta también emplea recursos psicológicos, que permiten explotar las características psicométricas del lenguaje humano. Los resultados experimentales confirman la robustez de la propuesta, que en términos generales ha obtenido un buen rendimiento, alcanzando el primer puesto en la tarea de clasificación de tópicos.
**Palabras clave:** Twitter, Análisis del sentimiento, Clasificación de tópicos, Tendencias políticas, Análisis sintáctico de dependencias, Dimensiones del lenguage

**Abstract:** This article describes the approach developed by our group in order to resolve the sentiment analysis at a global level, topic identification and political tendency classification tasks on Spanish tweets; proposed at the Workshop of Sentiment Analysis at SEPLN (TASS 2013). As a preliminary step, we carry out an *ad-hoc* preprocessing in order to normalise the tweets. We then apply part-of-speech tagging and dependency parsing algorithms to the tweets to obtain their syntactic structure. Our proposal also employs psychological resources in order to exploit the psychometric properties of human language. The experimental results confirm the robustness of the proposal, which has achieved good performance in general, being the best-performing approach in the topic classification task.
**Keywords:** Twitter, Sentiment Analysis, Topic Classification, Political Leanings, Dependency Parsing, Language Dimensions

## 1 Introduction

The analysis and comprehension of user reviews have always been a key asset when making effective decisions. A few years ago, it was difficult to obtain broad and reliable information about products, services or other issues: there was no place where one could retrieve and analyse a large amount of reviews, and surveys usually were a costly and very limited solution. With the appearance of the Web 2.0, and especially the rise of social media, millions of users express and share their opinions in these sites, making it possible to obtain an overview on virtually any topic. This is useful for individuals, but even more so for companies and institutions, which monitor social networks, such as Facebook or Twitter, to poll their sphere of influence, plan their strategy and make decisions.

In this context, *sentiment analysis* (SA), also known as *opinion mining* (OM), has become a growing field of research, in both academia and industry, which focusses on developing techniques to automatically analyse subjective content in texts. The Workshop of Sentiment Analysis at SEPLN (TASS 2013)[1] proposes four tasks related to this area: per-

[1]http://www.daedalus.es/TASS2013/about.php

forming sentiment analysis to determine the global polarity of tweets, creating a classifier able to identify the topic (or topics) of the tweets, building an approach to categorise the polarity of the different entities that appear in a message; and determining the political tendency of different users. We took part in all the tasks except polarity classification at the entity level,[2] and we employed a similar approach for those three activities. We combine unsupervised and supervised techniques, taking into account lexical, syntactic, semantic and psychometric information. To optimise the performance, we also apply domain adaptation to Twitter.

The remainder of the paper is organised as follows. Section 2 reviews related research on sentiment analysis. In Section 3, we describe the applications and resources used to address the workshop tasks. Section 4 briefly outlines each of the proposed activities and explains the approaches adopted to treat them. Experimental results are shown in Section 5. Finally, we present the conclusions and future work in Section 6.

## *2 Related research*

In the last decade, SA has become one of the major technological challenges in the field of Natural Language Processing (NLP), given its applications in social, market research and business intelligence areas. One of the basic tasks in sentiment analysis is classifying the polarity of texts, which focusses on determining the attitude of an author of a review as positive, mixed or negative. This challenge has been tackled mainly from two different angles: semantic-based (Turney, 2002) and supervised (Pang, Lee, and Vaithyanathan, 2002). Semantic approaches are characterised by the use of semantic orientation (SO) dictionaries or opinion lexicons. On the other hand, machine learning (ML) solutions involve building classifiers from a collection of annotated texts, where each text is usually represented as a bag-of-words.

These two main approaches commonly take linguistic knowledge into account in order to optimise their performance. Turney (2002) uses a part-of-speech (PoS) tagger to identify subjective phrases (adjectives and adverbs) in reviews to then estimate their polarity, based on the SO of the phrases. Taboada et al. (2011) also use morphological tags to identify and treat relevant phenomena, such as intensification or negation, in SA in an unsupervised way. With the same aim, there are also studies that investigate the utility of parsing in this area (Joshi and Penstein-Rosé, 2009; Nakagawa, Inui, and Kurohashi, 2010), showing that opinion mining performance can be improved by incorporating syntactic knowledge. A general problem of polarity classification is that this task becomes harder when the topic is more abstract (Turney, 2002). In this respect, topic classification is another NLP challenge by itself, which is usually tackled by exploiting the content of the message to detect its subject, although other approaches follow different methods, such as using the URL of reviews (Baykan et al., 2009).

But sentiment analysis not only involves polarity classification. Other related tasks have gained importance in recent years, such as identification of political leanings. In this respect, Dalvean (2013) classifies Australian MPs into the two main political parties by means of their speech. Mullen and Malouf (2006) present a preliminary study to classify informal political discourse. They conclude that OM using traditional word-based classification is not adequate to perform effective political discourse analysis.

## *3 Text analytics tools*

In order to carry out the different tasks proposed in TASS 2013, we employ several resources and applications, described below.

### 3.1 Dependency Parsing

We rely on MaltParser (Nivre et al., 2007) for analysing the syntactic structure of each tweet. As a first step, we run a PoS-tagger to then obtain the syntactic structure of the tweets by means of dependency parsing algorithms. As a result, we obtain a *dependency tree* for each sentence, consisting of a set of *head/dependent* binary relations, called *dependencies*, between words. Each dependency has a label with a given *dependency type*, which denotes the existing syntactic relation between head and dependent.

In this way, we are able to extract at the present time the following information: the number of occurrences of each PoS-tag, each dependency type and each dependency triplet for each tweet. As we detail in Section

[2]Due to time restrictions, we could not submit results for this task.

4, all this linguistic knowledge is used to perform the sentiment analysis tasks proposed in the workshop.

### 3.2 The syntactic SO analyser

We apply an approach presented in (Vilares, Alonso, and Gómez-Rodríguez, 2013b) to estimate the global SO of tweets from the dependency tree obtained by our parser. To calculate the SO for each individual term, we rely on the opinion lexicon of Brooke, Tofiloski, and Taboada (2009). This lexicon is a collection of dictionaries of subjective common nouns, adjectives, adverbs and verbs where each word is annotated with its SO, between -5 (the most negative) and +5 (the most positive). The SO corresponds to a generic assignment, without considering a specific domain. It also provides a dictionary of intensifiers, where the label assigned to each intensifying expression represents the value (positive or negative) of its modification.

The dependency tree of each sentence is used to address some relevant linguistic constructions for polarity classification: *intensification*, *subordinate adversative clauses* and *negation*. We identify the scope of influence of these phenomena in the sentence defining a set of syntactic-based rules which modify the SO of the corresponding fragment of the sentence, according to human language intuition.

In this manner, the SO analyser returns three features for each tweet: the global SO, the number of positive words and the number of negative ones.

### 3.3 Psychometric properties

Linguistic Inquiry and Word Count (LIWC) is a software presented in (Pennebaker, Francis, and Booth, 2001); which identifies emotions, causal words and psychometric properties present in different types of texts, by means of dictionaries. It includes a dictionary for Spanish (Ramírez-Esparza et al., 2007) that distinguishes between different *psychological aspects* of the human language (cognition mechanisms, anxiety, sexuality, ...), but also considering *topics* (TV, family, religion, ...) or even *linguistic information* (past, present and future tense, exclamations, questions, ...).

We hypothesise these psychometric features can be helpful to perform effective sentiment analysis, as we explain in Section 4.

### 3.4 WEKA

The training of the models used for the tasks relies on the WEKA data mining software (Hall et al., 2009). Concretely, we chose an SMO, an implementation of a support vector machine (SVM) presented in (Platt, 1999). The attribute selection tools provided by this software are used to perform domain adaptation and noise reduction, given the number of features that we are extracting (PoS-tags, dependencies, dependency types, the psychometric properties extracted from the LIWC and the text of each tweet represented as a bag of words). We rank the features by measuring the information gain with respect to the class and we establish a minimum threshold in an empirical way for each task, as we detail in Section 4.

## *4 Description of the tasks*

The Workshop on Sentiment Analysis at SEPLN (TASS 2013) proposed four tasks:

1. *Sentiment Analysis at a global level*: it consists of automatic classifying the polarity of tweets.
2. *Topic classification*: it focusses on identifying the subject of tweets.
3. *Sentiment Analysis at entity level*: it has a similar aim to task 1, but determining the polarity for the different entities mentioned in the tweets.
4. *Political tendency identification*: this task consists of classifying the political discourse of users given their set of tweets.

Due to time constraints, we have only taken part in tasks 1, 2 and 4. A description of the approaches adopted for each activity is provided in the following subsections. To evaluate and train the models, TASS 2013 provides two corpora:

- *General corpus*: It is a collection of Spanish tweets written by public figures that is composed of a training and a test set which contain 7,219 and 60,798 tweets, respectively. Each one is annotated with one of these six categories: *strong positive* (P+), *positive* (P), *neutral* (NEU), *negative* (N), *strong negative* (N+) or *without opinion* (NONE). In addition, each tweet is annotated with a set of topics. The corpus distinguishes of up ten topics: *film*, *football*,[3] *economics*, *entertainment*, *literature*, *music*, *politics*, *sports*, *technology* and *other*.

- *Politics corpus*: It contains 2,500 tweets posted during the electoral campaign of the 2011 Spanish general election, which mention one of the four main national political parties. This corpus was only created for task 3, so we did not used it in our experiments.

### 4.1 Sentiment analysis at global level

This task consists of identifying the global polarity of a tweet according to two different criteria: classification into six categories (P+, P, NEU, N, N+ and NONE) and classification into four categories (the classes P+ and N+ are included in the classes P and N, respectively).

Our strategy to resolve this task is as follows. We combine the lexical, syntactic, semantic and psychometric features provided by our analytics tools; ranking them according to their information gain and selecting those with gain at least 0.001.[4] We then use the selected features to train an SMO classifier. A more detailed description about how our approach addresses some of these functionalities is shown in Vilares, Alonso, and Gómez-Rodríguez (2013a). Table 1 shows the best selected discriminating terms, taking into account their information gain with respect to the class.

Moreover, we trained a second model, which includes as features some information about the author of the tweet. In particular, we used the user screen name (*e.g.* the screen name of *'@twitter'* is *'twitter'*) and the user type (*journalist*, *politician* or *public figure*).[5]

### 4.2 Topic classification

As explained previously, this task focusses on identifying the topic (or the topics) of tweets. Since a message can cover different topics, we carry out a *one vs all* strategy to be able to perform multinomial classification. In this manner, we trained ten SMO classifiers, where each one distinguishes a topic from the others (*e.g.* economics vs other, politics vs other, etc). We applied oversampling in order to balance the categories. To create each classifier, we follow a similar approach as in the *Sentiment analysis at a global level* task. We first rank the features provided by the dependency trees, the LIWC dictionaries and the tweet itself represented as a bag of words; by measuring the information gain with respect to the topic. Table 2 and 3 shows the most discriminating features of the classifiers *film vs other* and *sports vs other*, respectively. Experiments suggested that the best topic classification models are obtained selecting the features with an information gain greater than 0.

| Ranking | Feature | Category |
|---|---|---|
| 1 | SO | SEMANTIC ORIENTATION |
| 2 | positive emotion | LIWC DIMENSION |
| 3 | affective | LIWC DIMENSION |

Table 1: Ranking of the best polarity discriminating features

| Ranking | Feature | Category |
|---|---|---|
| 1 | proper name | FINE POSTAG |
| 2 | job | LIWC DIMENSION |
| 3 | 'película'(film) | LEMMA |

Table 2: Ranking of the best discriminating terms for *film vs other*

| Ranking | Feature | Category |
|---|---|---|
| 1 | sports | LIWC DIMENSION |
| 2 | pleasure | LIWC DIMENSION |
| 3 | 'nadal' | LEMMA |

Table 3: Ranking of the best discriminating terms for *sports vs other*

As we did in the global sentiment analysis task, we also trained an alternative model which includes the user screen name and the user type as features for a supervised classifier. We hypothesise that a Twitter user is

[3]This category refers to association football, also known as soccer, which is the most popular sport in Spain.

[4]The information gain threshold to the global sentiment task was empirically established testing the following values: 1,0.1,0.001, 0.0001, 0.00005, 0

[5]This information is provided by the TASS 2013 organisation.

more likely to write about only a few topics. In the same line, we think that some user types, such as *politician*, can help determine what a user tweets about.

### 4.3 Political tendency identification

The goal of this task is to identify the political leaning of an user (LEFT, RIGHT, CENTRE or UNDEFINED) by analyzing her tweets. To do so, we used the same machine learning setup as in the global sentiment classification task, except that we now classify sets of tweets authored by a given user, rather than individual tweets. However, as the provided corpus is not annotated with political tendencies, we had to build our own annotated training corpus for this task.

To obtain such a corpus, we downloaded tweets from the official Twitter user accounts of the four most-voted nationwide political parties (PP, PSOE, IU and UPyD), as well as those of prominent politicians from those parties. In particular, we obtained 27,367 tweets from 16 user accounts officially associated with the PP, 28,180 tweets from 11 users linked to the PSOE, 28,418 tweets from 9 IU users, and 18,953 tweets from 6 UPyD users.

The downloaded tweets were from dates ranging from April 2012 to June 2013. We believe that it would have been better to train on tweets from the same dates as the target tweets to analyze, as it has been noted in the literature that political language changes according to whether a party is in power or in the opposition, affecting classification tasks (Hirst, Riabinin, and Graham, 2010). However, this was not possible due to the limits imposed by the Twitter API, which only allows recent tweets to be downloaded.

To annotate the users with their tendencies, it is worth noting that the ascription of parties to political categories such as "left" or "right" is a controversial matter, with no clear social consensus as to where some parties lie on the left-right spectrum. For instance, this is reflected in the discrepancy between the positions perceived by citizens in polls and those officially stated in each party's statutes and manifestos: while the PSOE defines itself as a left-wing party, the citizens polled in the latest CIS Autonomic Barometer (CIS, 2012) located it in the centre. Therefore, we decided to train two different models, based on different criteria: Model 1 classifies the parties according to the results of the CIS Autonomic Barometer (placing IU on the left, PP on the right, and both PSOE and UPyD in the centre) while Model 2 prioritizes each party's official allegiance (with IU and PSOE on the left, PP on the right, and UPyD in the centre[6]). For the UNDEFINED category, we used accounts from news outlets that report news without commentary.

Since the user accounts from which we downloaded would constitute a too small training corpus if we used each of them directly as a training instance, we instead built training instances by taking groups of 200 tweets from each user.[7] To further enlarge the training set, we also generated synthetic data using interpolation, mixing tweets from different user accounts with the same leaning to create artificial accounts.

## 5 *Empirical results*

To develop our approaches, we splitted the training set of the general corpus, and we did the same for the our ad-hoc political tendency corpus. We used the 80% of each corpus as our training set and the remaining 20% as the development set.

Table 4 shows the empirical results for the sentiment analysis at a global level task in our development set.[8] With respect to the four-category classification task, the performance for positive, negative and none tweets seems to be consistent, but the same is not true for neutral tweets. We hypothesise this is due to the lack of neutral tweets in the training set. We tried to solve the problem applying oversampling, but experimental results were not satisfactory. The same tendency was observed at the six-category classification task. We also presented a second model, which considers specific user information as features, but we obtained no improvement in empirical results.

[6]While the PP declares itself a "reformist centre" party, we still locate it on the right because it is still the rightmost party among the four considered. On the other hand, UPyD does not literally define itself as centre, but as a cross-ideology party, but we believe centre is the closest among the four categories considered.

[7]This amount of tweets is close to the median number of tweets per user in the TASS training corpus.

[8]We trained two different models to solve the classification with four and six categories in the development set, although in the test set, the TASS organisation used the results obtained with six categories to evaluate both approaches.

| Measure | 4 classes | 6 classes |
|---|---|---|
| $F_{p+}$ | - | 0.596 |
| $F_p$ | 0.698 | 0.244 |
| $F_{neu}$ | 0.124 | 0.213 |
| $F_n$ | 0.633 | 0.428 |
| $F_{n+}$ | - | 0.399 |
| $F_{none}$ | 0.584 | 0.585 |
| *Accuracy* | 0.619 | 0.463 |

Table 4: Global sentiment task: Results (F-scores) on our TASS 2013 development set

Table 5 shows the results on the test set. Due to the number of experiments submitted by the participants, we only include the best run for each group.

| Group | Average F. 6 cat | Average F. 4 cat |
|---|---|---|
| DLSI-UA | 0.616 | 0.663 |
| ELHUYAR | 0.601 | 0.686 |
| UPV | 0.576 | 0.684 |
| CITIUS-USC | 0.558 | 0.668 |
| | **0.553** | **0.657** |
| JRC | 0.519 | 0.612 |
| ITA | 0.439 | 0.543 |
| UNED-LSI | 0.402 | 0.479 |
| UNED-JRM | 0.393 | 0.496 |
| TECNALIA-UNED | 0.348 | 0.496 |
| ETH-ZURICH | 0.328 | 0.466 |
| SINAI-EMML | 0.314 | 0.409 |
| SINAI-CESA | 0.135 | 0.389 |

Table 5: Global sentiment task: Results (F-scores) on the TASS 2013 test general set

The performance of our topic classification approach on the development set is shown in Table 6, both without (basic model) and with user information features. As expected, the best performance was obtained for the predominant classes in the original training set such as politics, entertainment or others. The Hamming loss distance, the label-based accuracy and the exact match are calculated according to equations 1, 2 and 3 where: $L$ is the number of different labels, $D$ is the number of instances, $Y_i$ are the labels expected for an instance $i$ and $Z_i$ are the labels predicted for an instance $i$ :

$$Hamming\ loss = \frac{1}{|D|}\sum_{i=1}^{|D|}\frac{Y_i \triangle Z_i}{L} \quad (1)$$

$$Label\text{–}based\ accuracy = \frac{1}{|D|}\sum_{i=1}^{|D|}\frac{Y_i \cap Z_i}{Y_i \cup Z_i} \quad (2)$$

$$Exact\ match = \frac{\#instances\ exactly\ labelled}{\#instances} \quad (3)$$

The model that considers specific user information seems to improve the performance over the basic model. Although we have also sent this second model to the workshop, we think it would not be applicable in a real-life environment. The user type is not provided by anny Twitter API method and the user name would be not effective, due to the impossibility of training models with tweets from every user.

| Measure | Basic Model | Model with user info |
|---|---|---|
| $F_{film}$ | 0.208 | 0.265 |
| $F_{politics}$ | 0.697 | 0.707 |
| $F_{technology}$ | 0.085 | 0.045 |
| $F_{entertainment}$ | 0.458 | 0.513 |
| $F_{sports}$ | 0.250 | 0.258 |
| $F_{other}$ | 0.532 | 0.532 |
| $F_{economy}$ | 0.413 | 0.437 |
| $F_{music}$ | 0.447 | 0.475 |
| $F_{football}$ | 0.290 | 0.472 |
| $F_{literature}$ | 0.412 | 0.378 |
| *Hamming loss* | 0.101 | 0.090 |
| *Label-based acc.* | 0.563 | 0.599 |
| *Exact match* | 0.385 | 0.435 |

Table 6: Topic classification task: Results (F-scores) on our TASS 2013 development set

Table 7 shows the results obtained on the test set on the topic classification task, where our system obtained the best score in the workshop. We hypothesise that the LIWC dictionaries were specially useful, because they allowed the model to identify words that refer to specific topics, such as politics or sports.

| Group | Average F. |
|---|---|
| | **0.804** |
| UPV | 0.756 |
| FHC25-IMDEA | 0.710 |
| ETH-ZURICH | 0.562 |
| UNED-LSI | 0.501 |
| UNED-JRM | 0.479 |
| SINAI-CESA | 0.160 |

Table 7: Topic classification task: Results (F-scores) on the TASS 2013 general test set

The results for the political tendency identification task on our development set can be seen on Table 8. We obtained an accuracy of 1, meaning that the development set

| Measure | Model 1 | Model 2 |
|---|---|---|
| $F_{left}$ | 1.000 | 1.000 |
| $F_{centre}$ | 1.000 | 1.000 |
| $F_{right}$ | 1.000 | 1.000 |
| $F_{undefined}$ | 1.000 | 1.000 |
| *Accuracy* | 1.000 | 1.000 |

Table 8: Political tendency identification task: Results (F-scores) on our TASS 2013 development set

was easy to classify. This is not surprising, given its small size and the fact that it is made with official tweets from political parties, which will obviously have a clear-cut ideological content. Table 9 shows the results on the test set. We obtained the last position in terms of F-measure, but the third place in terms of recall. We hypothesise this is due to following a radically different strategy that the proposed by the TASS organisation, and to the complications classifying the undefined users.

| Group | Average F. |
|---|---|
| ETH-ZURICH (manual) | 0.734 |
| UPV | 0.703 |
| SINAI-CESA | 0.474 |
| | **0.424** |

Table 9: Political tendency identification task: Results (F-scores) on the TASS 2013 general test set

## 6 Conclusions and future work

This paper tests the effectiveness of employing linguistic features to resolve three classification tasks proposed at the Workshop on Sentiment Analysis at SEPLN (TASS 2013): categorising the global polarity of the tweets, identifying their topics and determining the political tendency of a user, given their messages. We use PoS-tag information, dependency relations between words with semantic information provided by opinion lexicons and psychological knowledge extracted from LIWC dictionaries. We combine these features with a pure machine learning approach which takes the tweets as a bag a words. We adopted a similar approach to carry out the three tasks, although there are some particularities. With respect to the topic classification task, we did not use our syntactic SO analyser nor opinion lexicons, and we followed a *one vs all* strategy in order to be able to apply multi-topic classification. To perform the political tendency identification task, we semi-automatically created a corpus of tweets from politicians, political parties and newspapers (to identify when an user has no political tendency), and we grouped each user's tweets into a single document in order to try to directly classify their tendency. In general terms, the evaluation on both development and test sets reinforce the robustness and the generality of our approach, which was the best-performing system in the topic classification task, and scored a performance very near that of the best systems on the global polarity classification challenge.

As future work, there is room for improvement. We plan to implement exhaustive tweet normalization, in order to improve the accuracy of our tools (the tagger, parser and syntactic SO analyser). We would also like to explore how to adapt dependency parsing to micro texts, especially tweets. Finally, we want to focus on a more complex treatment of dependencies. The employment of composite back-off dependencies (Joshi and Penstein-Rosé, 2009), or the introduction of semantic dependencies (*e.g.* combining the semantic and psychological knowledge provided by the words with their syntactic relations) are some of our short-term aims.


## *Acknowledgments*

Research reported in this article has been partially funded by Ministerio de Economía y Competitividad and FEDER (Grant TIN2010-18552-C03-02) and by Xunta de Galicia (Grants CN2012/008, CN2012/319).


## *Bibliografía*


Baykan, Eda, Monika Henzinger, Ludmila Marian, and Ingmar Weber. 2009. Purely url-based topic classification. In *Proceedings of the 18th international conference on World wide web*, WWW '09, pages 1109–1110, New York, NY, USA. ACM.

Brooke, J., M. Tofiloski, and M. Taboada. 2009. Cross-Linguistic Sentiment Analysis: From English to Spanish. In *Proceedings of International Conference on Recent Advances in Natural Language Processing (RANLP)*, pages 50–54, Borovets, Bulgaria.

CIS. 2012. Barómetro autonómico (iii). Technical Report 2956, Centro de Investigaciones Sociológicas, September-October.

Dalvean, Michael Coleman. 2013. Who's who on the australian political left-right spectrum? Working/technical paper, Australian National University, Canberra. Open Access Research http://hdl.handle.net/1885/9755.

Hall, M., E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I.H. Witten. 2009. The WEKA data mining software: an update. *SIGKDD Explor. Newsl.*, 11(1):10–18, November.

Hirst, Graeme, Yaroslav Riabinin, and Jory Graham. 2010. Party status as a confound in the automatic classification of political speech by ideology. In *Proceedings, 10th International Conference on Statistical Analysis of Textual Data / 10es Journées internationales d'Analyse statistique des Données Textuelles (JADT 2010)*, pages 731–742, Rome, June.

Joshi, M. and C. Penstein-Rosé. 2009. Generalizing dependency features for opinion mining. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, ACLShort '09, pages 313–316, Suntec, Singapore. Association for Computational Linguistics.

Mullen, Tony and Robert Malouf. 2006. A preliminary investigation into sentiment analysis of informal political discourse. In *AAAI 2006 Spring Symposium on Computational Approaches to Analysing Weblogs (AAAI-CAAW 2006)*.

Nakagawa, T., K. Inui, and S. Kurohashi. 2010. Dependency tree-based sentiment classification using CRFs with hidden variables. In *NAACL HLT'10 Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics. Proceedings of the Main Conference*, HLT '10, pages 786–794, Stroudsburg, PA, USA. Association for Computational Linguistics.

Nivre, J., J. Hall, J. Nilsson, A. Chanev, G. Eryigit, S. Kübler, S. Marinov, and E. Marsi. 2007. Maltparser: A language-independent system for data-driven dependency parsing. *Natural Language Engineering*, 13(2):95–135.

Pang, B., L. Lee, and S. Vaithyanathan. 2002. Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing - Volume 10*, EMNLP '02, pages 79–86, Stroudsburg, PA, USA. Association for Computational Linguistics.

Pennebaker, J.W., M.E. Francis, and R.J. Booth. 2001. Linguistic inquiry and word count: Liwc 2001. *Mahway: Lawrence Erlbaum Associates*, page 71.

Platt, J.C. 1999. Advances in kernel methods. MIT Press, Cambridge, MA, USA, chapter Fast training of support vector machines using sequential minimal optimization, pages 185–208.

Ramírez-Esparza, N., J.W. Pennebaker, F.A. García, and R. Suria. 2007. La psicología del uso de las palabras: Un programa de computadora que analiza textos en Español (The psychology of word use: A computer program that analyzes texts in Spanish). *Revista Mexicana de Psicología*, 24:85–99.

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

Turney, P. D. 2002. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, ACL '02, pages 417–424, Stroudsburg, PA, USA. Association for Computational Linguistics.

Vilares, D., M. A. Alonso, and C. Gómez-Rodríguez. 2013a. Supervised polarity classification of spanish tweets based on linguistic knowledge. In *Proceedings of the 2013 ACM symposium on Document engineering*, DocEng '13, New York, NY, USA. ACM.

Vilares, D., M. A. Alonso, and C. Gómez-Rodríguez. 2013b. A syntactic approach for opinion mining on spanish reviews. *Natural Language Engineering*, FirstView:1–25, 8.