LyS at TASS 2014: A Prototype for Extracting and Analysing Aspects from Spanish tweets *

LyS en TASS 2014: Un prototipo para la extracción y análisis de aspectos en tuits.

David Vilares, Yerai Doval, Miguel A. Alonso and Carlos Gómez-Rodríguez
Grupo LyS, Departamento de Computación, Campus de A Coruña s/n
Universidade da Coruña, 15071, A Coruña, Spain
{david.vilares, yerai.doval, miguel.alonso, carlos.gomez}@udc.es

Resumen: Este artículo describe nuestra participación en la tercera edición del taller de análisis del sentimiento de tuits escritos en castellano, el TASS 2014. En la evaluación competitiva de este año, se han propuesto cuatro retos: (1) análisis del sentimiento a nivel global, (2) clasificación de tópicos, (3) extracción de aspectos y (4) análisis del sentimiento a nivel aspectual. Para las tareas 1 y 2 empleamos una aproximación basada en aprendizaje automático, donde distintos recursos lingüísticos e información extraída del conjunto de entrenamiento son utilizados para entrenar un clasificador supervisado. Para abordar la tarea 3, nuestra aproximación recolecta una lista de representaciones que es empleada para identificar los aspectos requeridos por los organizadores. Por último, la tarea 4 delega en heurísticas para identificar el alcance de cada aspecto, para después determinar su sentimiento a través de un clasificador supervisado. Los resultados experimentales son prometedores y nos servirán para desarrollar técnicas más complejas en el futuro.
Palabras clave: Análisis del sentimiento, Clasificación de tópicos, Extracción de aspectos, Análisis del sentimiento a nivel aspectual.

Abstract: This paper describes our participation at the third edition of the workshop on Sentiment Analysis focused on Spanish tweets, TASS 2014. This year's evaluation campaign includes four challenges: (1) global sentiment analysis, (2) topic classification, (3) aspect-extraction and (4) aspect-based sentiment analysis. Tasks 1 and 2 are addressed from a machine learning approach, using several linguistic resources and other information extracted from the training corpus to feed to a supervised classifier. With respect to task 3, we develop a naive approach, collecting a set of representations to identify the predefined aspects requested by the organisers. Finally, task 4 uses heuristics to identify the scope of each aspect, to then classify their sentiment via a supervised classifier. The experimental results are promising and will serve us as the starting point to develop more complex techniques.
Keywords: Sentiment Analysis, Topic-classification, Aspect-extraction, Aspect-based Sentiment Analysis.

1 Introduction

In the age of the Web 2.0, many companies and organisations have the need to learn from the data shared by users in this medium. Among the main practical applications of discovering and understanding knowledge from the opinions published on these sites are: measuring the perception of the public with respect to a product, service or event; and identifying its strengths and weaknesses, allowing to make better marketing and business decisions. One of the most interesting spaces for monitoring trends is Twitter, especially for real-time purposes, given the number of active users and messages published each day. However, it is also one of the most challenging ones. Users are restricted to express their views in messages of up to 140 characters. This requires to build precise systems, since arguments are limited to one or

* Research reported in this paper has been partially funded by Ministerio de Educación, Cultura y Deporte (FPU13/01180) and Xunta de Galicia (Grant CN2012/008).

two sentences. Moreover, Twitter presents a distinctive jargon, including elements such as hashtags, usernames or special tags. Since a manual monitoring of the content generated by the users in the website is not viable, industry and academia have become interested in designing and developing automatic techniques to solve this new challenge.

In this respect, *sentiment analysis* (SA) is the field of research focused on the automatic processing of subjective information, such as that shared by users in social networks. Most of the related work on this area is focused on English, since it is the predominant language on the Internet. However, languages such as Spanish are also playing an important role. The aim of the workshop on sentiment analysis focused on Spanish language, TASS 2014, is to provide a standard framework for creating and evaluating systems. Held since 2012, the participants of this third edition are encouraged to participate in up to four tasks. Two of them are legacy tasks: (1) *sentiment analysis at global level* and (2) *topic classification*. In addition, this year's edition proposes two new challenges: (3) *aspect extraction* and (4) *aspect-based sentiment analysis*. To evaluate them, the organisation 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 ten topics: *film*, *football*,[1] *economics*, *entertainment*, *literature*, *music*, *politics*, *sports*, *technology* and *other*. This collection was used to carry out both tasks 1 and 2.

- *Balanced general corpus*: A test subset containing 1 000 tweets with a similar distribution to the training corpus. It is used for an alternate evaluation of the performance of systems.

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

- *Social-TV corpus:* It is composed of a training and a test set of 1 773 and 1 000 tweets, respectively. They all refer to the 2014 Final of the Copa del Rey, a Spanish soccer championship. Each tweet is labelled with the aspects that occur in it as well as their corresponding sentiment. For example, in the sentence *'The team is good, but the coach is really bad'*, *'team'* would be labelled as positive and *'coach'* as negative. The organisation only takes into account a set of predefined aspects, including: players (*e.g. Lionel Messi*, *Cristiano Ronaldo* or *Iker Casillas*), three teams (*Barcelona*, *Real Madrid* and *Atlético de Madrid*) and some miscellaneous aspects (general aspects such as *team*, *player* or *referee*).

2 Related work

Many researchers had evaluated the performance of a variety of approaches, especially focused on sentiment analysis at global level, on the TASS corpora. Saralegi and San Vicente (2013) present a supervised learning approach, employing lexicons and linguistic information to address the problem. Pla and Hurtado (2013) follow a similar strategy, remarking the need of having specific tools for tokenising tweets. The contribution of Fernández et al. (2013) consists of combining two different approaches: a version of a ranking algorithm (RA-SR) and a new proposal using a skipgram scorer, which are used to create sentiment lexicons able to retain the context of the terms. Balahur and Perea-Ortega (2013) also tackle the challenge of sentiment analysis at the tweet level from a machine learning approach initially designed for English, combining and testing different sets of features. A different strategy is proposed by Martínez-Cámara et al. (2013), where an unsupervised approach is used to obtain interesting findings which are used to improve the performance of a supervised model used by this team at the first edition of TASS. Montejo-Ráez, Díaz Galiano, and García-Vega (2013) try to solve the problem from an Information Retrieval angle, by applying Latent Semantic Analysis (LSA), although they concluded results are not promising in comparison to machine learning techniques.

3 NLP pipeline

To address this evaluation campaign we rely on a natural language processing (NLP) pipeline as a starting point, which includes steps such as:

Preprocessing: The way people express themselves in web environments, in addition to the Twitter jargon, makes it necessary to adapt NLP tools to these messages. In recent years, some resources have appeared for tokenising and tagging English tweets (Gimpel et al., 2011). However, there is a lack of this kind of tools for Spanish. Therefore, we carry out an *ad-hoc* preprocessing of tweets in order to be able to apply, in an appropriate manner, standard NLP tools. This includes a normalisation of Twitter usernames, hashtags, interjections or emoticons. A detailed description of our preprocessing algorithm is available in Vilares, Alonso, and Gómez-Rodríguez (2014b).
Part-of-speech tagging: To the best of our knowledge, there exists no Twitter corpus annotated with part-of-speech (PoS) tags for Spanish language. This means that we cannot train a tagger specifically built for this social network. Instead, we rely on an implementation based on the Brill (1992) tagger provided by the NLTK[2] framework, using the Ancora corpus (Taulé, Martí, and Recasens, 2008) as our training set.
Dependency parsing: Let be a sentence S of the form $w_1w_2...w_{n-1}w_n$, where w_i indicates the *ith* word within the sentence; the result of applying a dependency parsing algorithm is a graph $G = \{(w_i, arc_{ij}, w_j)\}$, where w_i and w_j are the *head* and the *dependent* terms, and arc_{ij} is the existing syntactic relation between those terms, known as *dependency type*. We rely on MaltParser (Nivre et al., 2007) and the Ancora corpus to build the parser which will serve us to obtain the syntactic structure of the tweets.

4 Task 1: Sentiment Analysis at the tweet level

The task focuses on classifying the sentiment of a tweet into six (P+, P, NEU, N, N+ and NONE) and four (P, N, NEU, NONE) polarities. This year, our aim was to evaluate different sets of features in order to measure their impact on SA at the tweet level. Our approach is constrained, since we only use the official training set as the sole labelled corpus. More specifically, we use the training corpus to take some linguistic information provided by our NLP pipeline. As features, we tested: a bag-of-words, as well as lemmas or part-of-speech tags. We also include several lexicons available for Spanish language, in order to measure their effectiveness on the evaluation of micro-texts.

- *Ramírez-Esparza et al. (2007)*: It relates terms with psychometric properties (*e.g. anger* or *anxiety*) and topics (*e.g. football*, *sports* or *jobs*). The number of words that appear in a tweet referring to each property are used as features for our supervised classifier.

- *Saralegi and San Vicente (2013)*: This manually-built lexicon includes a collection of *positive* and *negative* words. It classifies terms into different classes: *general terms*, *terms extracted from Twitter*, *tourism domain terms*, *colloquialims* and *interjections*. We consider the number of positive and negative words of each class that appear in a tweet as features to feed to a supervised classifier.

- *Brooke, Tofiloski, and Taboada (2009)*: They provide a dictionary of polarity words based on their part-of-speech tag distinguishing between: nouns, adjectives, verbs and adverbs. Each word is labelled with a semantic orientation (SO) from -5 to 5. We again consider the number of words with a specific PoS-tag and SO, as features for the supervised classifier.

- *LYSA Twitter lexicon v-0.1*: We introduce a new automatically-built lexicon. The goal is to obtain a list of usual subjective terms employed by Spanish users in Twitter. We took the technique successfully applied by Mohammad, Kiritchenko, and Zhu (2013) for English tweets. We downloaded tweets from May 1 to June 31, 2014, containing a list of seed hashtags (the synonyms of *'#good'* and *'#bad'*). We split the tweets into word unigrams and we compute the pointwise mutual information (PMI) for the hypothetical classes *positive* (P) and *negative* (N). We finally compute the SO of a term x as follows,

[2] www.nltk.org

like Mohammad et al.:

$$SO(x) = PMI(x, P) - PMI(x, N) \quad (1)$$

The resulting SO is then normalised to a number between 5 and -5, following an strategy applied by other authors (Taboada et al., 2011). This automatically-built dictionary can be freely downloaded at `http://www.grupolys.org/software/LYSA/LYSA-v-0.1.txt`

The collected features are used to train a LibLINEAR classifier (Fan et al., 2008). We also consider applying feature selection filters, as we did last year. Specifically, we include an information gain (IG) filter, which measures the relevance of each feature with respect to the class, where features with an IG of zero are irrelevant to classify an instance. Readers are encouraged to test a model based on this approach at `miopia.grupolys.org`. It is free to use via an API.

4.1 Results

Tables 1 and 2 illustrate the performance of different models when performing a classification into 6 and 4 polarities. In both cases, we carried out a 5-fold cross-validation (cv) over the official training set. Values are computed as micro-averages. The performance of the different models maintains a similar ranking for the two evaluations proposed by the organisation.

Table 3 compares the performance on these tasks with respect to the rest of participants of the TASS 2014.

Model	P	R	F1	Acc.
LPTESM*	**0.439**	**0.471**	0.441	**0.471**
LPTES	**0.439**	0.470	0.440	0.469
LPTE	0.435	0.467	0.435	0.467
LPTESM* (no IG)	0.438	0.463	**0.446**	0.463
LPT	0.420	0.450	0.418	0.449
LP	0.422	0.451	0.417	0.451
L	0.406	0.422	0.386	0.423
W	0.410	0.416	0.379	0.416
E	0.333	0.408	0.331	0.408
P	0.354	0.395	0.341	0.395
S	0.355	0.3410	0.272	0.342
M	0.112	0.2690	0.156	0.269

Table 1: Performance on the 6 polarities training set (5-fold cv) of different models. W stands for features obtained from a bag-of-words, L from lemmas, P from (Ramírez-Esparza et al., 2007), T from PoS-tags, E from (Saralegi and San Vicente, 2013), S from (Brooke, Tofiloski, and Taboada, 2009) and M from our automatically-built lexicon. Models marked with an '*' indicate our official runs (task 1).

Model	P	R	F1	Acc.
LPTESM* (no IG)	**0.617**	**0.643**	**0.626**	**0.643**
LPTESM*	0.610	0.638	0.615	0.638
LPTES	0.608	0.636	0.613	0.636
LPTE	0.603	0.633	0.613	0.633
LPT	0.583	0.615	0.592	0.615
LP	0.583	0.613	0.590	0.613
L	0.559	0.589	0.565	0.589
W	0.564	0.590	0.565	0.589
E	0.545	0.558	0.538	0.558
P	0.516	0.556	0.529	0.556
S	0.525	0.473	0.437	0.473
M	0.342	0.459	0.375	0.459

Table 2: Performance on the 4 polarities training set (5-fold cv) of different models (task 1).

5 *Task 2: Topic Classification*

This is a legacy task proposed by the organisation since the first TASS edition. The aim is determining what a tweet is talking about. Ten topics are taken into account: *film*, *football*, *economics*, *entertainment*, *literature*, *music*, *politics*, *sports*, *technology* and *other*. This is a multi-label classification problem, since a user can relate more than one topic in the same tweet. In Vilares, Alonso, and Gómez-Rodríguez (2014a) we presented an evaluation of different supervised models for topic classification using linguistic knowledge. We chose a *one vs all* strategy to address the problem of assigning multiple labels to the same instance: given a set of labels, N, we create |N| classifiers where each one is able to distinguish a class i from the other classes j, where $i \in N \wedge j \in N \wedge i \neq j$. The experimentation was made on the TASS 2013 corpus, obtaining that a model which used a bag-of-words and bi-grams of lemmas as features, was able to achieve a consistent performance. Thus, we used that model for this year's edition. Two runs were submitted, where the only difference was the inclusion (or not) of an IG filter as a previous step.

Team	Accuracy			
	6	6 (1k)	4	4 (1k)
ELIRF-UPV	0.643_1	0.48_1	0.709_1	0.659_1
ELHUYAR	0.610_2	0.474_2	0.699_2	0.635_3
LYS	0.578_3	0.455_4	0.675_3	0.637_2
SINAI	0.513_4	0.464_3	0.612_4	0.633_4
JRC	0.484_6	0.422_5	0.611_5	0.562_5
SINAI-ESMA	0.509_5	0.368_6	0.606_6	0.523_6
IPN	0.373_7	0.347_7	0.564_7	0.518_7

Table 3: Ranking for task 1 of TASS 2014. Some teams submitted more than one run, although we only show the best performance obtained for each participant. Subscripts indicate their rank for each corpus. The columns *6 (1k)* and *4 (1k)* refer to the classification on the balanced general corpus for six and four polarities.

5.1 Results

In this case, we made an 80-20 split of the official training set, to create a a custom training and development set.[3] To evaluate our models we used standard metrics for multi-label classification: *Hamming loss distance*, *label-based accuracy* and *exact-match*. They are calculated according to equations 2, 3 and 4, 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 :

$$Hamming\ loss = \frac{1}{|D|} \sum_{i=1}^{|D|} \frac{Y_i \triangle Z_i}{L} \quad (2)$$

$$Label\text{–}based\ accuracy = \frac{1}{|D|} \sum_{i=1}^{|D|} \frac{Y_i \cap Z_i}{Y_i \cup Z_i} \quad (3)$$

$$Exact\ match = \frac{\#instances\ exactly\ labelled}{\#instances} \quad (4)$$

We show in Table 4 the performance of our two runs on the development set.

Model	EM	LBA	HL
WBL	0.454	0.495	0.684
WBL (no IG)	0.470	0.511	0.094

Table 4: Performance for our models on the development set (task 2).

[3] A one vs all strategy is costly in terms of time if it is addressed from a sequential perspective, so we decided not to use cross-validation.

On the other hand, Table 5 compares our approach with the rest of the participants.

Model	P	R	F1
ELIRF-UPV	0.700	0.706	0.703
LYS	0.683	0.596	0.636
IPN	0.271	0.332	0.299

Table 5: Ranking for task 2 of TASS 2014.

6 Task 3: Aspect extraction

The aim of task 3 is detecting the aspects from a predefined list, related with the football domain. Those aspects may refer to a player or a team, as we explained in previous sections.

Given the training set, we first obtain for each *aspect* a set of its *representations*. A representation is a particular form of referring to a concept, entity or a target. For example, the aspect *Football Player Lionel Messi* may have different representations in the training set, such as *'Messi'*, *'Leo'* or *'Leo Messi'*. The resulting collection of aspects and their representations is used to identify those same aspects when processing new instances. If a representation matches in a tweet, we hypothesise that we have found an aspect. This is a naive approach which will serve as a future starting point to build more complex techniques. However, since the task was limited to an small set of predefined aspects, we achieved an acceptable performance both on the development and the official test set, as shown below.

6.1 Results

To evaluate our model, we made again a random split of the original training set, employing the 80% of the corpus as our training set and the remaining 20% as the development set. We detail the performance obtained for each aspect in Table 6. The comparison with respect to the rest of participants can be found in Table 7.

7 Task 4: Aspect-based Sentiment Analysis

Once we have identified a representation of an aspect, the next step consists of detecting its *scope of influence*, *i.e.* the fragment of the text which is talking about the aspect that was referred to. In this paper, we propose

Aspect	P	R	F1	#
Isco	1.000	1.000	1.000	7
Dani Carvajal	1.000	1.000	1.000	5
Xavi Hernández	1.000	1.000	1.000	3
Sergio Busquets	1.000	1.000	1.000	1
Andrés Iniesta	1.000	1.000	1.000	1
Carles Puyol	1.000	1.000	1.000	1
Gareth Bale	1.000	1.000	1.000	45
Karim Benzema	1.000	1.000	1.000	1
Pepe	1.000	1.000	1.000	3
Neymar Jr.	1.000	1.000	1.000	14
Jesé Rodríguez	1.000	1.000	1.000	3
Sergio Ramos	1.000	0.958	0.979	48
Lionel Messi	0.980	0.961	0.970	51
Pinto	1.000	0.923	0.960	13
Marc Bartra	1.000	0.909	0.952	11
Iker Casillas	0.950	0.905	0.927	21
Real Madrid	0.896	0.938	0.916	128
Cristiano Ronaldo	1.000	0.833	0.909	24
Dani Alves	0.833	1.000	0.909	5
Barcelona	0.889	0.914	0.901	70
Entrenador	0.947	0.857	0.899	21
Angel Di María	1.000	0.625	0.769	8
Partido	0.597	0.875	0.709	88
Árbitro	1.000	0.400	0.571	5
Afición	0.550	0.512	0.530	43
Cesc Fábregas	1.000	0.333	0.499	6
Jugador	0.308	0.500	0.381	16
Autoridades	0.120	0.500	0.194	6
Retransmisión	0.000	0.000	0.000	4
Equipo	0.000	0.000	0.000	3
Javier Mascherano	0.000	0.000	0.000	2
Asier Ilarramendi	0.000	0.000	0.000	1
Atlético de Madrid	0.000	0.000	0.000	2
Weighted average	0.799	0.858	0.827	660

Table 6: Performance on the development set (task 3).

Model	P	R	F1
ELIRF-UPV	0.906	0.911	0.909
LYS	0.810	0.903	0.854

Table 7: Ranking for task 3 of TASS 2014.

and evaluate different heuristics to detect the scope of an aspect:

- *Whole tweet scope (baseline)*: The whole content of the tweet where an aspect appears is taken as the scope.
- *Sentence scope*: The sentence where the aspect was found is taken as the hypothetical scope. Since tweets are short, this heuristic would be in most cases equivalent to the baseline.
- *Syntactic scope*: Given a word of the graph, w_i, which corresponds to the root of a subtree referring to an aspect, its scope is the subtree rooted at the ancestor node k levels above it, $w_i^{(k)}$, where $1 \leq k \leq 3$. Figure 1 illustrates how these heuristics work. The example contains three representations of aspects, highlighted in bold letters: (1) *'Madrid fans'*, (2) *'they'* (referring also to the Real Madrid supporters) and (3) *'keeper'* (which in this case would be Iker Casillas). The solid-line box indicates the scope for each aspect, when $k = 1$. The obtained scope seems to be reliable for *'they'*, but it is too large for the term *'keeper'*. And it is clearly incomplete for the aspect *'Madrid fans'*. On the other hand, the dashed line boxes show the resulting scope when $k = 2$. Finally, the dotted line shows the scope obtained for *'Madrid fans'*, when $k = 3$.

The sentiment of each scope is then obtained by a supervised classifier. The model employed was the same as in our official runs for task 1, although the training corpus was different. We built a custom training set for each heuristic from the social TV corpus. Taking the output of task 3, where we identify a set of aspects for each tweet, we apply our scope heuristics, obtaining the potential fragments of text related with them. We use those fragments as our training sets (one for each heuristic), assigning to them the polarity corresponding to the referred aspect in the original training set.

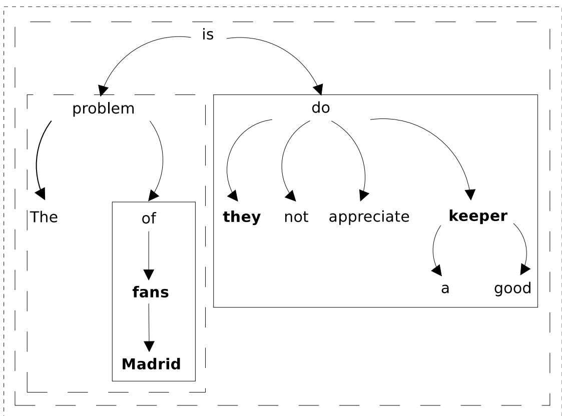

Figure 1: Example for different scope detection heuristics.

7.1 Results

Table 8 shows the performance for the different scopes proposed. We performed a 5-fold cross-validation on our custom training sets. The employment of syntactic structure is useful in terms of precision, recall and F-measure, compared with snippets such as the tweet or the sentence.

Scope	P	R	F1	Acc.
head term (no IG)	**0.620**	**0.627**	**0.622**	**0.628**
head term	0.594	0.600	0.596	0.600
3-levels up	0.564	0.582	0.565	0.582
2-levels up	0.565	0.581	0.569	0.581
sentence	0.562	0.578	0.562	0.577
tweet	0.561	0.576	0.559	0.576

Table 8: Performance on the custom training set (5-fold cv) for different scope heuristics (task 4).

Finally, Table 9 ranks the participant systems for task 4.

Team	P	R	F1
ELIRF-UPV	0.578	0.596	0.586
LYS	0.518	0.577	0.546

Table 9: Ranking for task 4 of TASS 2014.

8 *Conclusions and Future work*

This paper has described the participation of the LyS research group at TASS 2014. The proposed challenges are addressed from a natural language processing perspective. Tasks 1 (global sentiment analysis) and 2 (topic classification) rely on a machine learning approach to obtain the classifications. On the other side, task 3 (aspect-detection) follows an unsupervised perspective. Finally, task 4 (aspect-based sentiment analysis) employs heuristics to identify the scope of each aspect, to then apply a supervised classifier to obtain their sentiment. The official results reinforce the robustness of the proposed models. As future work, we are especially interested in addressing problems related with aspect-based sentiment analysis. The workshop has focused on identifying a predefined list of aspects related with the football domain. We plan to explore semi-supervised techniques to be able to detect and extract new aspects. In this respect, studies such as (Vechtomova, 2014) could help to enrich our future approaches. In this paper we also proposed a naive algorithm to identify the scope of influence of the aspects, showing the usefulness of employing syntactic information. We would like to explore more complex techniques to push performance beyond lexical-based approaches.

Bibliografía

Balahur, Alexandra and José M. Perea-Ortega. 2013. Experiments using varying sizes and machine translated data forsentiment analysis in twitter. In A. Díaz Esteban, I. Alegría Loinaz, and J. Villena Román, editors, *XXIX Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN 2013). TASS 2013 - Workshop on Sentiment Analysis at SEPLN 2013*, pages 172–178, Madrid, Spain, September.

Brill, E. 1992. A simple rule-based part of speech tagger. In *Proceedings of the workshop on Speech and Natural Language*, HLT'91, pages 112–116, Stroudsburg, PA, USA. Association for Computational Linguistics.

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.

Díaz Esteban, A., I. Alegría Loinaz, and J. Villena Román, editors. 2013. *XXIX Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN 2013). TASS 2013 - Workshop on Sentiment Analysis at SEPLN 2013*, Madrid, Spain, September. SEPLN.

Fan, Rong-En, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. 2008. LIBLINEAR: A library for large linear classification. *The Journal of Machine Learning Research*, 9:1871–1874.

Fernández, Javi, Yoan Gutiérrez, José M Gómez, Patricio Martınez-Barco, Andrés Montoyo, and Rafael Munoz. 2013. Sentiment analysis of spanish tweets using a ranking algorithm and skipgrams. In Alberto Díaz Esteban, Iñaki Alegría Loinaz, and Julio Villena Román, editors, *XXIX*

Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN 2013). TASS 2013 - Workshop on Sentiment Analysis at SEPLN 2013, pages 133–142, Madrid, Spain, September.

Gimpel, Kevin, Nathan Schneider, Brendan O'Connor, Dipanjan Das, Daniel Mills, Jacob Eisenstein, Michael Heilman, Dani Yogatama, Jeffrey Flanigan, and Noah A. Smith. 2011. Part-of-speech tagging for Twitter: annotation, features, and experiments. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers - Volume 2*, HLT '11, pages 42–47, Stroudsburg, PA, USA. Association for Computational Linguistics.

Martínez-Cámara, Eugenio, Miguel Ángel García Cumbreras, M. Teresa Martín-Valdivia, and L. Alfonso Ureña López. 2013. Sinai-emml: Combinación de recursos lingüísticos para elanálisis de la opinión en twitter. In Alberto Díaz Esteban, Iñaki Alegría Loinaz, and Julio Villena Román, editors, *XXIX Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN 2013). TASS 2013 - Workshop on Sentiment Analysis at SEPLN 2013*, pages 187–194, Madrid, Spain, September.

Mohammad, Saif M., Svetlana Kiritchenko, and Xiaodan Zhu. 2013. Nrc-canada: Building the state-of-the-art in sentiment analysis of tweets. In *Proceedings of the seventh international workshop on Semantic Evaluation Exercises (SemEval-2013)*, Atlanta, Georgia, USA, June.

Montejo-Ráez, A., Manuel C. Díaz Galiano, and M. García-Vega. 2013. LSA based approach to TASS 2013. In Díaz Esteban et al. (Díaz Esteban, Alegría Loinaz, and Villena Román, 2013), pages 195–199.

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.

Pla, Ferran and Lluís-F. Hurtado. 2013. ELiRF-UPV en TASS-2013: Análisis de sentimientos en Twitter. In Díaz Esteban et al. (Díaz Esteban, Alegría Loinaz, and Villena Román, 2013), pages 220–227.

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.

Saralegi, Xabier and Iñaki San Vicente. 2013. Elhuyar at TASS 2013. In Alberto Díaz Esteban, Iñaki Alegría Loinaz, and Julio Villena Román, editors, *XXIX Congreso de la Sociedad Española de Procesamiento de Lenguaje Natural (SEPLN 2013). TASS 2013 - Workshop on Sentiment Analysis at SEPLN 2013*, pages 143–150, Madrid, Spain, September.

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

Taulé, M., M. A. Martí, and M. Recasens. 2008. AnCora: Multilevel Annotated Corpora for Catalan and Spanish. In Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odjik, Stelios Piperidis, and Daniel Tapias, editors, *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco.

Vechtomova, Olga. 2014. A method for automatic extraction of multiword units representing business aspects from user reviews. *Journal of the Association for Information Science and Technology*, 65(7):1463–1477.

Vilares, David, Miguel A Alonso, and Carlos Gómez-Rodríguez. 2014a. Evaluación de aproximaciones lingüısticas para la asignación de temas a tuits. In *III Congreso Español de Recuperación de la Información, A Coruna, Spain.*

Vilares, David, Miguel A. Alonso, and Carlos Gómez-Rodríguez. 2014b. On the usefulness of lexical and syntactic processing in polarity classification of Twitter messages. *Journal of the Association for Information Science Science and Technology*, to appear.