Using relational discourse structure information in Basque sentiment analysis

El uso de la información de la estructura retórica en el analisis de sentimiento

Jon Alkorta, Koldo Gojenola, Mikel Iruskieta y **Alicia Perez**
IXA Group. University of the Basque Country
jon.alkorta@ehu.eus, koldo.gojenola@ehu.eus, mikel.iruskieta@ehu.eus, alicia.perez@ehu.eus

Resumen: En este artículo presentamos un estudio sobre el análisis del sentimiento que explota información extraida de la estructura relacional discursiva en un corpus en euskera sobre crítica literaria. Para el análisis discursivo hemos utilizado la *Rhetorical Structure Theory* (RST) y para la polaridad el método QWN-PPV. Los resultados preliminares demuestran que el análisis del discurso es efectivo para el análisis de opiniones.
Palabras clave: Análisis del sentimiento, polaridad, relaciones de coherencia, unidad central, RST, crítica literaria

Abstract: This paper presents a study in sentiment analysis which exploits information of the relational discourse structure in a Basque corpus consisting of literature reviews. The QWN-PPV method was employed to label all the texts at element level and the *Rhetorical Structure Theory* (RST) was used to extract discourse structure information. The preliminary results show that discourse structure is effective for opinion mining.
Keywords: Sentiment analysis, polarity, coherence relations, central unit, RST, literary criticism

1 Introduction

Sentiment analysis is nowadays a well known topic where the opinion, sentiment or subjectivity (Pang and Liu, 2008) are studied. The opinion about films (Pang and Lee, 2004), the success of politicians (Tumasjan et al., 2010) and the opinion of consumers about products are some of the topics studied.

Different levels of language has been studied in Sentiment Analysis. Hatzivassiloglou and McKeown (1997) studied the word level, Yu and Hatzivassiloglou (2003) the sentence level and Pang, Lee, and Vaithyanathan (2002) the discourse level. These are some examples of these three levels extracted form our corpus[1]:

i) Lexical level: where words[2] and entities have their own polarity[3], as in Example (1).

(1) [...] *literatura ona*$_{(+)}$ *sortu ahal izateko.* BER01
[...] to create good$_{(+)}$ literature.

In the example the word *ona* (good) has a positive polarity, because its entry in a dictionary has a positive value.

ii) Syntactic level: where the function in the word ordering or the clause's syntactic function can change the polarity assigned in the lexical level.

(2) *xede*$_{(+)}$ *onak*$_{(+)}$ *ez* *dira nahikoak*$_{(+)}$ *izaten literatura ona*$_{(+)}$ *sortu ahal izateko.* BER01
good$_{(+)}$ goals$_{(+)}$ are not enough$_{(+)}$ to create a good$_{(+)}$ literature.

In Example (2), the negation *ez* (not) changes the polarity assigned by the dictionary entries to a negative polarity determined by the negation of an otherwise positive statement.

[1] References and links to see the annotated text are at the end of the examples.

[2] For example, SentiWordNet (Esuli and Sebastiani, 2010) is a lexical resource for opinion mining which assigns three sentiment scores to each synset of WordNet: positivity, negativity, objectivity.

[3] In the following examples the polarity will be marked with: (+) when positive, (−) when negative and (∗) when neutral. All the examples were analyzed with the QWN-PPV method, explained below.

iii) Discourse level: where the coherence relations can highlight or even change a clause polarity (micro-structure), or the overall polarity of a text (macro-structure).

In Example (3) the change of polarity is out of the sentence scope, at discourse level.

(3) *Dokumentazio lana, esan bezala, nabarmena*$_{(+)}$ *da, eta baliabideen erabileran idazleak duen ahalmena eta egindako lana bereziki azpimarratzekoak*$_{(+)}$ *dira. Baina, horiek horrela izanik ere, emaitza zalantzagarria*$_{(*)}$ *da.* BER04
The documentation work, as mentioned before, is spectacular$_{(+)}$, and the capacity of the writer to use the resources and the work done especially are to underline$_{(+)}$. But, although that is so, the result is doubtful$_{(*)}$.

In this example, there are several words with a positive polarity, but the polarity of the example is not positive because a contrast discourse relation signaled by the adversative connector *baina* (but) has changed it.

This example demonstrates the importance of rhetorical relations, which can change the polarity of the sentence. For that reason, it is necessary, from our point of view, to also consider the discourse structure information in sentiment analysis.[4]

Currently there exists an Opinion Mining system for Basque. We have used this tool[5], that assigns automatically a positive or negative polarity to words[6]. The system makes use of the QWN-PPV method (Vicente, Agerri, and Rigau, 2014), that automatically generates polarity lexicons annotated at synset and lemma level. For that purpose, QWN-PPV uses a Lexical Knowledge Base (WordNet) and a list of positive and negative elements. QWN-PPV is a method that detects elements from lexical level and, consequently, the method is unable to correctly detect the polarity of the examples mentioned before —see examples (2) and (3).

[4]Example (4) below also shows the importance of discourse structure considering the main topic of the text and the rhetorical relation, when assigning a polarity score.

[5]Ber2Tek Opinion Mining can be tested at http://iritzierauzketa.ber2tek.eus/

[6]The method does not signal a neutral polarity.

With the aim of fulfilling this gap, we want to develop a method based in two language levels: the lexical and the discourse level. To that end, we will estimate the importance that the discourse structure has in sentiment analysis. Our study, which is based on a theoretical approach based on discourse, exploits information from the relational discourse structure in a Basque corpus consisting of literature reviews. The theory we employ to that end is the *Rhetorical Structure Theory* (RST) (Mann and Thompson, 1987)[7]. This theory describes the structure and coherence of text and it has been useful in Sentiment Analysis and in other many NLP advanced tasks (Taboada and Mann, 2006). In this respect, this work is a first approximation to Opinion Mining using discourse structure in Basque. This study, as far as we know, is the first work on Basque and sentiment analysis from a discourse point of view. Therefore, it fullfils this gap in Basque, but it is also relevant for RST, because it includes a different language to some recent works like (Trnavac and Taboada, 2014) in English and (Zhou et al., 2011) in Chinese.

The rest of the paper is structured as follows. Section 2 lays out the related work. Section 3 sets out the methodology we used and Section 4 presents the results. Finally, Section 5 presents a discussion and establishes directions for future work.

2 Related work

As we have explained before, Sentiment Analysis has three linguistic different levels, while we will focus on the lexical level and the discourse level interaction. Inside the discourse level, there are various categorizations according to different viewpoints.

In the discourse level there are two possible methods: language model and knowledge-based model. The first one determines if a span of a text is subjective, while the second one finds words with its polarity in a dictionary and calculates a sentiment score for all the text with an algorithm.

i) In the language model approach, Alistair and Diana (2005) use Support Vector Machines to classify sentiment expressed by movie reviews. Firstly, they use unigram fea-

[7]Other theories worth to mention are the *Segmented Discourse Representation Theory* (SDRT) (Asher, 1993) and the *Penn Discourse TreeBank* (PDTB) (Miltsakaki et al., 2005).

tures and then, they couple bigrams. The bigrams are composed by a valence shifter and another word. The results are acceptable and adding a term-counting method helps get better results.

ii) The knowledge-based approximation can be divided in four approaches (Cambria et al., 2013): keyword spotting, lexical affinity, statistical methods, and concept-level techniques. In the experiments, we will use the QWN-PPV method, which is an (almost) unsupervised method, i.e., a statistical method.

According to (Zhou, 2013), from a theoretical point of view, there are two approaches to Sentiment Analysis: discourse-based and aspect-based.

a) In a discourse-based approach not all the sentences have the same importance. Several researchers have tried to measure the contribution of sentences or phrases to the polarity of the text. Discourse Structure based Sentiment Analysis is divided in two approaches: rule based and weight based ones. In both approaches the results have improved with the addition of discourse relations. Moreover, these works have shown that the combination of two paradigms can bring an overall improvement.

In the rule based approach, Somasundaran et al. (2009) use a supervised collective classification and a supervised optimization framework in order to improve polarity classification. Text spans are extracted according to their importance in discourse structure.

Vanzo, Croce, and Basili (2014) assign a sentiment polarity to entire tweet sequences using a Markovian formulation of the Support Vector Machine discriminative model, SVM_{hmm}. In contextual information, they take into account two aspects: the conversation and the user attitude or the overall attitude of the last tweets. The individual perspective is independent in context, so they consider the tweet as a multifaceted entity. Consequently, each vector contributes in one aspect of the overall representation. The evaluation shows that sequential tagging effectively improves the detection precision approximately 20% in F1 measure.

In the weight based approach, Polanyi and Zaenen (2006) demonstrate that the structure of the text gives important information to extract the opinion. They have found that connectors increase or decrease the intensity of polarity. In this way, discourse relations can also increase or decrease the intensity.

Inspired in this previous work, Taboada, Voll, and Brooke (2008) extract the most important spans of a text and then, they calculate the semantic orientation in two ways, where the most important spans weight more. First, they use RST and they extract all the nuclei of the text. After that, they use a topic classifier based in support vector machines, improving results.

The rhetorical information of a text can be extracted automatically with a discourse parser and then this information can be used in sentiment analysis to determine the polarity of the text in a more reliable way. For example, Taboada, Voll, and Brooke (2008) and Heerschop, Goossen, and Hogenboom (2011) use the Sentence-level PArsing of DiscoursE (SPADE) tool in order to extract the discourse relations automatically from the text (Soricut and Marcu, 2003).

b) In Aspect-based Sentiment Analysis, the subject of a review is important because it helps to predict the "relevant" polarity expressed in the text. In this way, words related with the aspect help to give more accurate results.

Lim and Buntine (2014) combine language model and aspect creating the LDA-based[8] Twitter Opinion Topic Model. This model uses strong sentiment words, hashtags, mentions and emoticons to predict the opinion modeling the target.

The OpeNER project (Opener, 2013) makes use of three components: *i*) opinion express, *ii*) opinion holder and *iii*) opinion target. There are four tag levels in the Annotation Tool of the project. Tagging is based in three parameters: positive / negative attitude, sentence "on-topic" and "to-the-point"[9]. In the project, topic sentence can be a touristic attraction, a restaurant or a hostel. The opinions indirectly linked with the reviewed entity are also considered as "on-topic". The "to-the-point" category implies that a reviewer gives an opinion of the annotation object and then expresses a lot of details of it.

The Replab project developed the *Reputation online*. Spina, Gonzalo, and Amigó (2014) added Twitter signals to content sig-

[8]Latent Dirichlet Allocation is a opinion model used for Opinion Mining

[9]The OpeNER project can be consulted at `http://www.opener-project.eu/`.

nals to improve topic detection and they learnt a similarity function in order to supervise the topic detection clustering process. The last aim of this work is to solve the reputation monitoring problem automatically. They made use of different entities (*Maroon 5, Yamaha, Ferrari*) in the task. The difference with the current work is that their text is unordered (tweets) while ours is ordered (literary criticism).

Our method is a combination of two approaches: based on discourse structure and based on aspect. On the one hand, our approach is based on discourse structure because we will use RST in order to put different weights to Elementary Discourse Units (EDUs) according to their position in a discourse tree. On the other hand, our approximation is also aspect-based because we want to identify the words related with the main topic in order to get better results.

We think that the implementation of discourse structure together with the QWN-PPV method can improve the results. In other words, the polarity of the text will be better assessed. In this paper, we will do a first approach analyzing discourse topic and its influence on structures of attitude. In the following Example (4), we show the results of the QWN-PPV tool on the whole AIZ02 text.

(4) Number of words containing sentiment found: 7
Polarity score: 0.22
Polarity (if threshold > 0.0): positive
Gustura$_{(+)}$ *irakurtzen da nobela, protagonistaren*$_{(+)}$ *joko*$_{(+)}$ *bikoitza nola bukatuko ote den, nahiz eta amaiera horren zantzuak aurretik eskaintzen*$_{(+)}$ *dizkigun*$_{(+)}$ *idazleak. Idazkerak ere laguntzen*$_{(+)}$ *du aurrera plazerez*$_{(+)}$ *egiten.* AIZ02
The novel is read with pleasure$_{(+)}$, how is going to finish the$_{(+)}$ double$_{(+)}$ game$_{(+)}$ of$_{(+)}$ the$_{(+)}$ protagonist$_{(+)}$, although the narrator previously gives$_{(+)}$ us$_{(+)}$ some clues of the ending. Writing also helps to go along with pleasure$_{(+)}$.

The QWN-PPV method determines any positive word with a positive value (+1) and any negative word with a negative value (−1). Then, a polarity score (0.22) is estimated for the text. To do so, both positive and negative words are counted and divided by the number of the words in the text. If there are more positive words, as in Example (4), the polarity score will be higher (0.23) than the threshold (zero) and, therefore, the overall polarity of the text will be positive.

On the other hand, the QWN-PPV method estimates a lower polarity score (0.11) for all the AIZ02 text. So, the example shows how coherence relations related to the discourse topic can contribute to a better assignment of the text polarity.

3 Methodology

We have used the Rhetorical Structure Theory (RST) (Mann and Thompson, 1988) to achieve the rhetorical information of the text. The main concepts of RST are nuclearity of text spans[10] and coherence relations among text spans. With these two concepts a hierarchical tree structure (RS-tree) of coherence can be build, where all the text spans have a function in the tree, because relations are recursive (one relation can work as one of the spans in another relation). RST relations are hypotactic —hierarchical relations with one nucleus (N) and one satellite (S), e.g.: ELABORATION, JUSTIFY, EVALUATION, CAUSE. . . — and paratactic —discourse coordination where all the discourse units are nucleus, e.g. CONTRAST, DISJUNCTION, CONJUNCTION. . . [11]

In a hierarchical RS-tree there is always a Central Unit which is the most salient EDU (Iruskieta, Diaz de Ilarraza, and Lersundi, 2014). For example, in scientific abstracts authors often show explicitly the discourse topic as follows: “the principal aim of this paper is to investigate. . . ” (Paice, 1980).

To show an example of the discourse structure we want to use, a partial RS-tree of AIZ02 in Figure 1 is presented. The central unit of the tree structure is represented with straight vertical lines (the unit 2-2 in the example). The annotator interpreted the RS-tree as follows:

a) PREPARATION for the article, by means of the title ([1−1 > 2−10]).

[10] Discourse structure is recursive so there are formally two different units: *a*) Elementary Discourse Unit and *b*) group of EDUs.

[11] A more detailed explanation of RST can be found at `http://www.sfu.ca/rst/` and in Basque at `http://www.sfu.ca/rst/07basque/index.html`.

b) with the highest EVALUATION linked to the central unit she interprets that it is evaluating all the propositions mentioned before, that can be taken as all the work ([2−14 < 15−20]).

c) with the lowest EVALUATION linked to the central unit, she is evaluating an aspect of the work, only the proposition mentioned in the central unit ([6−6 < 7−7]).

These are the steps taken to carry out this study:

i) Building a corpus. We have collected a corpus of 28 texts, where 19 of them will be used for training and the remaining 9 for testing.[12] The texts are reviews of Basque Literature works. The size of the texts is not uniform: the shortest one has 106 words and the longest one 485 words. A corpus description is presented in Table 1 and the annotated corpus can be consulted in the Basque RST Treebank at `http://ixa2.si.ehu.es/diskurtsoa/en/` (Iruskieta et al., 2013).

Text	**Doc.**	**EDUs**	**Words**
CRITICS	28	1038	8823

Table 1: Corpus description

ii) EDU segmentation of texts. Before preparing the experiment, we have processed the texts. Firstly, we used EusEduSeg (Iruskieta and Zapirain, 2015) a discourse segmenter to segment all the texts automatically.[13] After that, the segmentation has been corrected manually to avoid losing rhetorical information in subsequent phases.

iii) Corpus annotation. After segmentation, we have annotated the most salient EDU or the central unit and after it we have tagged all rhetorical structures of the text with the RSTTool (O'Donnell, 1997) using the Basque extended relations of RST.

iv) Central Unit (CU) and Polarity gold standard. To do so, we have made up a questionnaire based on Google Forms, where 20 annotators participated in the annotation. This was done in order to have a gold standard which we could use to compare the results.
Our gold standard was collected as follows: *i*) the central unit must be selected at least by four participants. If not, we selected the three most voted central units.[14] *ii*) The polarity of each text was conformed with the average of all the annotators, in two ways:
- Polarity 1 (P1): quantitative polarity annotation from 1 to 5.
- Polarity 2 (P2): qualitative polarity description with three values: negative, neutral and positive.

v) Manual extraction of text spans composed with the text of the central unit and the EVALUATION relation. We have manually built different features based on the rhetorical structure tree:
- ALL (F1): the result of QWN-PPV on the full text.
- CU (F2): only the central unit.
- CU-H-EV (F3): the central unit and the highest EVALUATION relation linked to it.
- CU-ALL-EV (F4): the central unit and all the EVALUATION relations linked to it.

Table 2 shows all the glosses we have used to perform the analysis.

Gloss	**Meaning**
P1	Polarity of five categories
P2	Polarity of three categories
F1	QWN-PPN for all the text
F2	The central unit
F3	The central unit and the highest EVALUATION relation
F4	All the EVALUATION relations of the text

Table 2: Glosses of the predicted variables and features.

vi) Lemmatization and polarity extraction with the QWN-PPV. Before running QWN-PPV, we run Eustagger (Ezeiza et al., 1998) to divide the sentences into unambiguous tokens.[15]
After that, we run the QWN-PPV

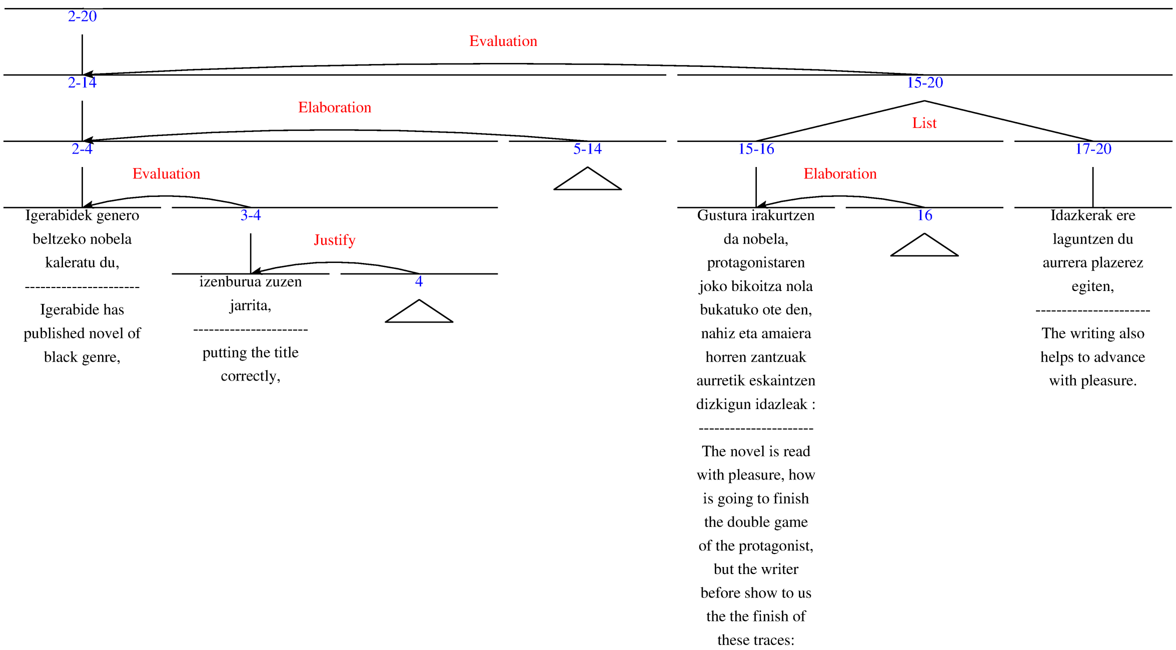
[12] All the texts are available in *Kritiken Hemeroteka* at `http://kritikak.armiarma.eus/`

[13] EusEduSeg can be tested at `http://ixa2.si.ehu.es/EusEduSeg/EusEduSeg.pl`.

[14] Hearst (1997) considered that a subtopic boundary was true if at least three out of seven (42.86%) annotators placed a boundary mark.

[15] Eustagger is a lemmatizer and tagger for Basque based on Stochastic and Rule-Based Methods. Eustagger can be tested at `http://ixa2.si.ehu.es/demo/analisimorf.jsp`.

Figure 1: A partial RS-tree of AIZ02



method (Vicente, Agerri, and Rigau, 2014) and obtained the polarity for each of the four features.

vii) Analysis of results. We have used two methods in order to analyze the results: *Logistic Regression* (LR) and *Sequential Minimal Optimization* (SMO).
They are well known though efficient techniques that have often been used as baseline. The first is adequate for regression problems as in this case, where the P1 class is a numeric polarity from 1 to 5. The second one tackles classification, that is, the class to guess (P2) is nominal and it was obtained by a straightforward discretization mechanism (positive, negative and neutral). Both methods are implemented with open libraries.
We calculate percent agreement and precision, recall and f-measure as follows:

$$precision = \frac{correct_{polarity}}{correct_{polarity} + excess_{polarity}}$$

$$recall = \frac{correct_{polarity}}{correct_{polarity} + missed_{polarity}}$$

$$F_1 = \frac{2 * precision * recall}{precision + recall}$$

where $correct_{polarity}$ is the number of correct *polarity items*, $excess_{polarity}$ is the number of overpredicted *polarity items* and $missed_{polarity}$ is the number of *polarity items* the system missed to tag.

A summary of the methodology we have employed is presented in Figure 2.

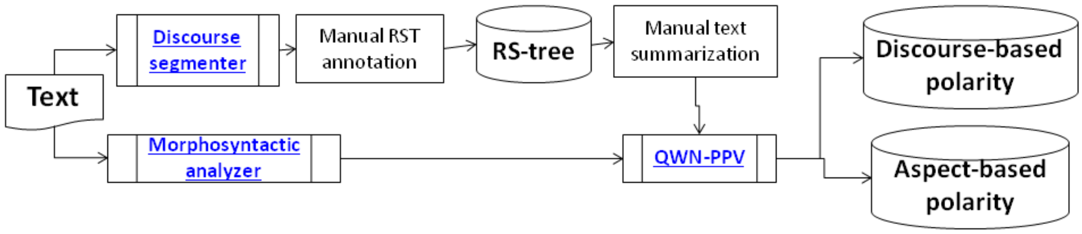
4 Results

Firstly, we present the results of all the features when trying to guess P1 (five categories). The results of each feature and the best combinations are presented in Table 3.

When individual features were considered, F1 with LR obtained the best results (0.37), while F2, F3 and F4 obtain lower results, with F4's contribution near to zero. But when combinations were considered, using features F1, F3 and F4 together (F134) with SMO obtained a better result (0.40). The results show that, in guessing P1, there is a gain employing combinations of features based on discourse structure.

Secondly, we will test the same algorithms to try to guess P2, based on three categories. The results are presented in Table 4. When using individual features, F1 and F2 with LR obtain the best results (0.47). Looking at the combinations, F1234 with SMO gives the

Figure 2: System architecture



Method	Feature	Fm
LR	**F1**	**0.37**
	F2	0.21
	F3	0.32
	F4	-
	F134	0.28
	F1234	0.30
SMO	**F1**	0.23
	F2	0.19
	F3	0.27
	F4	-
	F134	**0.40**
	F1234	0.38

Table 3: Results guessing P1 (five categories, cross-validation on the development set).

Feature	Method	Fm
LR	**F1**	**0.47**
	F2	**0.47**
	F3	0.44
	F4	-
	F134	**0.52**
	F1234	0.39
SMO	**F1**	0.37
	F2	0.36
	F3	0.36
	F4	-
	F134	0.50
	F1234	**0.53**

Table 4: Results on P2 (three categories, cross-validation on the development set).

best results (0.53). Overall, the results show that when using discourse structure (combinations), the results on P2 improve considerably.

To sum up, in the case of P1 with five categories (see Table 3) SMO is the best single algorithm for prediction. In contrast, SMO gave the best results when using a combination of features, with an F-measure of 0.40.

In the case of P2 with three categories (see Table 4), LR continues to be the best method using a single feature with an F-measure of 0.47. In combinations, SMO gives the best results (0.53).

After examining the results on the development set, we test the best methods (LR on the individual features and SMO on the combinations) on the test set. We show the results in Table 5.

	Feature	Method	Fm
P1	**LR**	**F1**	0.09
	SMO	**F134**	0.09
P2	**LR**	**F1**	0.59
	LR	**F134**	**0.84**
	SMO	**F1**	0.40
	SMO	**F1234**	0.73

Table 5: Test set results for P1 and P2

When guessing P1, the best results are obtained with F1 using LR and H134 using SMO. That means that the algorithms based in discourse structure we have used are not able to guess P1 accurately, possibly because the small size of the corpus to deal with five categories. In contrast, the results to guess P2 (three categories) using combinations of features based on discourse structure are considerably better than considering all the text (F1), with 0.84 and 0.73 for LR and SMO, respectively. Therefore, it seems that the implementation of discourse features can improve the results in opinion mining.

Performing a first error analysis, the confusion matrix of F134 guessing P2 with SMO shows that a text with neutral polarity has been classified as having a negative polarity. The error is not specially important, as a matter of fact, the QWN-PPV method puts only positive or negative polarity to the texts. So, the difference is that the method has two main categories and we use three categories.

If we analyze Example (5), we can see that the Central Unit of the text is neutral but the method considers it a negative text.

(5) *Aho gustu*$_{(+)}$ *gazi-gozoa utzi*$_{(*)}$ *dit*$_{(+)}$ *[...]ren bigarren ipuin liburuak.* [...] BER03
The second storybook of [...] has$_{(+)}$ left$_{(*)}$ me a sweet-and-sour taste$_{(+)}$ in the mouth.[...]

In the example, the word *gazi-gozoa* 'sweet-and-sour' is a neutral word but the QWN-PPV does not detect it. Moreover, some words of the remaining text are not tagged with their correct polarity.

5 Conclusions and future work

We have presented a set of algorithms in which we have tried to examine the importance of discourse structure information in Opinion Mining. Firstly, we have concluded that guessing the polarity of the text of literary reviews based on three categories gives better results than the one with five categories for Opinion Mining. This could be due to the fact that the task is easier and also to the reduced size of the training set.

The second conclusion is that combining several discourse structures is the best method for Logistic Regression giving an F-measure of 0.84 (and also for SMO with an F-measure of 0.73). An error analysis has shown that the errors were soft: a text with neutral polarity has been misclassified as having negative polarity. Looking at the importance of each individual feature, we can say that an important weight related to polarity is situated in the EVALUATION discourse relation.

In future works, our aim is to:

- Annotate a bigger corpus. The perliminary experiments performed in this work should be validated using a bigger corpus.
- Implement a full set of experiments on combinations of the central unit and the EVALUATION relation. We want to study if there is any difference with the relations which are attached to the central unit and the relations which are not.
- Test other phenomena of discourse structure such as the nuclearity (satellite vs. nucleus) and INTERPRETATION relations, to check if they have any influence on polarity.
- Automatize all the system, by testing in partial RS-trees, where only the central unit and EVALUATION relations linked to it were considered.
- Study which syntactic and discourse structures are more important (i.e., they change the polarity of lower levels).
- Implement an automatic annotator of word level polarity based on a supervised dictionary, to solve some problems observed in the QWN-PPV.

Bibliografía

[Alistair and Diana2005] Alistair, Kennedy and Inkpen Diana. 2005. Sentiment classification of movie and product reviews using contextual valence shifters. *Proceedings of FINEXIN*.

[Asher1993] Asher, Nicholas. 1993. *Reference to abstract objects in discourse*, volume 50. Springer Science & Business Media.

[Cambria et al.2013] Cambria, Erik, Bjorn Schuller, Yunqing Xia, and Catherine Havasi. 2013. New Avenues in Opinion Mining and Sentiment Analysis. In *IEEE Intelligent Systems*, volume 28, pages 15–21, Piscataway, NJ, USA, March.

[Esuli and Sebastiani2010] Esuli, A. and F. Sebastiani. 2010. SentiWordNet: A publibly available lexical resource for opinion mining. In *Proceedings of 5th International Conference on LREC*, pages 417–422.

[Ezeiza et al.1998] Ezeiza, Nerea, Itziar Aduriz, Iñaki Alegria, Jose Mari Arriola, and Ruben Urizar. 1998. Combining Stochastic and Rule-Based Methods for Disambiguation in Agglutinative Languages. In *COLING-ACL'98*, volume 1, pages 380–384, Canada, August 10-14.

[Hatzivassiloglou and McKeown1997] Hatzivassiloglou, Vasileios and Kathleen R McKeown. 1997. Predicting the semantic orientation of adjectives. In *Proceedings of the 35th annual meeting of the association for computational linguistics and eighth conference of the european chapter of the association for* *computational linguistics*, pages 174–181. Association for Computational Linguistics.

[Hearst1997] Hearst, M. 1997. TextTiling: Segmenting Text into Multi-Paragraph Subtopic Passages. In *Computational Linguistics*, volume 23, pages 33–64, March.

[Heerschop, Goossen, and Hogenboom2011] Heerschop, B., F. Goossen, and A. Hogenboom. 2011. Polarity Analysis of Texts using Discourse Structure. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, pages 1061–1070, Glasgow, Scotland, UK, October 24–28.

[Iruskieta et al.2013] Iruskieta, Mikel, María Jesus Aranzabe, Arantza Diaz de Ilarraza, Itziar Gonzalez, Mikel Lersundi, and Oier Lopez de la Calle. 2013. The RST Basque TreeBank: an online search interface to check rhetorical relations. In *4th Workshop "RST and Discourse Studies"*, Brasil, October 21-23.

[Iruskieta, Diaz de Ilarraza, and Lersundi2014] Iruskieta, Mikel, Arantza Diaz de Ilarraza, and Mikel Lersundi. 2014. The annotation of the central unit in rhetorical structure trees: A key step in annotating rhetorical relations. In *COLING*, pages 466–475. Dublin City University and ACL, Dublin, Ireland.

[Iruskieta and Zapirain2015] Iruskieta, Mikel and Beñat Zapirain. 2015. EusEduSeg: A Dependency-Based EDU Segmentation for Basque. In *SEPNL*, Alicante, September 16-18.

[Lim and Buntine2014] Lim, Kar Wai and Wray Buntine. 2014. Twitter opinion topic model: Extracting product opinions from tweets by leveraging hashtags and sentiment lexicon. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pages 1319–1328. ACM.

[Mann and Thompson1987] Mann, William C and Sandra A Thompson. 1987. *Rhetorical structure theory: A theory of text organization.* University of Southern California, Information Sciences Institute.

[Mann and Thompson1988] Mann, William C. and Sandra A. Thompson. 1988. Rhetorical structure theory: Toward a functional theory of text organization. In *Text*, volume 8, pages 243–281.

[Miltsakaki et al.2005] Miltsakaki, Eleni, Nikhil Dinesh, Rashmi Prasad, Aravind Joshi, and Bonnie Webber. 2005. Experiments on sense annotations and sense disambiguation of discourse connectives. In *Proceedings of the Fourth Workshop on Treebanks and Linguistic Theories (TLT2005), Barcelona, Spain, December.*

[O'Donnell1997] O'Donnell, Michael. 1997. Rst-tool: An rst analysis tool. In *Proceedings of the 6th European Workshop on Natural Language Generation*, Duisburg, Germany.

[Opener2013] Opener. 2013. OpeNER annotation guidelines. Annotation of Costumer Reviews in the Touristic Domain V5.0. pages 1–19.

[Paice1980] Paice, Chris D. 1980. The automatic generation of literature abstracts: an approach based on the identification of self-indicating phrases. In *3rd annual ACM conference on Research and development in information retrieval*, pages 172–191, Cambridge, June. Butterworth and Co.

[Pang and Liu2008] Pang and N. Liu. 2008. Opinion Mining and Sentiment Analysis. In *Foundations and trends in information retrieval*, volume 2, pages 1–135.

[Pang and Lee2004] Pang, B. and L. Lee. 2004. A sentimental education: Sentiment analysis using subjectivity summarization based on minimun cuts. In *Proceedings of the 42nd annual meeting on Association for Computational Linguistics*, page 271.

[Pang, Lee, and Vaithyanathan2002] Pang, Bo, Lillian Lee, and Shivakumar 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*, pages 79–86. Association for Computational Linguistics.

[Polanyi and Zaenen2006] Polanyi, Livia and Annie Zaenen. 2006. Contextual valence shifters. In *Computing attitude and affect* *in text: Theory and applications*. Springer, pages 1–10.

[Somasundaran et al.2009] Somasundaran, Swapna, Galileo Namata, Janyce Wiebe, and Lise Getoor. 2009. Supervised and unsupervised methods in employing discourse relations for improving opinion polarity classification. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*, pages 170–179. Association for Computational Linguistics.

[Soricut and Marcu2003] Soricut, Radu and Daniel Marcu. 2003. Sentence level discourse parsing using syntactic and lexical information. In *Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*.

[Spina, Gonzalo, and Amigó2014] Spina, Damiano, Julio Gonzalo, and Enrique Amigó. 2014. Learning similarity functions for topic detection in online reputation monitoring. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pages 527–536. ACM.

[Taboada, Voll, and Brooke2008] Taboada, M., K. Voll, and J. Brooke. 2008. Extracting sentiment as a function of discourse structure and topicality. In *Simon Fraser Univeristy School of Computing Science Technical Report*.

[Taboada and Mann2006] Taboada, Maite and William C Mann. 2006. Applications of rhetorical structure theory. *Discourse studies*, 8(4):567–588.

[Trnavac and Taboada2014] Trnavac, Radoslava and Maite Taboada. 2014. Discourse structure and attitudinal valence of opinion words in sentiment extraction. *LSA Annual Meeting Extended Abstracts*.

[Tumasjan et al.2010] Tumasjan, A., T. O. Sprenger, P. G. Sandner, and I. T. Welpe. 2010. Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. In *ICWSM*, number 10, pages 178–185.

[Vanzo, Croce, and Basili2014] Vanzo, Andrea, Danilo Croce, and Roberto Basili. 2014. A context-based model for Sentiment Analysis in Twitter. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 2345–2354, Dublin, Ireland, August.

[Vicente, Agerri, and Rigau2014] Vicente, Iñaki San, Rodrigo Agerri, and German Rigau. 2014. Simple, Robust and (almost) Unsupervised Generation of Polarity Lexicons for Multiple Languages. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2014)*, Gothenburg, Sweden, April 26-30.

[Yu and Hatzivassiloglou2003] Yu, Hong and Vasileios Hatzivassiloglou. 2003. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of the 2003 conference on Empirical methods in natural language processing*, pages 129–136. Association for Computational Linguistics.

[Zhou et al.2011] Zhou, Lanjun, Binyang Li, Wei Gao, Zhongyu Wei, and Kam-Fai Wong. 2011. Unsupervised discovery of discourse relations for eliminating intrasentence polarity ambiguities. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 162–171. Association for Computational Linguistics.

[Zhou2013] Zhou, Yudong. 2013. Fine-grained Sentiment Analysis with Discourse Structure. In *Master Thesis. Saarland University*, October.