Negation Detection for Sentiment Analysis:
A Case Study in Spanish

S A L U D   M A R Í A   J I M É N E Z - Z A F R A,[1] N O A   P.   C R U Z - D Í A Z,[2]

M A I T E   T A B O A D A,[3] M A R Í A   T E R E S A   M A R T Í N - V A L D I V I A[1]

[1]*SINAI, Centro de Estudios Avanzados en TIC (CEATIC), Universidad de Jaén, Spain*
[2]*Centro de Excelencia de Inteligencia Artificial, Bankia, Madrid, Spain*
[3]*Discourse Processing Lab, Simon Fraser University, Burnaby, BC, Canada*
sjzafra@ujaen.es, contact@noacruz.com, mtaboada@sfu.ca, maite@ujaen.es

Abstract

Accurate negation identification is one of the most important tasks in the context of sentiment analysis. In order to correctly interpret the sentiment value of a particular expression, we need to identify whether it is in the scope of negation. While much of the work on negation detection has been focused on English, we have seen recent developments that provide accurate identification of negation in other languages. In this paper, we provide an overview of negation detection systems and describe an implementation of negation cue and negation scope detection for Spanish. We apply this system to the sentiment analysis task, showing that improvements can be gained from accurate negation detection. The paper contributes an implementation of negation detection for sentiment analysis in Spanish and a detailed error analysis. This is the first work in Spanish in which a machine learning negation processing system is applied to the sentiment analysis task. Existing methods have used negation rules that have not been assessed, perhaps because the first Spanish corpus annotated with negation for sentiment analysis has only recently become available.

1 Introduction

Negation is a complex phenomenon in natural language. It usually changes the polarity of a sentence, creating an opposition between positive and negative counterparts of the same sentence (Horn, 1989). Negation is primarily a syntactic phenomenon, but it also has pragmatic effects, leading to asymmetry in the effect of positive and negative statements (Israel, 2004; Potts, 2011a) and to difficulties in interpretation, especially in Natural Language Processing (NLP) systems (Blanco and Moldovan, 2014).

Four tasks are usually performed in relation to negation processing: i) negation cue detection, in order to find the words that express negation; ii) scope identification, to find which parts of the sentence are affected by the negation cues; iii) negated event recognition, to determine which events are affected by the negation cues; and iv) focus detection, to find the part of the scope that is most prominently negated.

Processing negation is relevant for a wide range of NLP applications, such as information retrieval (Liddy et al., 2000), information extraction (Savova et al., 2010), machine

To appear in *Natural Language Engineering*. Current version: May 19, 2020.

translation (Baker et al., 2012) or sentiment analysis (Liu, 2015; Wiegand et al., 2010; Kennedy and Inkpen, 2006; Benamara et al., 2012). In this paper we focus on treating negation for sentiment analysis, specifically, for the polarity classification task, which aims to determine the overall sentiment-orientation (positive, negative or neutral) of the opinion given in a document.

In this work, we study the use of a negation cue (*no, not, n't*) to change the polarity of a sentence. The cue affects a part of the sentence, labelled as the scope. For instance, in Example (1) the negation cue *n't* negates the adjective *scary*, establishing an opposition with *enthralling*.

(1)   It isn't scary, but it is enthralling.

Negation presents specific challenges in NLP. Despite great strides in recent years in detecting negation cues and their scope (Vincze et al., 2008; Councill et al., 2010; Morante and Sporleder, 2012), many aspects of negation are still unsolved. For instance, Blanco and Moldovan (2014) provide numerous examples of how difficult it is to correctly interpret implicit positive meaning, using examples such as Example (2) and (3). In the first case, the implicit negative meaning is that cows eat something other than meat, i.e., that the negation only affects *meat,* not *eat.* In the second example, the implication is that cows eat, and that they eat grass, but do so with something other than a fork (or that they don't use utensils at all).

(2)   Cows don't eat meat.

(3)   Cows don't eat grass with a fork.

Existing methods for detecting negation and—the most difficult part—its scope, can be classified into those that are rule-based and those that rely on some form of machine-learning classifiers, work which we will review in Section 2.

Negation is an interesting research topic in its own right. At the same time, processing negation has been pursued as a way to improve applications in NLP. In the well-developed field of biomedical text mining, the detection of negation is crucial to understanding the meaning of a text (e.g., did the patient improve after treatment or not). In sentiment analysis, the object of our study, negation also plays a vital role in accurately determining the sentiment of a text.

A great deal of the research on negation, whether on its processing or application, has focused on English. However, the study of this phenomenon in languages other than English is a necessity, since negation is a language-dependent phenomenon. Although general concepts related to negation, such as negation cue, scope, event and focus, can probably be applied to all languages, the morphological, syntactic, semantic and pragmatic concrete mechanisms to express them vary depending on the language (Payne, 1997). Therefore, each language requires a specific way of treating negation.

In this paper, we address negation in Spanish, starting with existing research in English and highlighting the slightly different methods that are needed to capture the expression of negation in Spanish. We describe our implementation of a state-of-the-art negation processing method for Spanish (Jiménez-Zafra et al., 2020), based on our previous work on

English and its results. The implementation is then applied to the task of sentiment analysis by integrating the negation system in a well-known lexicon-based sentiment analysis system, the Semantic Orientation CALculator[1] (SO-CAL). Finally, a detailed analysis of different types of errors is performed, which can be attributed either to the negation processing or to the sentiment analysis system.

This is, to our knowledge, the first work in Spanish in which a machine learning negation processing system is applied to the sentiment analysis task. Existing methods for Spanish sentiment analysis have used negation rules that have not been assessed, perhaps because the first Spanish corpus annotated with negation for sentiment analysis have only recently become available. Comparison with previous works is not possible because there is no work that incorporates a negation processing system into SO-CAL. Therefore, we have used as baselines SO-CAL without negation and SO-CAL with built-in negation.

The rest of the paper is organized as follows. In Section 2, related research on negation detection and its application to sentiment analysis is outlined. Section 3 and Section 4 present the corpus and method adopted for processing negation in Spanish, respectively. In Section 5, it is described how the negation detector is integrated in a well-known sentiment analysis system, SO-CAL. Moreover, the different experiments conducted in order to evaluate the effect of accurate negation detection and the error analysis are also provided in this section. Finally, Section 6 summarizes the conclusions and future work that suggest possible avenues of research in order to improve the systems we describe.

2 Related work

This section reviews relevant literature related to processing of negation and its application to improve sentiment analysis. Since much of this work has been carried out for English, the language taken as a reference, we first provide an overview of negation detection for English, before reviewing work for Spanish.

2.1 Negation detection for English

Negation detection in English has been a productive research area during recent years in the NLP community as shown by the challenges and shared tasks held (e.g., BioNLP'09 Shared Task 3 (Kim et al., 2009), i2b2 NLP Challenge (Uzuner et al., 2011), *SEM 2012 Shared Task (Morante and Blanco, 2012) and ShARe/CLEF eHealth Evaluation Lab 2014 Task 2 (Mowery et al., 2014)). It is worth noting that the initial solutions that arose for sentiment analysis are not accurate enough since they have relatively straightforward conceptualizations of the scope of negation and have traditionally relied on rules and heuristics. For example, Pang and Lee (2004) assumed that the scope of a negation keyword consists of the words between the keyword and the first punctuation mark following it (see also Polanyi and Zaenen, 2006).

The first work we are aware of detecting negation and its scope using a more robust approach is presented by Jia et al. (2009). They develop a rule-based system that uses information derived from a parse tree. This algorithm computes a candidate scope, which is

---

[1] https://github.com/sfu-discourse-lab/SO-CAL

then pruned by removing the words that do not belong to the scope. Heuristic rules, which include the use of delimiters (i.e., unambiguous words such as *because*) and conditional word delimiters (i.e., ambiguous words like *for*), are used to detect the boundaries of the candidate scope. Situations in which a negation cue does not have an associated scope are also defined. The authors evaluate the effectiveness of their approach on polarity determination showing that the identification of the scope of negation improves both the accuracy of sentiment analysis and the effectiveness of opinion retrieval.

Regarding the impact of negation identification on sentiment analysis using machine learning techniques, this has not been sufficiently investigated. As Díaz and López (2019) point out, this is perhaps because reasonably sized standard corpora annotated with this kind of information have only recently become available. However, there is relevant work that shows the suitability of applying negation modelling to the task of sentiment analysis in other languages and that, therefore, has inspired the experimentation presented in this paper. For example, Councill et al. (2010) develop a system that can precisely recognize the scope of negation in free text. The cues are detected using a lexicon (i.e., a dictionary of 35 negation keywords). A Conditional Random Field (CRF) algorithm is used to predict the scope. This classifier incorporates, among others, features from dependency syntax. The approach is trained and evaluated on a product review corpus. Using the same corpus, Lapponi et al. (2012) present a state-of-the-art system for negation detection. Their proposal is based on the application of CRF models for sequence labelling, which makes use of a wealth of lexical and syntactic features, together with a fine-grained set of labels that capture the scopal behaviour of tokens. With this approach, they also demonstrate that the choice of representation has a significant effect on performance. Cruz et al. (2016) also conduct research into machine learning techniques in this field. They define a system which automatically identifies negation cues and their scope in the Simon Fraser University (SFU) Review corpus (Konstantinova et al., 2012), showing results in line with the results of other authors in the same task and domain.

Another type of approaches worth mentioning are the composition models that implicitly learn negation. For instance, Socher et al. (2013) generate the Stanford Sentiment Treebank corpus and apply Recursive Neural Tensor Networks to it, improving the state-of-the-art in single sentence positive/negative classification. This model also accurately captures the effects of negation and its scope at various tree levels for both positive and negative phrases.

Deep learning techniques have also been applied to the task of negation detection. Work by Fancellu et al. (2016) and Qian et al. (2016), although not focused on the sentiment analysis domain, should be highlighted. Fancellu et al. (2016) present two different neural network architectures, i.e., a hidden layer feed-forward neural network and a bidirectional Long Short-Term Memory (LSTM) model. Training, development and tests are done on the negative sentences of the Conan Doyle corpus (Morante and Blanco, 2012). The results show that neural networks perform on a par with previously developed classifiers. Qian et al. (2016) propose a Convolutional Neural Network (CNN)-based model with probabilistic weighted average pooling to address negation scope detection. This system first extracts path features from syntactic trees with a convolutional layer and concatenates them with their relative positions into one feature vector, which is then fed into a soft-max layer to compute the confidence scores of its location labels. It is trained on the abstract sub-

collection of the BioScope corpus, achieving the second highest performance for negation scope on abstracts.

In the field of sentiment analysis, one of the latest published works proposes a multi-task approach to explicitly incorporate information about negation (Barnes et al., 2019). Similarly to Fancellu et al. (2016), this system consists of a BiLSTM-based model, relying only on word embeddings as input, but also adding a CRF for the prediction layer. This configuration outperforms learning negation implicitly in a data-driven manner and shows that explicitly training the model with negation as an auxiliary task helps improve the main task of sentiment analysis.

In all, what negation detection for English has unveiled is that it is a complex phenomenon. Rules and heuristics for detecting scope significantly improve the accuracy of sentiment analysis systems. A full understanding of negation scope and the role of negation in evaluative language is still elusive. Most recent methods with deep learning show promise, but more annotated corpora are needed.

2.2 Negation detection for Spanish

Negation processing in NLP for Spanish has started relatively recently compared to English. We find systems such as those proposed by Costumero et al. (2014), Stricker et al. (2015) and Cotik et al. (2016) aimed at automatically identifying negation in the clinical domain by adapting the popular rule-based algorithm NegEx (Chapman et al., 2001), which uses regular expressions to determine the scope of some negation cues.

In the review domain, negation has also been taken into account for Spanish sentiment analysis. One of the first systems, developed by Brooke et al. (2009), adapts an English lexicon-based sentiment analysis system, SO-CAL (Taboada et al., 2011), to Spanish. The original implementation uses simple rules and heuristics for identifying the scope of negation. It is this method that we improve upon for this paper.

More sophisticated is the work of Vilares et al. (2013, 2015), which incorporates dependency parses to better pinpoint the scope of several intensifiers and negation cues.[2] Their results show that taking into account the syntactic structure of the text improves accuracy in the review domain, whether the sentiment analysis system uses machine learning or lexicon-based approaches. They do not, however, analyze the gain obtained using negation individually (separate from intensification) and, therefore, it is not possible to determine the relative contribution of negation by itself to the improvement obtained.

Jiménez-Zafra et al. (2015) study the most important cues[3] according to La Real Academia Española (Bosque et al., 2009) and propose a set of rules based on dependency trees for identifying the scope of these negation cues. Later, Jiménez-Zafra et al. (2017) apply the detection negation module for sentiment analysis on Twitter, incorporating some changes to address the peculiarities of the language used in this social medium. They also use a lexicon-based system and statistically demonstrate that the results obtained

---

[2] Vilares et al. (2015) study the negation cues *no* ('not'), *sin* ('without') and *nunca* ('never').

[3] The cues are: *no* ('not'), *tampoco* ('neither'), *nadie* ('nobody'), *jamás* ('never'), *ni* ('nor'), *sin* ('without'), *nada* ('nothing'), *nunca* ('never') and *ninguno* ('none').

considering the negation module are significantly higher than those obtained without taking negation into account. Moreover, they compare the proposed method with the method most used to determine the scope of negation in English tweets (Potts, 2011b), showing that the classification with their approach is better.

Other work that has clearly demonstrated the benefits of negation detection for Spanish sentiment analysis includes Mitchell et al. (2004) or Amores et al. (2016). In most works (Taboada et al., 2011; Vilares et al., 2013, 2015; Jiménez-Zafra et al., 2015; Miranda et al., 2016; Amores et al., 2016; Jiménez-Zafra et al., 2017), negation detection is applied, but without a detailed error analysis or without ablation experiments to determine the gains of negation and how errors in negation detection affect the accuracy of the sentiment analysis system. This is, in part, due to the lack of an annotated corpus for negation in the review domain.

However, after the annotation of the SFU Review$_{SP}$-NEG corpus (Jiménez-Zafra et al., 2018) and the organization of the 2018 and 2019 editions of NEGES (Jiménez-Zafra et al., 2019a,b), the Workshop on Negation in Spanish, we find some systems for processing negation in the review domain. The aim of this workshop is to promote the identification of negation cues in Spanish and the application of negation for improving sentiment analysis. For the negation cue detection task, six systems has been developed (Fabregat et al., 2018; Loharja et al., 2018; Giudice, 2019; Beltrán and González, 2019; Domínguez-Mas et al., 2019; Fabregat et al., 2019). All of them address the task as a sequence labeling problem using machine learning approaches. Deep learning algorithms and CRF algorithm are predominant, with CRF performing best.

Existing works addressing the clinical domain provide methods for the identification of negated entities and negated findings and those focusing on the review domain detect negation cues. None of them, however, focuses on the identification of the scope. To address this gap, we have recently developed a system for the identification of negation cues and their scopes in Spanish texts in which the SFU Review$_{SP}$-NEG corpus is also used (Jiménez-Zafra et al., 2020), which will be explained in detail in Section 4.

3 Data: The SFU Review$_{SP}$-NEG corpus

This section describes the corpus used in the experimentation, the SFU Review$_{SP}$-NEG corpus (Jiménez-Zafra et al., 2018), a Spanish review corpus annotated for negation cues and their scope. It is a labelled version of the Spanish SFU Review corpus. Both are comparable Spanish versions to their English counterparts. The English version of the corpus, the SFU Review corpus (Taboada et al., 2006) is a corpus extensively used in opinion mining (Taboada et al., 2011; Martínez-Cámara et al., 2013). It consists of 400 documents (50 of each type: 25 positive and 25 negative reviews) of movie, book, and consumer product reviews (i.e., cars, computers, cookware, hotels, music and phones) from the now-defunct website Epinions.com. This English corpus has several annotated versions (e.g., for appraisal and rhetorical relations), including one where all 400 documents are annotated at the token level with cues for negation and speculation, and at the sentence level with their linguistic scope (Konstantinova et al., 2012). The annotation guidelines follow closely the BioScope corpus guidelines (Vincze et al., 2008). The annotations of negation in English

have been used to develop a machine learning method for negation identification, which was then applied in a sentiment analysis system, SO-CAL (Cruz et al., 2016).[4]

The SFU Review$_{SP}$-NEG (Jiménez-Zafra et al., 2018)[5] is a Spanish corpus composed of 400 product reviews, 25 positive reviews and 25 negative reviews from eight different domains: cars, hotels, washing machines, books, cell phones, music, computers and movies. The reviews were collected to be comparable to the English version of the corpus, from Ciao.es, a site no longer available. Each review was automatically annotated at the token level with PoS-tags and lemmas using Freeling (Padró and Stanilovsky, 2012), and manually annotated at the sentence level with negation cues, their corresponding scopes and events, and how negation affects the words within its scope, that is, whether there is a change in the polarity or an increase or decrease of its value. In this corpus we distinguish four types of structures: *neg*, *noneg*, *contrast* and *comp*. The structures with a cue that negates the words in its scope have the label *neg*, whereas the other labels do not express negation (*noneg*, *contrast* and *comp*). These labels are necessary because words that are typically used as negation cues do not always act exclusively as such. They are also frequently used in rhetorical or tag questions (4), or in contrastive (5) and comparative (6) structures.

(4) Viniste a verlo, ¿**no**?
    You came to see him, didn't you?

(5) **No** hay **más** solución **que** comprar una lavadora.
    There is no other solution but to buy a washing machine.

(6) **No** me gusta **tanto como** lo otro.
    I don't like it as much as the other.

The corpus consists of 221,866 tokens and 9,446 sentences, out of which 3,022 (31.99%) were annotated with some of the structures mentioned before and out of which 2,825 (29.91%) contain at least one structure of type *neg*. Table 1 shows the distribution of the annotated structures in the corpus.

As we mentioned above, 2,825 sentences of the corpus contain at least one structure of type *neg*. We find sentences with one negation cue (2,028), two negation cues (578) and even three or more negation cues (219). In Table 2, we show the total and percentage of sentences by number of negations.

Negation cues in this corpus can be *simple*, if they are composed of a single token (e.g., *no* ('not'), *nunca* ('never')); *contiguous*, if they have two or more contiguous tokens (e.g., *casi no* ('almost not'), *en mi vida* ('never in my life')); or *non-contiguous*, if they consist of two or more non-contiguous tokens (e.g., *no-en absoluto* ('not-at all'), *no-nada* ('not-nothing')). Non-contiguous tokens are common in Spanish, as sentences with post-verbal negation words such as *nada* ('nothing') or *nunca* ('never') also have the enclitic *no* preceding the verb, e.g., *Ustedes no pueden hacer nada* ('You cannot do anything').

---

[4] The English components of the corpus, raw and annotated, are freely available at https://www.sfu.ca/~mtaboada/SFU_Review_Corpus.html.
[5] This version of the corpus is freely available at http://sinai.ujaen.es/sfu-review-sp-neg-2/

Table 1. Total annotated structures by type in the SFU Review$_{SP}$-NEG corpus

Type structure	Total	Example
neg	3,941	***No** tengo **nada** en contra de Opel*
noneg	181	*Viniste a verlo, no?*
contrast	175	***No** hay **más** solución **que** comprar una lavadora*
comp	30	***No** me gusta **tanto como** lo otro*

Table 2. Total and percentage of sentences by number of negations in the SFU Review$_{SP}$-NEG corpus

	# sent.	% sent.	Example
0 negations	6,621	70.09	*Lo que mejor funciona es el aire acondicionado*
1 negation	2,028	21.47	*Para una pareja **sin** hijos es cómodo*
2 negations	578	6.12	*Audi **no** se hace cargo de vehículos que **no** estén en garantía*
≥3 negations	219	2.32	*De lo ofrecido por la oferta de nuestra reserva **nada de nada**: **ni** "Special Amenity", **ni** "Free access to Sauna" puesto que estaba cerrada*
All	9,446	100.00	

Table 3 shows the total and percentage of negation cues grouped by type. We can see that most of the negation cues of the corpus are *simple* (3,147). However, we also find some *contiguous* cues (186) and a considerable amount of *non-contiguous* cues (608).

Table 4 provides the most frequent cues in the corpus, with the token *no* ('not') being the most common negation cue with a total of 2,317 occurrences.

In relation to the scopes annotated in the corpus, they correspond to a syntactic component, that is, a phrase, a clause or a sentence. They always include the corresponding

Table 3. Total and percentage of negation cues by type in the SFU Review$_{SP}$-NEG corpus

	# neg. cues	% neg. cues	Example
Simple	3,147	79.85	*El problema es que **no** saben arreglarlo*
Contiguous	186	4.72	***Ni nunca** quiso ser de nadie*
Non-contiguous	608	15.43	***No** tengo **nada** en contra de Opel*
All	3,941	100.00	

Table 4. Most frequent negation cues in the SFU Review$_{SP}$-NEG corpus

Cue	#	%	Example
no	2,317	58.79	***No** nos dieron ninguna facilidad*
sin	282	7.16	*El hotel está muy bien, **sin** grandes lujos*
ni	151	3.83	*Las habitaciones..., **ni** tienen terraza*
nada	125	3.17	*Pedimos lo normal, **nada** raro*
no-nada	120	3.04	***No** entiendo **nada***
nunca	76	1.93	*Ideal para recorrer en cualquier época, **nunca** pasarás calor*
nadie	57	1.45	***Nadie** salió de su puesto para ayudarme*
tampoco	50	1.27	*Las habitaciones **tampoco** tienen doble puerta*
no-ni	38	0.96	***No** hay en la habitación **ni** una hoja para decidir qué comer*
Other	725	18.40	*Sin poder entrar ya en **ningún** programa*

Table 5. Total and percentage of scopes grouped by type in the SFU Review$_{SP}$-NEG corpus

Scope span	#	%	Example
Before cue	230	5.84	*Tiene fiabilidad **cero***
After cue	2,702	68.56	*La pila de ropa [**sin** lavar] sigue subiendo*
Before and after cue	1,009	25.60	*[El ordenador **no** funciona bien]*

negation cue and the subject when the word directly affected by the negation is the verb of the sentence. We can find three types of scopes: i) scopes that span before the cue, ii) scopes that span after the cue, and iii) scopes that span before and after the cue. In Table 5 we present the total and percentage of scopes distributed by type. Most of the scopes span after the cue (2,720), although there is also an important amount of scopes that span before and after the cue (1,009) and a small amount before the cue (230).

The SFU Review$_{SP}$-NEG corpus constitutes an invaluable resource for the study of negation in Spanish. Given the opinionated nature of the texts (reviews), this corpus is also very useful to test the influence of negation for Spanish sentiment analysis. In the next section, we describe the method for negation processing, which we apply to the sentiment analysis task in Section 5.

4 Method for negation processing

To process negation, we use the system presented in Jiménez-Zafra et al. (2020). This system works on the SFU Review$_{SP}$-NEG corpus and models the task of detecting cues and scopes as two consecutive classification tasks, using a supervised machine learning method, the CRF classifier.

In the first phase of negation detection, a BIO representation is used to decide whether each word in a sentence is the beginning of a cue (B), the inside (I), or no cue (O). The BIO representation is useful in detecting multiword cues (MWCs), because those constitute a significant part of the SFU Review corpus (20.15% of the total number of negation cues in Spanish). In the second phase another classifier determines which words in the sentence are affected by the cues identified in the first phase. Similarly to the first phase, this involves, for every sentence that has a cue, identifying which of the other words in the sentence are inside (I) or outside (O) the scope of the cue.

The authors conduct experiments on the SFU Review$_{SP}$-NEG (Jiménez-Zafra et al., 2018), using the partitions (training, development and test) of the corpus in CoNLL format provided in *NEGES 2018: Workshop on Negation in Spanish* (Jiménez-Zafra et al., 2019a). They carry out the evaluation with the script of the **SEM-2012 Shared Task* (Morante and Blanco, 2012). Each token of the corpus is represented with a set of features because most researchers have defined negation cue detection and scope identification as token-level classification tasks. Specifically, they define a different set of features for each task by a selection process using the development set, and they use these features on the test set to report results.

The feature set for the negation cue detection task is composed of 31 features (Table 6). The feature selection process starts with using the feature set proposed by Cruz et al. (2016) for English texts: lemma and PoS tag of the token in focus, boolean tags to indicate if the token in focus is the first/last in the sentence, and the same features for the token before and after the token in focus (12 features in total). However, during the feature selection process in the development phase, Jiménez-Zafra et al. (2020) find that the most important features are lemmas and PoS tags. The authors conduct different experiments, finding that the lemma and PoS tags of seven tokens before and after the token in focus are useful. They also detect that discontinuous cues are difficult to classify. Therefore, they define as feature set the following: lemma and PoS tag of the token in focus as well as those of the seven tokens before and after it (features 1-30), and a string value stating whether the token in focus is part of any cue in the training set and whether it appears as the first token of a cue in the training set (B), as any token of a cue except the first (I), as both the first token of a cue and other positions (B_I), or if it does not belong to any cue of the training set (O) (feature 31). The motivation of this last feature is that many cues appear as a single token (e.g., *ni*, 'neither') and are also part of multiword cues (e.g., *ni siquiera*, 'not even').

For detecting scope, 24 features are used (Table 7). They are the same features used by Cruz et al. (2016) for detecting scopes in English. Although Jiménez-Zafra et al. (2020) conduct experiments on the development set filtering out those features that are not significant according to the chi-square test, results do not improve. Therefore, they define the same feature set as the one of Cruz et al. (2016): lemma and PoS tag of the current token and the cue in focus (features 1-4), location of the token with respect to the cue (feature 5) (before, inside or after), distance in number of tokens between the cue and the current token (feature 6), chain of PoS tags and chain of types between the cue and the token (features 7-8), lemma and PoS tags of the token to the left and right of the token in focus (features 9-12), relative position of the cue and the token in the sentence (features 13-14), dependency relation and direction (head or dependent) between the token and the cue (features 15-16), PoS tags of the first and second order syntactic heads of the token (features 17-18), whether

the token is ancestor of the token and vice versa (features 19-20), dependency shortest path from the token in focus to the cue and vice versa (features 21-22), dependency shortest path from the token in focus to the cue including direction (up or down) (feature 23), and length of the short path between the token and the cue (feature 24).

Table 6. Feature set for negation cue detection

	Feature name	Description
1,2	token	Lemma and PoS tag of t (token to be predicted)
3-30	token_window	Lemmas and PoS tags of 7 tokens before and after t
31	known_cue	Whether t was seen as a cue during training (B, I , B_I, or O)

Table 7. Feature set for scope identification

	Feature name	Description
1,2	token	Lemma and PoS tag of t (token to be predicted)
3,4	cue	Lemma and PoS tag of nc (negation cue)
5	location	Location of current t with respect to nc (before, inside or after)
6	distance	Number of tokens between t and nc
7	chain_pos_f	Sequence of fine PoS tags between t and nc
8	chain_pos_c	Same as *chain_pos_f* but with coarse tags
9–12	{l,r}_tokens	Lemma and PoS tags of the tokens to the left and right of t
13,14	rel_positions	Position of nc and t in the sentence over number of tokens in the sentence
15,16	dep_rel	Dependency type and direction (head or dependent) between t and nc
17,18	heads	PoS tags of the first and second order syntactic heads of t
19,20	is_ancestor	Whether t is an ancestor of nc and vice versa
21,22	path_types	Dependency types in the syntactic path from t to nc and vice versa
23	path_types_dir	Same as path_types but including direction (up or down) and only for t
24	path_length	Length of path_types

It should be noted that the system of Jiménez-Zafra et al. (2020) outperforms state-of-the-art results for negation cue detection in Spanish (Fabregat et al., 2018; Loharja et al., 2018; Giudice, 2019; Beltrán and González, 2019; Domínguez-Mas et al., 2019; Fabregat

et al., 2019)[6] and provides the first results for scope identification in Spanish. Therefore, we select this system for our experiments in this paper.

5 An application task: Sentiment analysis

Sentiment analysis is a mature field at the intersection of computer science and linguistics devoted to automatically determining evaluative content of a text (e.g., a review, a news article, a headline, a tweet). Such content may be whether the text is positive or negative, usually called polarity detection; whether it contains different types of evaluation or appraisal; or whether it contains emotion expressions and the categories of those emotions (see Taboada, 2016, for a survey). We focus here on the problem of polarity detection.

Approaches to this problem can be broadly classified into two types: lexicon-based or machine learning (Taboada, 2016; Taboada et al., 2011). In lexicon-based methods, dictionaries of positive and negative words are compiled, perhaps adding not just polarity, but also strength (e.g., *accolade* is strongly positive, whereas *accept* is mildly positive). When a new text is being processed, the system extracts all the words in the text that are present in the dictionary and aggregates them using different rules; for instance, a simple average of the values of all the words may be taken. The system may also take into account intensification and negation, changing the value of, respectively, *good, very good* and *not good.* Lexicons may, of course, also be compiled using machine learning methods (Taboada et al., 2011).

Most machine learning methods are a form of supervised learning, where enough samples of positive and negative texts are collected, and the classifier learns to distinguish them based on their features. Common features include n-grams (individual words and phrases), parts of speech or punctuation (Kennedy and Inkpen, 2006). In some of these cases, the features are lexicons of words, but these methods are different, in that the processing of texts in lexicon-based approaches typically involves rules (even rules as simple as averaging the values of the words in the text). In machine learning methods, negation may be picked up by unigrams as a single feature (a negative text may contain more instances of *not* and thus have a higher frequency of that unigram), or by bigrams and trigrams (*not good; not very good*), but otherwise the method is not able to detect whether an individual phrase is being negated. For this reason, we will focus on discussing negation in lexicon-based methods.

Assuming that negation and its scope have been adequately identified, lexicon-based methods may employ different strategies to account for its presence. A simple strategy is to reverse the polarity of the word or words in the scope of the negation, an approach that has been labelled as *switch negation* (Saurí, 2008). When the polarity is binary, this is simple. When the individual words in the dictionary have a more fine-grained scale, this becomes more complex. We know that negation is not symmetrical (Horn, 1989; Potts, 2011a), so simply changing the sign on any given word will not fully capture the contribution of negation. For instance, intuitively, *not good* and *not excellent* are not necessarily the exact opposite of *good* and *excellent*. This is more pronounced for strongly positive words

---

[6] It outperforms by evaluating on the same corpus. Specifically, on the SFU Review_SP-NEG corpus (Jiménez-Zafra et al., 2018) using the partitions provided in *NEGES 2018* (Jiménez-Zafra et al., 2019a).

like *excellent*. To address this imbalance, *shift negation* may be implemented, where the negated word is simply shifted along the scale by a fixed term. Thus, a very positive word like *excellent* may be negated to a mildly positive term.

In our experiments for this paper, we have made use of an existing lexicon-based method for sentiment analysis, SO-CAL (Taboada et al., 2011) and tested different ways to handle negation within the system. The next subsection provides an overview of SO-CAL and in particular of the implementation of negation in the system in Spanish. Then, the next subsection will describe our experiments and results with a new approach to negation detection.

5.1 SO-CAL

SO-CAL, The Semantic Orientation CALculator,[7] is a lexicon-based sentiment analysis system that was specifically designed for customer reviews, but has also been shown to work well on other texts such as blog posts or headlines (Taboada et al., 2011). It contains dictionaries[8] classified by part of speech (nouns, verbs, adjectives and adverbs), for a total of about 5,000 words for English and just over 4,200 for Spanish. SO-CAL takes into account intensification by words such as *very* or *slightly,* with each intensifier having a percentage associated with it, which increases or decreases the word it accompanies.

Negation in the standard SO-CAL system for both English and Spanish takes the shift method, i.e., any item in the scope of negation sees its polarity shifted by a fixed amount, 4 points in the best-performing version of the system. Thus, *excellent* (a +5 word in the dictionary) becomes *not excellent,* +1, and *sleazy,* which is a −3 word also becomes +1 when negated.

Negation in SO-CAL is handled by first identifying a sentiment word from the dictionaries. If a word is found, then the system tracks back to the previous and searches for a negation keyword. If a negation keyword is present before the sentiment word, then negation is applied to the sentiment word. Scope is not explicitly identified, i.e., the system assumes that a sentiment-bearing word is in the scope of negation if it is after the negation keyword in the same sentence. The system may continue to track back and keep looking left for negation keywords if a 'skipped' word is present, such as adjectives, copulas, determiners and certain verbs. Skipped words allow the system to look for keywords in cases of raised negation, e.g., *I don't think it is good*, where the system would keep skipping backwards through the words *is, it* and *think* to find the raised negation that affects the sentiment of *good.*

Sentiment for a text is calculated by extracting all sentiment words, calculating intensification and negation for relevant phrases, and then averaging the values of all the words and phrases in the text. The accuracy of the original system is 80% for English (Taboada et al., 2011) and about 72% for Spanish (Brooke et al., 2009). Our goal in this paper is to investigate whether a more accurate method for negation detection can improve the Spanish results.

---

[7] https://github.com/sfu-discourse-lab/SO-CAL
[8] https://github.com/sfu-discourse-lab/Sentiment_Analysis_Dictionaries

5.2 Experiments

We conduct experiments on the corpus with negation annotations, the SFU Review_{SP}-NEG (Jiménez-Zafra et al., 2018). The experimentation was organized in the following phases:

- Phase A: Negation cue detection
    1. Prediction of the negation cues on the texts of the SFU Review_{SP}-NEG corpus using the system provided by Jiménez-Zafra et al. (2020) and 10-fold cross validation in order to classify all the reviews.
- Phase B: Scope identification
    1. Identification of the scopes corresponding to the predicted cues in Phase A - 1.
- Phase C: Sentiment analysis
    1. Classification of the texts of the SFU Review_{SP}-NEG corpus using the SO-CAL system without negation.
    2. Classification of the texts of the SFU Review_{SP}-NEG corpus using the SO-CAL system with built-in negation, i.e., using the rule-based method that incorporates the detection of cues and scopes in Spanish that is built in the SO-CAL system.
    3. Classification of the texts of the SFU Review_{SP}-NEG corpus using the SO-CAL system with the output of the negation processing system applied in Phase A and Phase B.

5.3 Evaluation measures

The output of the systems for negation cue detection and scope identification (Phases A and B) is evaluated with the script (https://www.clips.uantwerpen.be/sem2012-st-neg/data.html) used in the **SEM 2012 Shared Task "Resolving the Scope and Focus of Negation"* (Morante and Blanco, 2012): Precision (P), Recall (R) and F-score (F1). It is based on the following criteria:

- Punctuation tokens are ignored.
- A True Positive (TP) requires all tokens of the negation element (cue or scope) to be correctly identified.
- A False Negative (FN) is counted either by the system not identifying negation elements present in the gold annotations, or by identifying them partially, i.e., not all tokens have been correctly identified or the word forms are incorrect.
- A False Positive (FP) is counted when the system produces a negation element not present in the gold annotations.

For the evaluation of the sentiment analysis experiments (Phase C), the traditional measures used in text classification are applied: P, R and F1. They are measured per class (positive and negative) and averaged using the macro-average method.

5.4 Results

We evaluate the effect of the Spanish negation detection system on sentiment classification, using the SFU Review_{SP}-NEG corpus. Table 8 details the results for negation cue detection

and scope resolution. These results are obtained by employing 10-fold cross-validation with the same number of documents in all the folds.

In general, the results for negation cue detection and scope identification in Table 8 are encouraging. The cue detection module is very precise (92.70%) and provides a good recall (82.09%), although not as good as the English system, with 89.64% precision and 95.63% recall (Cruz et al., 2016). This is probably because negation expression in Spanish shows more variation than in English. We can find multiple negations in a sentence and they can be composed of two or more contiguous or non-contiguous tokens, increasing the difficulty of the task. On the other hand, the scope identification module is also very precise (90.77%), but its recall is not very high (63.64%) due to the fact that we can find three types of scopes: scopes than span before the cue, after the cue or before and after the cue, making scope resolution challenging. Moreover, we also need to consider the errors that the classifier introduces in the cue detection phase and which are accumulated in the scope recognition phase.

We observe that negation detection shows high accuracy, especially for cue detection. This is encouraging, as it will allow us to apply negation to several tasks.

Table 8. Results for negation cue detection (Phase A) and scope identification (Phase B) on the SFU Review$_{SP}$-NEG corpus using 10-fold cross validation - (P = Precision, R = Recall, F1 = F-score)

	Cue			Scope		
	P	R	F1	P	R	F1
Books	87.67	81.24	84.33	84.69	63.23	72.40
Cars	93.01	82.10	87.22	90.65	59.88	72.12
Cell phones	95.51	84.83	89.85	94.12	63.87	76.10
Computers	94.43	83.93	88.43	91.59	64.26	75.53
Hotels	92.97	80.83	86.48	91.47	65.56	76.38
Movies	91.84	81.73	86.49	89.96	65.06	75.51
Music	91.98	79.89	85.51	89.34	58.45	70.67
Washing machines	95.19	82.20	88.22	94.31	68.84	79.59
All	**92.70**	**82.09**	**87.07**	**90.77**	**63.64**	**74.79**

We then proceeded to use the output of these two phases to test the contribution of accurate negation identification to sentiment analysis. Results are shown in Tables 9 and 10. These tables show the results of our negation detection algorithm (SO-CAL with negation processing system), compared to using the search heuristics implemented in the existing SO-CAL system (SO-CAL with built-in negation) and to a simple baseline which involves not applying any negation identification (SO-CAL without negation). We discuss the most relevant aspects of these results in the next section, on error analysis.

Table 9. Results per class (positive and negative) on the SFU Review$_{\text{SP}}$-NEG corpus for sentiment analysis using SO-CAL without negation (Phase C - 1), SO-CAL with built-in negation (Phase C - 2) and SO-CAL with negation processing system (Phase C - 3) - (P = Precision, R = Recall, F1 = F-score)

| | SO-CAL without negation | | | | | | SO-CAL with built-in negation | | | | | | SO-CAL with negation processing system | | | | | |
| | Positive class | | | Negative class | | | Positive class | | | Negative class | | | Positive class | | | Negative class | | |
	P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1
Books	57.10	64.00	60.40	59.10	52.00	55.30	60.70	68.00	64.20	63.60	56.00	59.60	66.70	64.00	65.30	65.40	68.00	66.70
Cars	74.10	80.00	75.50	77.30	68.00	72.30	71.40	80.00	75.50	77.30	68.00	72.30	85.00	68.00	75.60	73.30	88.00	80.00
Cell phones	66.70	88.00	75.90	82.40	56.00	66.70	68.80	88.00	78.60	84.20	64.00	72.70	61.80	84.00	71.20	75.00	48.00	58.50
Computers	73.30	88.00	80.00	85.00	68.00	75.60	69.70	92.00	79.30	88.20	60.00	71.40	78.60	88.00	83.00	86.40	76.00	80.90
Hotels	77.40	96.00	85.70	94.70	72.00	81.80	74.40	96.00	85.70	94.70	72.00	81.80	80.00	96.00	87.30	95.00	76.00	84.40
Movies	65.20	60.00	62.50	63.00	68.00	65.40	66.70	56.00	60.90	62.10	72.00	66.70	76.50	52.00	61.90	63.60	84.00	72.40
Music	70.00	84.00	76.40	80.00	64.00	71.10	71.00	88.00	78.60	84.20	64.00	72.70	79.30	92.00	85.20	90.50	76.00	82.60
Washing machines	69.00	80.00	74.10	76.20	64.00	69.60	69.00	80.00	74.10	76.20	64.00	69.60	70.40	76.00	73.10	73.90	68.00	70.80
All	69.00	80.00	74.00	77.00	64.00	70.00	69.00	**81.00**	74.40	**78.70**	64.50	70.50	**74.80**	77.50	75.30	77.90	**73.00**	74.50

Table 10. Total results on the SFU Review_SP-NEG corpus for sentiment analysis using SO-CAL without negation (Phase C - 1), SO-CAL with built-in negation (Phase C - 2) and SO-CAL with negation processing system (Phase C - 3) - (P = Precision, R = Recall, F1 = F-score)

	SO-CAL without negation				SO-CAL with built-in negation				SO-CAL with negation processing system			
	Avg. P	Avg. R	Avg. F1	Acc.	Avg. P	Avg. R	Avg. F1	Acc.	Avg. P	Avg. R	Avg. F1	Acc.
Books	58.10	58.00	57.80	58.00	62.20	62.00	61.90	62.00	66.00	66.00	66.00	66.00
Cars	74.40	74.00	73.90	74.00	74.40	74.00	73.90	74.00	79.20	78.00	77.80	78.00
Cell phones	74.50	72.00	71.30	72.00	76.00	74.00	73.50	74.00	68.40	66.00	64.90	66.00
Computers	79.20	78.00	77.80	78.00	79.00	76.00	75.40	76.00	82.50	82.00	81.90	82.00
Hotels	86.10	84.00	83.80	84.00	84.00	84.00	83.80	84.00	87.50	86.00	85.90	86.00
Movies	64.10	64.00	63.90	64.00	64.40	64.00	63.80	64.00	70.10	68.00	67.20	68.00
Music	75.00	74.00	73.70	74.00	77.60	76.00	75.60	76.00	84.90	84.00	83.90	84.00
Washing machines	72.60	72.00	71.80	72.00	72.60	72.00	71.80	72.00	72.10	72.00	72.00	72.00
All	73.00	72.00	72.00	72.00	73.80	72.80	72.50	72.80	76.30	75.30	75.00	75.30

5.5 Error analysis

In this section we conduct an analysis of the SO-CAL system using our algorithm for negation detection, compared to SO-CAL's built-in detection system, which simply traces back until it finds a negation cue, without explicitly detecting scope (see Subsection 5.1). As expected and as shown in previous work, performance of the systems that integrate negation (SO-CAL with built-in negation and SO-CAL with negation processing system) outperform the baseline (SO-CAL without negation) in terms of overall precision, recall, F1 and accuracy. We are interested in studying how this improvement takes place and, in particular, in the cases where it hinders rather than helps the sentiment prediction.

In general, SO-CAL without negation is biased towards positive polarity in Spanish, with the F1-score for positive reviews higher than for negatives ones. We found the same result in English; see Cruz et al. (2016). This means that ignoring negation has an impact on the recognition of negative opinion in reviews. It is also the case, however, that the negative class has a lower overall performance, mostly due to low recall. It is well established that detecting negative opinions is more difficult than detecting positive ones (Ribeiro et al., 2016), for a host of reasons, including a possible universal positivity bias (Boucher and Osgood, 1969).

The configuration of SO-CAL with the negation processing system achieves the best performance, improving on the baseline by 3.3% and the search heuristic by 2.5% in terms of overall accuracy. These results can be explained by two factors. First, the negation detector that we propose benefits from a wider list of cues (the built-in search heuristics in SO-CAL include 13 different negation cues while the SFU Review$_{SP}$-NEG corpus contains 245 different negation cues). Second, the scope detection approach goes beyond the window-based heuristic that the SO-CAL system incorporates. Below, we illustrate with examples[9] these two situations. Note that in the examples there are two identifiers, 'NEGATED' and 'NEGATIVE'. The former refers to a word in the scope of negation. The latter is used for any word or phrase that is negative either from the dictionary (i.e., it has a negative value in the dictionary) or that becomes negative as a result of negation. When a negative expression is encountered, SO-CAL multiplies its value by 1.5. This accounts for the asymmetry of negation: a negative expression tends to be more saliently negative than a positive expression is positive (Taboada et al., 2011, 2017; Rozin and Royzman, 2001).

- Case 1: Negation cue predicted by the negation processing system, but not present in the SO-CAL list. For example, in Example (7), the negation cue *ningún* is identified by the negation detector but it is not present in the built-in SO-CAL list. Therefore, *ningun temazo* is correctly classified as negative (-1.5 points) by SO-CAL when we integrate the Spanish negation detector, but with the heuristic that it incorporates by default it is incorrectly classified as positive (3.0 points).

(7) Aqui tenemos un disco bastante antiguo de los smith... a mi gusto no cuenta con **ningun** temazo...
   *Here we have a pretty old Smith album... to my liking it doesn't have **any** hits...*

---

[9] Some of the examples contain grammatical errors in the original. Sentences are shown as written by users, to show that an added difficulty of the task is working with misspelled words.

    a. SO-CAL with built-in negation:
       temazo 3.0 = 3.0
    b. SO-CAL with negation processing system:
       ningun temazo 3.0 - 4.0 (NEGATED) X 1.5 (NEGATIVE) = -1.5

- Case 2: Scope correctly identified by the Spanish negation processing system, but not detected by SO-CAL due to its heuristic that checks if a word is negated based on looking for a negation cue in the previous word, unless the previous word is in the list of skipped words (see Section 5.1). The sentence in Example (8) is correctly classified as negative when we use the Spanish negation detector in SO-CAL, because the sentiment word *buena* is identified as negated by the negation cue *no*. However, using the search heuristic that SO-CAL incorporates by default, the sentence is incorrectly classified as positive. The search heuristic works as follows. The system detects that *buena* is a sentiment word and checks if the previous word is a negation cue of the list; *una* is not in the list, so the system checks if it is a skipped word in order to continue checking the previous words, but *una* is not in the skipped list either and therefore the sentence is incorrectly classified as positive.

(8) Han ahorrado en seguridad, lo que **no** es una buena politica.
    *They have saved on security, which is **not** a good policy.*

    a. SO-CAL with built-in negation:
       seguridad 2.0 = 2.0; buena 2.0 = 2.0
    b. SO-CAL with negation processing system:
       seguridad 2.0 = 2.0; no es una buena 2.0 - 4.0 (NEGATED) X 1.5 (NEGATIVE) = -3.0

The improvement we obtained on the sentiment classification of the reviews using the Spanish negation detector system is not as high as that achieved with the English system. In English, using the system by Cruz et al. (2016), we saw an improvement of 5% over the baseline and about 2% over the built-in search heuristics both in terms of overall F1 and accuracy. To determine where errors occurred in the Spanish analysis, and to tease apart those that were the result of a potentially under-developed Spanish SO-CAL as opposed to faulty negation detection, we have identified different error types. We describe each error type below, with examples.

Type 1 error: Words correctly identified as scope by the Spanish negation detector that are present in the SO-CAL dictionary, but are not sentiment words in the domain under study. In Example (9), the word *official* is in the scope of negation and belongs to the positive dictionary. However, in this context, *official* is not a positive word. The application of the sentiment heuristic of SO-CAL converts this word into a very negative one and consequently, the negative polarity of the sentence is increased in an incorrect way.

(9) De todas las mecánicas que puede montar, a mi la que más me gusta es el modelo de gasoil, de 1.9 cc pues creo que lo que pagas y las prestaciones que te da están muy bien, además su consumo es bastante equilibrado, si no subimos mucho el régimen de giro

(por encima de las 3500 vueltas), podemos gastar unos 6 litros y poco más de gasoil, estos datos **no** son los oficiales, son los reales obetnidos con este modelo, aunque por supuesto, dependiendo de muchos factores, este consumo varriará.

*Of all the mechanics one can configure, the one I like the most is the Diesel model, 1.9 cc because I think that what you pay and the performance that it gives you is very good, also its consumption is quite balanced, if we do not raise the rotation (above 3500 laps), we can consume just a bit more than 6 liters of Diesel, these data are **not** official, they are the real results obtained with this model, although of course, depending on many factors, this consumption will vary.*

a. SO-CAL with built-in negation:
   oficiales 1.0 = 1.0
b. SO-CAL with negation processing system:
   no son los oficiales 1.0 - 4.0 (NEGATED) X 1.5 (NEGATIVE) = -4.5

Type 2 error: Positive words in the SO-CAL dictionary whose sentiment value is low and the negation weighting factor is very high (-4). The sentiment heuristic of SO-CAL works as follows: if a positive word is negated, 4 points are subtracted from the scoring of the positive word and if the result is a negative value, it is multiplied by 1.5 points (this helps capture the asymmetric nature of negation; see above). On the other hand, if the word is negative, it is annulled, i.e., to the scoring of the word we add its opposite value. In Example (10), the positive word *mejor* has a value of 1 point in the SO-CAL dictionary. This is a low sentiment value and the negation weighting factor is very high (-4), consequently the polarity of the sub-string *sin ser el mejor* has a high negative value (-4.5), causing the sentence to be incorrectly classified as negative (0.67 - 4.5 + 1.25 + 1 = -1.58).

(10) Es una buena opción que **sin** ser el mejor ordenador del mercado, en relación calidad-precio es muy aceptable y durante un par de años (mínimo) estarás muy agusto con él, luego, quizás tengas que ampliar memoria, etc.
   *It is a good option that **without** being the best computer on the market, has a very acceptable quality-price relationship and for a couple of years (minimum) you will be very comfortable with it, then, you may have to expand memory, etc.*

a. SO-CAL with built-in negation:
   ampliar 1.0 = 1.0; buena 2.0 X 1/3 (REPEATED) = 0.67; mejor 1.0 X 1/2 (RE-PEATED) = 0.5; muy aceptable 1.0 X 1.25 (INTENSIFIED) = 1.25
b. SO-CAL with negation processing system:
   buena 2.0 X 1/3 (REPEATED) = 0.67; sin ser el mejor 1.0 - 4.0 (NEGATED) X 1.5 (NEGATIVE) = -4.5; muy aceptable 1.0 X 1.25 (INTENSIFIED) = 1.25; ampliar 1.0 = 1.0

Type 3 error: Sentiment words that are not included in the SO-CAL dictionary. In Example (11) the positive word *encanta* is not detected by SO-CAL because it is not in the positive dictionary. Therefore, the sentence is incorrectly classified with 0 points instead of being labelled as a positive sentence.

(11) A todos mis amigos les **encanta** mi movil y ahora están pensando en comprárselo ellos también, bueno os dejo amigos de Ciao!!
*All my friends **love** my mobile and now they are thinking of buying it too, well I leave you friends of Ciao!*

    a. SO-CAL with built-in negation: 0
    b. SO-CAL with negation processing system: 0

Type 4 error: Negation used in an ironic way. In Example (12), the sub-string *no nos íbamos a asfixiar porque tenía sus boquetitos* contains the negation cue *no*, that is correctly identified along with its scope by the Spanish negation detector. However, in this case, negation is used in an ironic way and it should not have been taken into account as negation. Therefore, instead of being classified with 0 points, it should have been assigned a negative score.

(12) INCREÍBLE , el cuarto era de moqueta y no brillaba la limpieza, la iluminación era del conde drácula y a mi me daba un agobio no poder abria la ventana increible, pero claro **no** nos íbamos a asfixiar porque tenía sus boquetitos por el que entraba el aire perfumado por lo que adornaba la ventana.
*INCREDIBLE, the room was carpeted and it was not clean, the illumination was Count Dracula-type and I felt claustrophobic because I could not open the incredible window, but of course we were **not** going to asphyxiate because it had holes adorning the window through which the perfumed air entered.*

    a. SO-CAL with built-in negation:
    agobio -4.0 X 1.5 (NEGATIVE) = -6.0; asfixiar -5.0 X 2.0 (HIGHLIGHTED) X 1.5 (NEGATIVE) = -15.0; increible -4.0 X 1.5 (NEGATIVE) = -6.0; claro 1.0 X 2.0 (HIGHLIGHTED) = 2.0
    b. SO-CAL with negation processing system:
    agobio -4.0 X 1.5 (NEGATIVE) = -6.0; claro no nos íbamos a asfixiar -5.0 + 5.0 (NEGATED) X 1.3 (INTENSIFIED) X 2.0 (HIGHLIGHTED) = 0; no poder abria la ventana increible -4.0 + 4.0 (NEGATED) = 0

Types 1-4 are errors of the sentiment analysis system. Now we focus on errors in the negation detection system.

Type 5 error: Negation cue detected by the SO-CAL system, but not predicted by the negation detector. In Example (13), the negation processing system has not predicted the word *falta* as negation cue. Therefore, the word *mejorar* has been classified as positive (6 points), but it should have been classified as negative due to the presence of negation.

(13) Es un gran telefono por la forma, pero **falta** mejorar lo muchisimo para mi gusto.
*It's a great phone based on the shape, but **it needs** a lot of improvement in my opinion.*

    a. SO-CAL with built-in negation:
    falta mejorar 3.0 - 4.0 (NEGATED) X 2.0 (HIGHLIGHTED) X 1.5 (NEGATIVE) = -3.0; gran 3.0 = 3.0

    b. SO-CAL with negation processing system:
       mejorar 3.0 X 2.0 (HIGHLIGHTED) = 6.0; gran 3.0 = 3.0

Type 6 error: Scope erroneously predicted by the negation detector. In Example (14), the negation processing system has predicted the following as scope of the last negation cue, *no*: *no dejaria de escribir sobre esta horrible experiencia*. However, this scope is not correct and, consequently, the sentiment word *horrible* has been negated, but it should not have been negated and it should have preserved its negative polarity.

(14) No hace falta hablar de la calidad de dicho aparato, PESIMA increiblemente malo, resulta que al abrir la tapa se ha roto la pantalla interior y no se ve nada, fui a el servicio tecnico ya que es sorprendente que solo me durara un mes y me dijeron que se habia roto por la presion ocasionada a el abrirlo, ALUCINANTE ya que no he ejercido ninguna presion en el movil ni he dado ningun golpe, pero bueno vamos a las prestaciones que tiene que de la rabia que tengo **no** dejaria de escribir sobre esta horrible experiencia.

*There is no need to talk about the quality of this device, it is TERRIBLE, incredibly bad, when I opened the lid the inner screen broke and I cannot see anything, I went to the technical service because it is amazing that it only lasted a month and I was told that it was broken by the pressure caused to open it, AMAZING, because I have not exerted any pressure on the mobile nor have I hit it, but okay, let's go ahead and talk about the good sides that it has, because the rage I feel would **not** stop me writing about this horrible experience.*

    a. SO-CAL with built-in negation:
       horrible -4.0 X 2.0 (HIGHLIGHTED) X 1.5 (NEGATIVE) = -12.0; increiblemente malo -3.0 X 1.35 (INTENSIFIED) X 1.5 (NEGATIVE) = -6.075; solo -1.0 X 1.5 (NEGATIVE) = -1.5; FACILIDAD 3.0 X 2.0 (CAPITALIZED) = 6.0; presion -3.0 X 1.5 (NEGATIVE) = -4.5; ninguna presion -3.0 + 3.0 (NEGATED) = 0; sorprendente 3.0 = 3.0; movil 1.0 = 1.0

    b. SO-CAL with negation processing system:
       increiblemente malo -3.0 X 1.35 (INTENSIFIED) X 1.5 (NEGATIVE) = -6.075; no dejaria de escribir sobre esta horrible -4.0 + 4.0 (NEGATED) X 2.0 (HIGH-LIGHTED) = 0; solo -1.0 X 1.5 (NEGATIVE) = -1.5; FACILIDAD 3.0 X 2.0 (CAP-ITALIZED) = 6.0; presion -3.0 X 1.5 (NEGATIVE) = -4.5; no he ejercido ninguna presion -3.0 + 3.0 (NEGATED) = 0; sorprendente 3.0 = 3.0; movil 1.0 = 1.0

In Table 11 we provide the results of an error analysis of 224 sentences from 12 different reviews, 88 of which (39.3%) contained at least one instance of negation. In those sentences, we found 23 errors with the negation or with how the negation was incorporated into SO-CAL. Type 1 and Type 2 errors are the most common, having to do with either lack of coverage in the SO-CAL dictionaries or with how negation was computed if it was found. The next highest category is Type 6, with the negation system mistakenly identifying a scope. There are, however, only 6 of those cases and they do not all affect the sentiment. It is clear that the negation processing system does its job fairly well, but it is hindered by the relatively less well developed Spanish SO-CAL (in comparison to the

Table 11. Error analysis of 224 sentences

Error type	# Errors	% Errors
Type 1	7	30.4%
Type 2	7	30.4%
Type 3	1	4.3%
Type 4	1	4.3%
Type 5	1	4.3%
Type 6	6	26.1%
Total errors	23	100%
Total number of sentences	224	
Total number of sentences with negation	88	

English version). Better performance, therefore, can be achieved by developing the system in conjunction with adopting a state-of-the-art negation processing system.

One interesting case that is rare, but indicative of the difficulty of sentiment analysis, is a case where the negation processing system detects negation accurately, and negates the right word, but the negation ends up hurting the score of the text overall because the negated word is not directly relevant to the product being discussed. In Example (15), SO-CAL's built-in negation method misses the negation. The negation processing system detects it, and changes the polarity of *activa*, literally 'active' to negative. In this case, however, the word does not refer to the phone being discussed in the review, or to the other phone the reviewer considered (PEBL U6), but to the phone company's available phones for their 'active pack'. We categorized this as a Type 1 error, having to do with the domain vocabulary, but it is clearly a more complex issue about word ambiguity in context (Benamara et al., 2018).

(15) yo a el principio keria el PEBL U6 negro pero era demasiado caro y **no** estaba en el pack activa de movistar.
*At first I wanted the black PEBL U6, but it was too expensive and it was **not** part of Movistar's active pack.*

Regarding the detection of negation, as identified in the work of Jiménez-Zafra et al. (2020), the Spanish language has some peculiarities. Negation cues can be simple (Example 16), continuous (Example 17) or discontinuous (Example 18). Moreover, some common negation cues, such as *no*, are also frequent in comparative (Example 19), contrasting (Example 20) and rhetoric structures (Example 21), making the task more difficult. In addition, the scope of negation[10] can span before the cue (Example 22), after the cue (Example 23), or before and after the cue (Example 24). In the texts analysed from the point of

---

[10] Scopes are marked between square brackets.

view of their application to sentiment analysis the errors found are due to: i) the system is trained to identify syntactic negation, and ii) it is difficult to determine whether the subject and complements of the verb are included in the scope or not. Depending on the negation cue used we can find different types of negation: syntactic negation (e.g. no [no/not], nunca [never]), lexical negation (e.g. negar [deny], desistir [desist]), and morphological or affixal negation, (e.g. ilegal [illegal], incoherente [incoherent]). The negation detector is trained to identify only syntactic negation so it is not able to correctly predict lexical nor morphological negation. As it is shown in Example (13), the negation processing system has not predicted the lexical negation cue *falta*. This could be solved by performing the annotation of the other types of negation, task that we plan to carry out in the future. On the other hand, the identification of scopes is a difficult task. Most of the errors found[11] are due to the fact that the scope does not always start with the cue and does not always end in a punctuation mark (Example 25), and the inclusion or not of the subject (Example 26) and the complements of the verbs within it (Example 27). In order to solve these errors we plan to experiment with adding more sophisticated syntactic features.

(16) El problema es que **no** saben arreglarlo.
      *The problem is they **don't** know how to fix it.*

(17) **Ni nunca** quiso ser de nadie.
      ***Nor** did he ever want to be anyone's.*

(18) **No** tengo **nada** en contra de Opel.
      *I have **nothing** against Opel.*

(19) **No** me gusta **tanto como** lo otro.
      *I **don't** like it **as much as** the other thing*

(20) **No** hay **más** solución **que** comprar una lavadora.
      *There is **no other** solution **than to** buy a washing machine*

(21) Viniste a verlo, no?
      *You came to see him, **didn't** you?*

(22) [El producto tiene fiabilidad **cero**].
      *[The product has **zero** reliability].*

(23) El problema es que [**no** saben arreglarlo].
      *The problem is [they **don't** know how to fix it].*

(24) Aunque [las habitaciones **no** están mal], la atención recibida me hace calificarlo mal.
      *[The rooms are **not** bad], but the attention received makes me rate it poorly.*

---

[11] Gold scopes are between square brackets and system scopes between curly brackets.

(25) Tiene 156.25 gibabytes de disco duro , que como os podreis imaginar [{**sin** ser exce-
sivo] estan muy bien a nivel usuario}, se tarda mucho, pero mucho tiempo en llenar
el disco duro...
*It has 156.25 gibabytes of hard disk, which as you can imagine [{**without** being excessive] are*
very good at the user level}, it takes a lot, but a long time to fill the hard drive...

(26) Los plásticos resultan demasiado evidentes y [la tapicería {**no** es nada del otro
mundo}].
*Plastics are too obvious and [upholstery is {**no** big deal}].*

(27) Vamos, [por 11900 euros {yo **no** me lo compraba}].
*[For 11900 euros {I **didn't** buy it}].*

In summary, we have shown that accurate negation detection is possible, and that the
system that we have adopted from previous work perform well. In addition, we show that
improvements in sentiment analysis can be gained from detecting negation and its scope
with sophisticated negation processing systems.

6 Conclusion

In this work we use a machine learning system that automatically identifies negation cues
and their scope in Spanish review texts and we investigate whether accurate negation detec-
tion helps to improve the results of a sentiment analysis system. Although it has long been
known that accurate negation detection is crucial for sentiment analysis, the novelty of this
work lies in the fact that, to the best of our knowledge, this is the first full implementation
of a Spanish machine learning negation detector in a sentiment analysis system. Another
contribution of the paper is the error analysis. We classify errors into different types, both
in the negation detection and the sentiment analysis phases.

The results obtained show that accurate recognition of cues and scopes is of paramount
importance to the sentiment classification task and reveal that simplistic approaches to
negation are insufficient for sentiment detection. In addition, the analysis of errors reveals
that Spanish presents additional challenges in negation processing such as double negation
or non-contiguous negation cues (Wang, 2006).

Future research for Spanish will focus on the improvement of the negation detection
system, especially in the correct identification of contiguous and non-contiguous cues, and
the exploration of some post-processing algorithm in order to cover the three types of
scopes that we can find (before the cue, after the cue, or before and after the cue). More-
over, we plan to check the SO-CAL Spanish dictionaries. There are some words that are
clearly sentiment words, such as *encanta* ('love'), that are not included in these dictionar-
ies. The Spanish dictionaries were translated from the English SO-CAL and then manually
inspected and corrected (Brooke et al., 2009), but more manual curation is probably neces-
sary. In addition, we should review the negation weighting factor of the sentiment heuristic
of SO-CAL. This was introduced because negative statements seemed to carry more weight
than negative ones. For a system that detects only a few negations, it may be appropriate,
but for a system that identifies a larger number, it may not be as useful, because it some-
times results in a very high negative score.

Besides the sentiment analysis task, accurate negation detection is useful in a number of domains: speculation detection, misinformation and 'fake' news or deception detection. The system presented here can be tested on such domains.

Acknowledgements

References

Mario Amores, Leticia Arco, and Abel Barrera. Efectos de la negación, modificadores, jergas, abreviaturas y emoticonos en el análisis de sentimiento. In *IWSW*, pages 43–53, 2016.

Kathryn Baker, Michael Bloodgood, Bonnie J Dorr, Chris Callison-Burch, Nathaniel W Filardo, Christine Piatko, Lori Levin, and Scott Miller. Modality and negation in SIMT use of modality and negation in semantically-informed syntactic MT. *Computational Linguistics*, 38(2):411–438, 2012.

Jeremy Barnes, Erik Velldal, and Lilja Øvrelid. Improving sentiment analysis with multi-task learning of negation. *arXiv preprint arXiv:1906.07610*, 2019.

Javier Beltrán and Mónica González. Detection of Negation Cues in Spanish: The CLiC-Neg System. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2019)*, CEUR Workshop Proceedings, Bilbao, Spain, 2019. CEUR-WS.

Farah Benamara, Baptiste Chardon, Yvette Yannick Mathieu, Vladimir Popescu, and Nicholas Asher. How do negation and modality impact opinions? In *Proceedings of the ACL-2012 Workshop on Extra-Propositional Aspects of Meaning in Computational Linguistics (ExProM-2012)*, pages 10–18, Jeju, Korea, 2012.

Farah Benamara, Diana Inkpen, and Maite Taboada. Introduction to the Special Issue on Language in Social Media: Exploiting Discourse and Other Contextual Information. *Computational Linguistics*, 44(4):663–681, 2018.

Eduardo Blanco and Dan Moldovan. Retrieving implicit positive meaning from negated statements. *Natural Language Engineering*, 20(4):501–535, 2014.

Ignacio Bosque, Victor García de la Concha, and Humberto López Morales. *Nueva gramática de la lengua española: morfología y sintaxis*. Real Academia Española, Madrid, 2009.

Jerry D. Boucher and Charles E. Osgood. The Pollyanna hypothesis. *Journal of Verbal Learning and Verbal Behaviour*, 8:1–8, 1969.

Julian Brooke, Milan Tofiloski, and Maite Taboada. Cross-linguistic sentiment analysis: From English to Spanish. In *Proceedings of the International Conference RANLP-2009*, pages 50–54, 2009.

Wendy W Chapman, Will Bridewell, Paul Hanbury, Gregory F Cooper, and Bruce G Buchanan. A simple algorithm for identifying negated findings and diseases in discharge summaries. *Journal of biomedical informatics*, 34(5):301–310, 2001.

Roberto Costumero, Federico López, Consuelo Gonzalo-Martín, Marta Millan, and Ernestina Menasalvas. An approach to detect negation on medical documents in Spanish. In *International Conference on Brain Informatics and Health*, pages 366–375. Springer, 2014.

Viviana Cotik, Vanesa Stricker, Jorge Vivaldi, and Horacio Rodríguez Hontoria. Syntactic methods for negation detection in radiology reports in Spanish. In *Proceedings of the 15th Workshop on Biomedical Natural Language Processing, BioNLP 2016: Berlin, Germany, August 12, 2016*, pages 156–165. Association for Computational Linguistics, 2016.

Isaac G. Councill, Ryan McDonald, and Leonid Velikovich. What's great and what's not: Learning to classify the scope of negation for improved sentiment analysis. In *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*, pages 51–59, Uppsala, Sweden, 2010.

Noa P Cruz, Maite Taboada, and Ruslan Mitkov. A machine-learning approach to negation and speculation detection for sentiment analysis. *Journal of the Association for Information Science and Technology*, 67(9):2118–2136, 2016.

Noa P Cruz Díaz and Manuel J Maña López. *Negation and Speculation Detection*, volume 13. John Benjamins Publishing Company, 2019.

Lluís Domínguez-Mas, Francesco Ronzano, and Laura I. Furlong. Supervised learning approaches to detect negation cues in Spanish reviews. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2019)*, CEUR Workshop Proceedings, Bilbao, Spain, 2019. CEUR-WS.

Hermenegildo Fabregat, Juan Martínez-Romo, and Lourdes Araujo. Deep Learning Approach for Negation Cues Detection in Spanish at NEGES 2018. In *Proceedings of NEGES 2018: Workshop on Negation in Spanish, CEUR Workshop Proceedings*, volume 2174, pages 43–48, 2018.

Hermenegildo Fabregat, Andrés Duque, Juan Martínez-Romo, and Lourdes Araujo. Extending a Deep Learning Approach for Negation Cues Detection in Spanish. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2019)*, CEUR Workshop Proceedings, Bilbao, Spain, 2019. CEUR-WS.

Federico Fancellu, Adam Lopez, and Bonnie Webber. Neural networks for negation scope detection. In *Proceedings of the 54th annual meeting of the Association for Computational Linguistics (volume 1: long papers)*, pages 495–504, 2016.

Valentino Giudice. Aspie96 at NEGES (IberLEF 2019): Negation Cues Detection in Spanish with Character-Level Convolutional RNN and Tokenization. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2019)*, CEUR Workshop Proceedings, Bilbao, Spain, 2019. CEUR-WS.

Laurence Horn. *A Natural History of Negation*. University of Chicago Press, Chicago, 1989.

Michael Israel. The pragmatics of polarity. In Laurence Horn and Gregory Ward, editors, *The Handbook of Pragmatics*, pages 701–723. Blackwell, Malden, MA, 2004.

Lifeng Jia, Clement Yu, and Weiyi Meng. The effect of negation on sentiment analysis and retrieval effectiveness. In *Proceedings of the 18th ACM conference on Information and knowledge management*, pages 1827–1830. ACM, 2009.

Salud María Jiménez-Zafra, Eugenio Martínez-Cámara, M. Teresa Martín-Valdivia, and

M. Dolores Molina-González. Tratamiento de la Negación en el Análisis de Opiniones en Español. *Procesamiento del Lenguaje Natural*, 54:367–44, 2015.

Salud María Jiménez-Zafra, María Teresa Martín-Valdivia, Eugenio Martínez-Cámara, and L. Alfonso Ureña-López. Studying the Scope of Negation for Spanish Sentiment Analysis on Twitter. *IEEE Transactions on Affective Computing*, 10(1):129–141, Jan 2017. ISSN 1949-3045. doi: 10.1109/TAFFC.2017.2693968. First published online on April 12, 2017.

Salud María Jiménez-Zafra, Mariona Taulé, M Teresa Martín-Valdivia, L Alfonso Ureña-López, and M Antónia Martí. SFU ReviewSP-NEG: a Spanish corpus annotated with negation for sentiment analysis. A typology of negation patterns. *Language Resources and Evaluation*, 52(2):533–569, 2018. doi: 10.1007/s10579-017-9391-x. URL http://dx.doi.org/10.1007/s10579-017-9391-x. First published online on May 22, 2017.

Salud María Jiménez-Zafra, Noa P. Cruz Díaz, Roser Morante, and María Teresa Martín-Valdivia. NEGES 2018: Workshop on Negation in Spanish. *Procesamiento del Lenguaje Natural*, 62:21–28, 2019a.

Salud María Jiménez-Zafra, Noa P Cruz Díaz, Roser Morante, and María Teresa Martín-Valdivia. NEGES 2019 Task: Negation in Spanish. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2019). CEUR Workshop Proceedings, CEUR-WS, Bilbao, Spain*, pages 329–341, 2019b.

Salud María Jiménez-Zafra, Roser Morante, Eduardo Blanco, María Teresa Martín-Valdivia, and L. Alfonso Ureña-López. Detecting Negation Cues and Scopes in Spanish. In *Proceedings of the Twelfth International Conference on Language Resources and Evaluation (LREC 2020)*, pages 1–10, Marseille, France, May 11-16 2020. European Language Resources Association (ELRA).

Alistair Kennedy and Diana Inkpen. Sentiment classification of movie and product reviews using contextual valence shifters. *Computational Intelligence*, 22(2):110–125, 2006.

Jin-Dong Kim, Tomoko Ohta, Sampo Pyysalo, Yoshinobu Kano, and Jun'ichi Tsujii. Overview of BioNLP'09 shared task on event extraction. In *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing: Shared Task*, pages 1–9. Association for Computational Linguistics, 2009.

Natalia Konstantinova, Sheila CM De Sousa, Noa P Cruz Díaz, Manuel J Mana López, Maite Taboada, and Ruslan Mitkov. A review corpus annotated for negation, speculation and their scope. In *Proceedings of LREC 2012*, pages 3190–3195, 2012.

Emanuele Lapponi, Jonathon Read, and Lilja Øvrelid. Representing and resolving negation for sentiment analysis. In *2012 IEEE 12th International Conference on Data Mining Workshops*, pages 687–692. IEEE, 2012.

Elizabeth D Liddy, Woojin Paik, Mary E McKenna, Michael L Weiner, S Yu Edmund, Theodore G Diamond, Bhaskaran Balakrishnan, and David L Snyder. User interface and other enhancements for natural language information retrieval system and method, February 15 2000. US Patent 6,026,388.

Bing Liu. *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge University Press, 2015.

Henry Loharja, Lluìs Padró, and Jordi Turmo. Negation Cues Detection Using CRF on Spanish Product Review Text at NEGES 2018. In *Proceedings of NEGES 2018: Work-*

shop on Negation in Spanish, CEUR Workshop Proceedings, volume 2174, pages 49–54, 2018.

Eugenio Martínez-Cámara, M Teresa Martín-Valdivia, M Dolores Molina-González, and L Alfonso Ureña-López. Bilingual experiments on an opinion comparable corpus. In *Proceedings of the 4th workshop on computational approaches to subjectivity, sentiment and social media analysis*, pages 87–93, 2013.

Carlos Henriquez Miranda, Jaime Guzmán, and Dixon Salcedo. Minería de Opiniones basado en la adaptación al español de ANEW sobre opiniones acerca de hoteles. *Procesamiento del Lenguaje Natural*, 56:25–32, 2016.

Kevin J Mitchell, Michael J Becich, Jules J Berman, Wendy W Chapman, John R Gilbertson, Dilip Gupta, James Harrison, Elizabeth Legowski, and Rebecca S Crowley. Implementation and Evaluation of a Negation Tagger in a Pipeline-based System for Information Extraction from Pathology Reports. In *Medinfo*, pages 663–667, 2004.

Roser Morante and Eduardo Blanco. * SEM 2012 shared task: Resolving the scope and focus of negation. In * *SEM 2012: The First Joint Conference on Lexical and Computational Semantics–Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)*, volume 1, pages 265–274, 2012.

Roser Morante and Caroline Sporleder. Special Issue on Modality and Negation. *Computational Linguistics*, 38(2), 2012.

Danielle L Mowery, Sumithra Velupillai, Brett R South, Lee Christensen, David Martinez, Liadh Kelly, Lorraine Goeuriot, Noemie Elhadad, Sameer Pradhan, Guergana Savova, et al. Task 2: ShARe/CLEF eHealth evaluation lab 2014. In *Proceedings of CLEF 2014*, 2014.

Lluís Padró and Evgeny Stanilovsky. FreeLing 3.0: Towards Wider Multilinguality. In *Proceedings of LREC 2012*, Istanbul, Turkey, May 2012.

Bo Pang and Lillian Lee. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the 42nd annual meeting on Association for Computational Linguistics*, page 271. Association for Computational Linguistics, 2004.

Thomas Edward Payne. *Describing morphosyntax: A guide for field linguists*. Cambridge University Press, 1997.

Livia Polanyi and Annie Zaenen. Contextual valence shifters. In James G. Shanahan, Yan Qu, and Janyce Wiebe, editors, *Computing Attitude and Affect in Text: Theory and Applications*, pages 1–10. Springer, Dordrecht, 2006.

Christopher Potts. On the negativity of negation. In *Proceedings of SALT 20: Semantics and Linguistic Theory*, pages 636–659, Vancouver, 2011a.

Christopher Potts. Sentiment symposium tutorial. In *Sentiment Analysis Symposium*, San Francisco, California, November, 2011b. Alta Plana Corporation. URL http://sentiment.christopherpotss.net/.

Zhong Qian, Peifeng Li, Qiaoming Zhu, Guodong Zhou, Zhunchen Luo, and Wei Luo. Speculation and negation scope detection via convolutional neural networks. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 815–825, 2016.

Filipe N. Ribeiro, Matheus Araújo, Pollyana Gonçalves, Marcos André Gonçalves, and

Fabrício Benevenuto. SentiBench: A benchmark comparison of state-of-the-practice sentiment analysis methods. *EPJ Data Science*, 5(23):1–29, 2016.

Paul Rozin and Edward B. Royzman. Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, 5(4):296–320, 2001.

Roser Saurí. *A Factuality Profiler for Eventualities in Text*. Ph.d. dissertation, Brandeis University, 2008.

Guergana K Savova, James J Masanz, Philip V Ogren, Jiaping Zheng, Sunghwan Sohn, Karin C Kipper-Schuler, and Christopher G Chute. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. *Journal of the American Medical Informatics Association*, 17(5):507–513, 2010.

Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D Manning, Andrew Ng, and Christopher Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 conference on empirical methods in natural language processing*, pages 1631–1642, 2013.

Vanesa Stricker, Ignacio Iacobacci, and Viviana Cotik. Negated Findings Detection in Radiology Reports in Spanish: an Adaptation of NegEx to Spanish. In *IJCAI-Workshop on Replicability and Reproducibility in Natural Language Processing: adaptative methods, resources and software, Buenos Aires, Argentina*, 2015.

Maite Taboada. Sentiment analysis: An overview from Linguistics. *Annual Review of Linguistics*, 2:325–347, 2016.

Maite Taboada, Caroline Anthony, and Kimberly D Voll. Methods for Creating Semantic Orientation Dictionaries. In *Proceedings of LREC 2016*, pages 427–432, 2006.

Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. Lexicon-based methods for sentiment analysis. *Computational Linguistics*, 37(2):267–307, 2011.

Maite Taboada, Radoslava Trnavac, and Cliff Goddard. On being negative. *Corpus Pragmatics*, 1(1):57–76, 2017.

Özlem Uzuner, Brett R South, Shuying Shen, and Scott L DuVall. 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. *Journal of the American Medical Informatics Association*, 18(5):552–556, 2011.

David Vilares, Miguel A Alonso, and Carlos Gómez-Rodríguez. Clasificación de polaridad en textos con opiniones en español mediante análisis sintáctico de dependencias. *Procesamiento del lenguaje natural*, 50:13–20, 2013.

David Vilares, Miguel A Alonso, and Carlos Gómez-Rodríguez. A syntactic approach for opinion mining on Spanish reviews. *Natural Language Engineering*, 21(1):139–163, 2015.

Veronika Vincze, György Szarvas, Richárd Farkas, György Móra, and János Csirik. The BioScope corpus: Biomedical texts annotated for uncertainty, negation and their scopes. *BMC bioinformatics*, 9(11):S9, 2008.

Ho-yen Wang. La negación en español, chino, inglés y alemán. *Encuentros en Catay*, 20: 129–157, 2006. ISSN 1023-6961.

Michael Wiegand, Alexandra Balahur, Benjamin Roth, Dietrich Klakow, and Andrés Montoyo. A survey on the role of negation in sentiment analysis. In *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*, pages 60–68, Uppsala, Sweden, 2010.