Procesamiento del Lenguaje Natural, Revista nº 54, marzo de 2015, pp 61-68 recibido 24-11-14 revisado 27-01-15 aceptado 10-02-15

TASS 2014 - The Challenge of Aspect-based Sentiment Analysis*

TASS 2014 - El Reto del Análisis de Opiniones a nivel de aspecto

Julio Villena Román, Janine García Morera
Daedalus, S.A.
E-28031 Madrid, Spain
{jvillena, jgarcia}@daedalus.es

Eugenio Martínez Cámara, Salud M. Jiménez Zafra
SINAI Research Group - University of Jaén
E-23071 Jaén, Spain
{emcamara, sjzafra}@ujaen.es

Resumen: El análisis de la reputación y el análisis de opiniones son dos tareas que están en boga actualmente. Pero esa moda viene justificada por la necesidad, cada vez más acuciante, de conocer la orientación de las opiniones que se publican diariamente en Internet. TASS es un taller de trabajo que tiene como fin fomentar la investigación en el descubrimiento de la orientación de la opinión de textos en español publicados en Internet. En este artículo se describe la tercera edición de TASS, en el que se han mantenido dos tareas propuestas en las dos ediciones anteriores, y se han planteado otras dos nuevas relacionadas con el análisis de opiniones a nivel de aspecto, y que se encuentran circunscritas en el fenómeno de la Televisión Social.
Palabras clave: TASS 2014, Análisis de Opiniones, Análisis de Opiniones a nivel de Aspecto, Televisión Social

Abstract: Currently, reputation and sentiment analysis are trendy tasks. However, the interest on these two tasks is growing by the need of knowing the polarity of the opinions published on the Internet. TASS is a workshop whose goal is to boost the research on sentiment analysis in Spanish. Hereinafter the third issue of TASS is described, in which four tasks have been proposed. Two of the proposed tasks are known by former participants and the other two ones are new. These new tasks are related to sentiment analysis at entity level, and they are circumscribed on the Social TV phenomenon.
Keywords: TASS 2014, Sentiment Analysis, Aspect Based Sentiment Analysis, Social TV.

1 *Introduction*

Workshop on Sentiment Analysis at SEPLN (TASS, in Spanish) is an experimental evaluation workshop on reputation and sentiment analysis (SA) focused on Spanish language, organized as a satellite event of the SEPLN Conference. After two successful editions in 2012 (Villena-Román et al., 2013) and 2013 (Villena-Román et al., 2014), TASS 2014[1] was held on September 16th, 2014 at University of Gerona, Spain.

The long-term objective of TASS is to foster the research on the field of reputation, i.e., the process of tracking, investigating and reporting an entity's actions and other entities' opinions about those actions, in Spanish language. As a first approach, reputation analysis encompasses at least two technological aspects: SA and text classification.

Nowadays, SA means the computational treatment of opinion, sentiment and subjectivity in text (Pang and Lee, 2008). It is a major technological challenge and the task is so hard that even humans often disagree on the sentiment of a given text, as issues that one individual may find acceptable or relevant may not be the same to others. And the shorter the text is (for instance, Twitter messages), the harder the task becomes.

On the other hand, automatic text classification (or categorization) is used to guess the topic of the text, among those of a predefined

* This work has been partially supported by a grant from the Fondo Europeo de Desarrollo Regional (FEDER), ATTOS project (TIN2012-38536-C03-0), Ciudad2020 (INNPRONTA IPT-20111006), MULTIMEDICA (TIN2010-20644-C03-01) from the Spanish Government, AORESCU project (P11-TIC-7684 MO) from the regional government of Junta de Andalucía, MA2VICMR (S2009/TIC-1542) from the regional government of Comunidad de Madrid and CEATIC-2013-01 project from the University of Jaén.

[1] www.daedalus.es/TASS2014

ISSN 1135-5948

set of categories, so as to be able to assign the reputation level into different axis or points of view of analysis. Text classification techniques, albeit studied for a long time, still need more research effort to be able to build complex models with many categories with less workload and increase the precision and recall of the results. In addition, these models should deal with specific text features in social media messages.

Up to now, TASS has proposed analyses at document level (tweet level), but the SA research community are beginning to go a step further, related to the fact that the society needs a fine-grained study of people attitude expressed on a tweet. Aspect-Based Sentiment Analysis (ABSA) is the task that is concerned with the extraction and classification of opinions on a specific entity. An entity can be decomposed into several parts or aspects, so that can be seen as a hierarchical structure whose head is the entity. ABSA is not only focused on opinions on entities, but also each of the aspects that are part of an entity. In a pragmatic way, an ABSA system does not take into account the hierarchical relation between the entity and the aspects, and both are considered in the same way. ABSA encompasses two subtasks, aspect extraction and aspect sentiment classification. The first one empathizes on the identification of the aspects presented on a text, and the second one comprehends the classification of the attitude of the opinion holder about the aspect.

The previous paragraphs described trendy, hard, and interesting tasks that are basic for a posterior study of reputation. Within this context, the aim of TASS is to provide a forum for discussion the latest research work in these fields. The setup is based on a series of challenge tasks intended to provide a benchmark forum for comparing different approaches. Moreover, the aim of TASS is to provide a common reference dataset for the research community, so it is generated and open-release the corpus fully tagged. Polarity classification and topic classification are two fixed tasks of TASS, but due to the relevance of ABSA, the 2014 edition of TASS has included two new tasks, aspect identification and aspect-based polarity classification, which is focused on the context of Social TV.

The rest of the paper is organized as follows. Section 2 describes the corpora provided to participants and used for the challenge tasks. The third section describes the different tasks proposed in the 2014 edition. Section 4 and 5 describes the participants and the analysis of the results, and the last section draws some conclusions and future directions.

2 Corpus

Experiments were based on two corpora. After the workshop, both were published only for research purposes.

2.1 General corpus

The general corpus, which is the same used in the previous two editions, contains over 68,000 tweets gathered between November 2011 and March 2012. The tweets are written in Spanish by about 150 well-known personalities and celebrities of the world of politics, economy, communication, mass media and culture.

The general corpus was divided into two sets: training (10%) and test (90%). Table 1 shows a summary of the training and test corpora provided to participants.

Attributes	**Value**
Tweets	68,017
Tweets (test)	60,798 (89%)
Tweets (train)	7,219 (11%)
Topics	10
Tweet languages	1
Users	154
User types	3
User languages	1
Date start (train)	2011-12-02 T00:47:55
Date end (train)	2012-04-10 T23:40:36
Date start (test)	2011-12-02 T00:03:32
Date end (test)	2012-04-10 T23:47:55

Table 1: Features of the General Corpus

Each message in both the training and test set was tagged with its global polarity, indicating whether the text expresses a positive, negative or neutral sentiment, or no sentiment at all. Five levels have been defined: strong positive (P+), positive (P), neutral (NEU), negative (N), strong negative (N+) and one additional no sentiment tag (NONE).

Furthermore, the level of agreement of the expressed sentiment within the text was also included, to clarify whether a neutral sentiment comes from neutral keywords (AGREE-

MENT) or else the text contains positive and negative sentiments at the same time (DISAGREEMENT).

On the other hand, a selection of a set of topics was made based on the thematic areas covered by the corpus, such as politics, literature or entertainment. Each message in both the training and test set was assigned to one or several of these topics.

All tagging was carried out semi automatically: a baseline machine learning model was first run (Villena-Román et al., 2011) and then all tags were manually checked by two human experts. For polarity at entity level, due to the high volume of data to check, this tagging was done just for the training set.

2.2 Social-TV corpus

Social-TV corpus was collected during the 2014 Final of Copa del Rey championship in Spain, between Real Madrid and F.C. Barcelona. It was played on 16 April 2014 at Mestalla Stadium in Valencia.

Over 1 million of tweets were collected from 15 minutes before to 15 minutes after the match. After filtering useless information, tweets in other languages than Spanish, a subset of 2773 was selected.

All tweets have been manually tagged with the aspects of the expressed messages and its sentiment polarity. Tweets may cover more than one aspect.

The general defined aspects were: *afición* (fans), *árbitro* (referee), *autoridades* (political authorities), *entrenador* (coach), *equipo* (team), *jugador* (player), *partido* (game) and *retransmisión* (broadcasting).

Some of the detailed aspects were: Equipo-Barcelona, Equipo-Real_Madrid, Jugador-Isco, Jugador-Dani_Álves, and the other players.

Sentiment polarity has been tagged from the point of view of the person who writes the tweet, using 3 levels: P (positive), NEU (neutral) and N (negative). No distinction was made in cases when the author does not express any sentiment or when he/she expresses a no-positive or no-negative sentiment.

The Social-TV corpus has been randomly divided into two sets: training (1773 tweets) and test (1000 tweets), with a similar distribution of both, aspects and sentiments. The training set was released with the aim of the participants could train and validate their models. The test corpus was provided without any tagging and was used to evaluate the results provided by the different systems. The list of the 31 aspects that have been defined can be read at the workshop webpage.

```
<tweet id="456544898791907328">
  <sentiment aspect="Equipo-Real_Madrid
      " polarity="P">#HalaMadrid</
      sentiment> ganamos sin <sentiment
       aspect="Jugador-
      Cristiano_Ronaldo" polarity="NEU"
      >Cristiano</sentiment>. .perdéis
      con <sentiment aspect="Jugador-
      Lionel_Messi" polarity="N">Messi<
      /sentiment>. Hala <sentiment
      aspect="Equipo-Real_Madrid"
      polarity="P">Madrid</sentiment>!
      !!!!!
</tweet>
<tweet id="456544898942906369">
  @nevermind2192 <sentiment aspect="
      Equipo-Barcelona" polarity="P">
      Barça</sentiment> por siempre!!
</tweet>
<tweet id="456544898951282688">
  <sentiment aspect="Partido" polarity=
      "NEU">#FinalCopa</sentiment> Hala
       <sentiment aspect="Equipo-
      Real_Madrid" polarity="P">Madrid<
      /sentiment>, hala <sentiment
      aspect="Equipo-Real_Madrid"
      polarity="P">Madrid</sentiment>,
      campeón de la <sentiment aspect="
      Partido" polarity="P">copa del
      rey</sentiment>
</tweet>
```

Figure 1: Sample tweet (Social TV corpus)

Figure 1 shows the information of three sample tweets in the training set.

3 Description of tasks

The main goal of TASS is to boost the research on reputation and SA in Spanish. With the aim of reaching it, the organization of TASS always proposed four tasks, two of them that have the purpose of analysing the evaluation of the investigation on SA and Topic Classification, and another two ones that are usually linked with needs of the society, which are usually voiced by a business demand. The two fixed tasks of TASS are Sentiment Analysis and Topic Classification at document level, which will be described hereinafter.

2014, in Spain, has been the year in which the TV channels have greatly taken advantage of the social networks with the objective of increasing the participation of the viewers in the TV shows. Last July, the CEO of Twitter Spain, José López de Ayala, asserted that the 66% of mobile phone users publish

tweets while they are watching TV. Also, he pointed out that the TV ads, whose last image is a Twitter hashtag, they achieve to increase by 60% the number of tweets related to that tag. Therefore, analyzing the sentiment related to the phenomenon of Social TV was a great candidate to be the target of a task. The level of analysis required by a reputation or SA system in the context of Social TV is deeper than the proposed one in the traditional tasks of TASS, in plain English, Social TV needs a fine-grained analysis. The level of analysis needed in a Social TV context is entity or aspect level. This is the reason why this year two new tasks were proposed. The new tasks required the development of SA systems at aspect level.

3.1 Task 1: SA at Document Level

The first of the two fixed tasks of TASS is the performing SA at document level. In the context of the workshop the task proposed the development of polarity classification systems at tweet level, in other words, build systems to classify tweets in several predefined polarity classes. Six is the number of polarity labels (P+, P, NEU, N, N+, NONE) in which the systems had to classify the tweets of the general corpus. But the systems had to be prepared to classify tweets in four labels (P, NEU, N, NONE), because the performance of the systems is evaluated in an environment of six classes and four classes.

Accuracy was used for ranking the systems. Precision, Recall and F1-measure will be used to evaluate each individual category.

Results were submitted in a plain text file with the following format:

```
tweetid \t polarity
```

where polarity could be: P+, P, NEU, N, N+ and NONE for the 6-labels case; P, NEU, N and NONE for the 4-labels case.

The same test corpus of previous years was used for the evaluation, to allow comparison between systems. Obviously, participants were not allowed to use any test data to train their systems. However, to deal with the problem reported last years of the imbalanced distribution of labels between the training and test set, a new selected test subset containing 1000 tweets with a similar distribution to the training corpus was extracted and used for an alternate evaluation of the performance of systems.

3.2 Task 2: Topic Classification

The challenge of this task is to automatically identify the topic of each message in the test set of the General corpus. Participants could use the training set of the General corpus to train and validate their models.

Participants were expected to submit up to 3 experiments, each one in a plain text file with the following format:

```
tweetid \t topic
```

A given tweet ID can be repeated in different lines if it is assigned more than one topic.

Micro averaged precision, Recall and F1-measure calculated over the full test set will be used to evaluate the systems. Systems were ranked by F1. To allow the comparison with previous years, the same test corpus will be used for the evaluation. Again, participants were not allowed to use any test data to train their systems.

3.3 Task 3: Aspect Detection

The main objective of this task is the automatic identification of the different aspects expressed by users, among a predefined list, in their opinions expressed in Twitter about a given topic. For example, in the following tweet:

```
CR7 jugó bien, Messi no, el Madrid
se mereció la victoria
(CR7 played well, Messi didn't, the
Madrid team deserved to win)
```

Three aspects can be identified CR7 as the player Cristiano Ronaldo, Messi as the player Lionel Messi and Madrid as the Real Madrid team. This task is a multi-label classification and tweets can have more than one aspect, as shown in the example.

A new Social-TV corpus was delivered and used for the training and evaluation of the systems (see description above).

Participants are expected to submit up to 3 experiments, each in a plain text file with the following format:

A given tweet id can be repeated in different lines if it is assigned more than one

```
tweetid \t aspect
```

aspect. We consider an aspect as the minimum set of words, not the detected terms or fragment in neither the text nor its offsets.

As evaluation measures micro averaged precision, recall and F1 were used, calculated over the full test set. The final list of participants was ranked by F1.

3.4 Task 4: Aspect-based SA

This task was similar to the first one, but sentiment polarity (using 3 levels) should be determined at aspect level of each tweet in the Social-TV corpus (fine-grained polarity detection). They worked with the aspects detected in the previous task. Again, participants were provided with this Social-TV corpus to train and evaluate their models. Aspects were tagged in the corpus to make participant focus on the polarity classification and not on aspect identification. The complex of the task arises from the fact that tweets can contain more than one sentence with more than one aspect per sentence, so more advance text processing techniques are needed.

Participants were expected to submit up 3 experiments, each in a plain text file with the following format:

```
tweetid \t aspect \t polarity
```

Allowed polarity values were P, NEU and N.

Accuracy, micro averaged Precision, Recall and F1-measure was used to evaluate the systems, considering a unique label combining aspect-polarity. Systems were ranked by F1.

4 Participants

This year 35 groups registered (as compared to 31 groups last year) and finally 7 groups (14 last year) sent their submissions. The list of active participant groups is shown in Table 2, including the tasks in which they have participated.

Along with the experiments, all participants were invited to submit a paper with the description of their experiments and the analysis of the results. These papers were reviewed by the program committee and were included in the workshop proceedings. References are listed in Table 3.

Group	**1**	**2**	**3**	**4**
LyS	✓	✓	✓	✓
SINAI-ESMA	✓			
Elhuyar	✓			
SINAIword2vec	✓			
JRC	✓			
ELiRF-UPV	✓	✓	✓	✓
IPN	✓	✓		
Total groups	7	3	2	2

Table 2: Participant groups

Vilares et al. (2014) used 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, they developed a naive approach, collecting a set of representations to identify the predefined aspects requested by the organizers. Jiménez Zafra et al. (2014) developed an unsupervised classification system which is based on the use of an opinion lexicon, and on the application of a syntactic heuristic for identifying the scope of Spanish negation words. San Vicente Roncal and Saralegi Urizar (2014) implemented a SVM algorithm that combines the information extracted from polarity lexicons with linguistic features. Montejo Ráez, García Cumbreras, and Díaz-Galiano (2014) used supervised learning with SVM over the summatory of word vectors in a model generated from the Spanish Wikipedia. Perea-Ortega and Balahur (2014) focused on different feature replacements carried out for both the development and test data sets provided. The replacements performed were mainly based on repeated punctuation signs, emoticons and affect words, by using an in-house built dictionary for SA. Then, they applied a machine learning approach to get the polarity of the tweets. Hurtado and Pla (2014) adapted the tweet tokenizer Tweetmotif (Connor, Krieger and Ahn, 2010) and they used Freeling (Padro y Stanilovsky, 2012) as lemmatizer, entity recognizer and morphosyntactic tagger. Hernández Petlachi and Li (2014) proposal is based on semantic ap-

proaches with linguistic rules for classifying polarity texts in Spanish. Polarity classification in the words is done according to a dictionary of semantic orientation where each term is labeled with a use value and emotional value, along with linguistic rules to solve various constructions that could affect the polarity of text.

Group	Report
LyS	(Vilares et al., 2014)
SINAI-ESMA	(Jiménez Zafra et al., 2014)
Elhuyar	(San Vicente Roncal and Saralegi Urizar, 2014)
SINAIword2vec	(Montejo Ráez, García Cumbreras, and Díaz-Galiano, 2014)
JRC	(Perea-Ortega and Balahur, 2014)
ELiRF-UPV	(Hurtado and Pla, 2014)
IPN	(Hernández Petlachi and Li, 2014)

Table 3: Participant reports

5 *Results*

After the submission deadline, runs were collected and checked and results were evaluated and made available to the participants using an automated web page in the password protected area in the website. Results for each task are described hereinafter.

5.1 Task 1: Sentiment Analysis at Document Level

Task 1 includes the experiments using the full test set and using the selected 1k test set.

Thirty-two runs for 5-level evaluation were submitted by 7 different groups. Results for the best-ranked experiment from each group are listed in the tables below. All tables show the precision (P), recall (R) and F1 value achieved in each experiment. Table 4 considers 5 polarity levels, with the whole test corpus. Accuracy values range from 0.64 to 0.37.

As previously described, an alternate evaluation of the performance of systems was done using a new selected test subset containing 1000 tweets with a similar distribution to the training corpus. Results are shown also in next Table 4. Accuracy values range from 0.48 to 0.33 (1k test corpus). Figures are much lower as compared to the previous evaluation, thus showing a high bias in the semi-automatic tagging of the whole test corpus.

In order to perform a more in-depth evaluation, results are calculated considering the classification only in 3 levels (POS, NEU, NEG) and no sentiment (NONE) merging P and P+ in only one category, as well as N and N+ in another one. The same double evaluation using the whole test corpus and a new selected corpus has been carried out, shown in Table 5.

The distributions of successful tweets per groups and per sentiment, for 3-level evaluation, are shown in Table 6 and Table 7.

# of groups	Correct tweets
7	13112 (21.6%)
6	11215 (18.5%)
5	9898 (16.3%)
4	7512 (12.4%)
3	5536 (9.1%)
2	4716 (7.8%)
1	4595 (7.6%)
0	4214 (6.9%)

Table 6: Distribution of successful tweets per groups, for 3-level evaluation

Label	Correct tweets
P	22007 (36.2%)
N	15655 (25.7%)
NONE	18076 (29.7%)
NEU	846 (1.4%)

Table 7: Distribution of successful tweets per sentiment, for 3-level evaluation

5.2 Task 2: Topic Classification

Table 8 shows the results for Task 2. Precision ranges from 67% to 27%. As in Task 1, different submissions from the same group usually have similar values.

Run Id	P	R	F1
ELiRF-UPV-run3	0.67	0.75	0.70
ELiRF-UPV-run2	0.70	0.71	0.70
ELiRF-UPV-run1	0.68	0.69	0.69
LyS-1	0.68	0.60	0.64
LyS-2	0.68	0.59	0.63
IPN-2	0.27	0.33	0.30

Table 8: Results for Task 2

Run Id	**Acc.**	**1k-Run Id**	Acc.
ELiRF-UPV-run3	0.64	ELiRF-UPV-run1-1k	0.48
ELiRF-UPV-run1	0.63	ELiRF-UPV-run3-1k	0.48
ELiRF-UPV-run2	0.63	Elhuyar-Run2-1k	0.47
Elhuyar-Run1	0.61	Elhuyar-Run3-1k	0.47
Elhuyar-Run3	0.61	ELiRF-UPV-run2-1k	0.47
Elhuyar-Run2	0.61	Elhuyar-Run1-1k	0.47
LyS-1	0.58	SINAIword2vec-1-1k	0.46
LyS-2	0.56	LyS-2-1k	0.46
SINAIword2vec-1	0.51	LyS-1-1k	0.45
SINAI-ESMA-1	0.51	JRC-run3-baseline-stop-1k	0.42
SINAI-ESMA-without_negation	0.51	JRC-run1-ER-1k	0.41
JRC-run1-ER	0.48	JRC-run2-RPSN-ER-AWM-4-all-2-skipbigrams-1k	0.40
JRC-run2-RPSN-ER-AWM-4-all-2-skipbigrams	0.48	SINAI-ESMA-without_negation-1k	0.37
JRC-run3-baseline-stop	0.48	SINAI-ESMA-1-1k	0.37
IPN-Linguistic_2	0.37	IPN-Linguistic_2-1k	0.35
IPN-1	0.37	IPN-1-1k	0.33

Table 4: Task 1, classification on 5 levels

Run Id	**Acc.**	**1k-Run Id**	Acc.
ELiRF-UPV-run2	0.71	ELiRF-UPV-run3-1k	0.66
ELiRF-UPV-run1	0.71	ELiRF-UPV-run1-1k	0.65
ELiRF-UPV-run3	0.70	LyS-2-1k	0.64
Elhuyar-Run1	0.70	Elhuyar-Run3-1k	0.64
Elhuyar-Run2	0.70	SINAIword2vec-2-1k	0.63
Elhuyar-Run3	0.70	Elhuyar-Run2-1k	0.63
LyS-1	0.67	LyS-1-1k	0.63
LyS-2	0.67	Elhuyar-Run1-1k	0.62
SINAIword2vec-2	0.61	SINAIword2vec-1-1k	0.61
JRC-run2-RPSN-ER-AWM-4-all-2-skipbigrams	0.61	ELiRF-UPV-run2-1k	0.60
SINAI-ESMA-1	0.61	JRC-run1-ER-1k	0.56
JRC-run1-ER	0.61	JRC-run3-baseline-stop-1k	0.56
SINAI-ESMA-without_negation	0.60	JRC-run2-RPSN-ER-AWM-4-all-2-skipbigrams-1k	0.55
JRC-run3-baseline-stop	0.60	SINAI-ESMA-1-1k	0.52
SINAIword2vec-1	0.59	SINAI-ESMA-without_negation-1k	0.52
IPN-Linguistic_2	0.55	IPN-Linguistic_2-1k	0.52

Table 5: Task 1, classification on 3 levels

5.3 Task 3: Aspect Detection

The aim of task 3 is detecting the aspects from a predefined list, related with the football domain, and using the Social-TV corpus. Results for Task 3 are shown in Table 9.

Run Id	**P**	**R**	**F1**
ELiRF-UPV-run1	0.91	0.91	0.91
LyS-1	0.81	0.90	0.85

Table 9: Results for Task 3

5.4 Task 4: Aspect-Based SA

Last, results for Task 4 are shown in Table 10. 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 and its SA.

The results are similar and robust and do not drop too low compared to those obtained on the while test corpus.

6 Conclusions and Future Work

Each edition of TASS has a positive conclusion because every year the Spanish SA research community improves their systems

Run Id	P	R	F1
ELiRF-UPV-run2	0.58	0.60	0.59
ELiRF-UPV-run1	0.57	0.59	0.58
ELiRF-UPV-run3	0.56	0.58	0.57
LyS-2	0.52	0.58	0.55
LyS-1	0.51	0.57	0.54
LyS-3	0.46	0.51	0.48

Table 10: Results for Task 4

and results. It is very important to note that the general results obtained are comparable to those of the international community. Two aspects are noteworthy, the first one is related to the fact that each edition is increasing the number of unsupervised or at least semi-supervised systems. This is important because they do not need prior knowledge to perform the classification, which is scarce in the dynamic context of social networks. The other issue is related to the fact that the systems submitted try to use the last methods in the state of the art, like classifiers based on deep learning.

The main purpose for future editions is to continue increasing the number of participants and the visibility of the workshop in international forums.

References

Hernández Petlachi, Roberto and Xiaoou Li. 2014. Análisis de sentimiento sobre textos en español basado en aproximaciones semánticas con reglas lingüísticas. In *Proceedings of the TASS workshop at SEPLN*. Gerona, Spain, September.

Hurtado, Lluís and Ferrán Pla. 2014. Elirf-upv en tass 2014: Análisis de sentimientos, detección de tópicos y análisis de sentimientos de aspectos en twitter. In *Proceedings of the TASS workshop at SEPLN*. Gerona, Spain, September.

Jiménez Zafra, Salud M., Eugenio Martínez Cámara, M. Teresa Martín Valdivia, and L. Alfonso Ureña López. 2014. Sinai-esma: An unsupervised approach for sentiment analysis in twitter. In *Proceedings of the TASS workshop at SEPLN*. Gerona, Spain, September.

Montejo Ráez, Arturo, M. Ángel García Cumbreras, and M. Carlos Díaz-Galiano. 2014. Participación de sinai word2vec en tass 2014. In *Proceedings of the TASS workshop at SEPLN*. Gerona, Spain, September.

Pang, Bo and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Found. Trends Inf. Retr.*, 2(1-2):1–135, January.

Perea-Ortega, José M. and Alexandra Balahur. 2014. Experiments on feature replacements for polarity classification of spanish tweets. In *Proceedings of the TASS workshop at SEPLN*. Gerona, Spain, September.

San Vicente Roncal, Iñaki and Xabier Saralegi Urizar. 2014. Looking for features for supervised tweet polarity classification. In *Proceedings of the TASS workshop at SEPLN*. Gerona, Spain, September.

Vilares, David, Yerai Doval, Miguel A. Alonaso, and Carlos Gómez-Rodríguez. 2014. Lys at tass 2014: A prototype for extracting and analysing aspects from spanish tweets. In *Proceedings of the TASS workshop at SEPLN*. Gerona, Spain, September.

Villena-Román, Julio, Sonia Collada-Pérez, Sara Lana-Serrano, and José Carlos González Cristóbal. 2011. Hybrid approach combining machine learning and a rule-based expert system for text categorization. In *Proceedings of the 24th International Florida Artificial Intelligence Research Society Conference (FLAIRS)*. AAAI Press.

Villena-Román, Julio, Janine García-Morena, Sara Lana-Serrano, and José Carlos González-Cristóbal. 2014. Tass 2013 - a second step in reputation analysis in Spanish. *Procesamiento del Lenguaje Natural*, 52(0):37–44.

Villena-Román, Julio, Sara Lana-Serrano, Eugenio Martínez-Cámara, and José Carlos González-Cristóbal. 2013. Tass - workshop on sentiment analysis at SEPLN. *Procesamiento del Lenguaje Natural*, 50:37–44.