Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), pages 6902–6911
Marseille, 11–16 May 2020

Detecting Negation Cues and Scopes in Spanish

Salud María Jiménez-Zafra[1], Roser Morante[2], Eduardo Blanco[3], María Teresa Martín-Valdivia[1], L. Alfonso Ureña-López[1]
[1]SINAI, Computer Science Department, CEATIC, Universidad de Jaén
{sjzafra, maite, laurena}@ujaen.es
[2]CLTL Lab, Computational Linguistics, VU University Amsterdam
r.morantevallejo@vu.nl
[3]University of North Texas
eduardo.blanco@unt.edu

Abstract

In this work we address the processing of negation in Spanish. We first present a machine learning system that processes negation in Spanish. Specifically, we focus on two tasks: i) negation cue detection and ii) scope identification. The corpus used in the experimental framework is the SFU Review$_{\text{SP}}$-NEG. The results for cue detection outperform state-of-the-art results, whereas for scope detection this is the first system that performs the task for Spanish. Moreover, we provide a qualitative error analysis aimed at understanding the limitations of the system and showing which negation cues and scopes are straightforward to predict automatically, and which ones are challenging.

Keywords: negation processing, negation cues detection, scope identification, Spanish, SFU Review$_{\text{SP}}$-NEG corpus

1. Introduction

Negation is a phenomenon that "relates an expression *e* to another expression with a meaning that is in some way opposed to the meaning of *e*" (Horn and Wansing, 2017). It is present in all languages and is always marked. Negation has been studied from a theoretical perspective in virtually all languages, and determining the scope, that is, the part of the sentence affected by the negation cue, is critical to understanding sentences with negation (Ladusaw, 1992). Within computational linguistics, however, most previous work on negation targets English texts, albeit in a variety of domains. While negation processing systems are available for a few other languages, for Spanish the effort has focused mostly and only recently on annotation tasks. Yet Spanish is the second most widely spoken language in terms of native speakers (Lewis and Gary, 2015) and the third most common language used on the Internet.[1]

Identifying negation cues and scopes in general—and in Spanish in particular—is challenging. First, negation cues may be discontinuous and ambiguous, i.e., the same sequence of characters does not always indicate a negation. Second, scopes may span tokens before and after the negation cue, and scopes do not need to cover full sentences or phrases. In the SFU Review$_{\text{SP}}$-NEG corpus (Jiménez-Zafra et al., 2018) most scopes (75.47%) cover less than 30% of the sentence they belong to. Third, two or more negations may be present in a sentence and their scopes may overlap. Ex. (1) and (2) illustrate some of these challenges with sentences selected from user-generated reviews of the SFU Review$_{\text{SP}}$-NEG corpus.[2] The first *no* in Ex. (1) does not express negation. Ex. (2) contains multiple negations (*no-ninguna* and *no*), a discontinuous cue (*no-ninguna*), and overlapping scopes.

1. "Hasta que <u>no</u> me dejó el disco para que lo escuchase, [**no** paró]."
   Until he left the record for me to listen to, he didn't stop.
2. "Además de todo esto, [**no**^{1} existe **ninguna**^{1} garantía de [que reparándo la lavadora ahora **no**^{2} nos vuelva a suceder esto u otra cosa por el estilo el año que viene]^2]^1."
   On top of this, there is no guarantee that if we fix the washing machine now, the same or similar problem will not happen again next year.

In this paper, we target the computational treatment of negation cues and scopes in Spanish. We address two research questions: i) whether the methods that have been previously used to process negation in English are transferable to Spanish, and ii) where lies the complexity of negation processing in Spanish. In order to address the first question we develop a new negation processing system for Spanish, which will be made publicly available. In order to answer the second question we perform a qualitative error analysis.

The rest of the paper is organized as follows. Section 2 presents related work. In Section 3, negation features of the corpus and the experiments carried out are described. Section 4 presents the obtained results and a comparison with other systems. The qualitative analysis is reported on Section 5 and, finally, conclusions are summarized in Section 6.

2. Related Work

Negation is a well-studied phenomenon from a theoretical perspective (Horn, 2010; Horn, 1989). However, its computational treatment has not been extensively studied for

[1]"Number of Internet Users by Language", Internet World Stats, Miniwatts Marketing Group, 31 December 2017, accessed 20 February 2019. URL: `https://www.internetworldstats.com/stats7.htm`

[2]In the examples we mark in bold negation cues and enclose negation scopes between square brackets.

languages other than English. Its automatic detection and treatment is relevant in a wide range of applications, such as information extraction (Savova et al., 2010), machine translation (Baker et al., 2012) or sentiment analysis (Liu, 2015), where it is crucial to detect when a fragment of text expresses a different meaning due to the presence of negation.

The first attempts to process negation in English were mostly rule-based and focused on finding negated terms in clinical texts (Chapman et al., 2001; Mutalik et al., 2001; Goldin and Chapman, 2003; Auerbuch et al., 2004; Elkin et al., 2005; Boytcheva et al., 2005; Goryachev et al., ; Sanchez-Graillet and Poesio, 2007; Huang and Lowe, 2007; Rokach et al., 2008). The task of detecting negation scopes was introduced in 2008 as a machine learning sequence labelling task (Morante et al., 2008). Subsequently, three main types of approaches have been applied to processing negation: (i) rule-based systems, in an attempt to improve the NegEx algorithm, such as ConText (Harkema et al., 2009), DEEPEN (Mehrabi et al., 2015), and NegMiner (Elazhary, 2017); (ii) machine learning techniques (Agarwal and Yu, 2010; Li et al., 2010; Cruz Díaz et al., 2012; Cotik et al., 2016); and (iii) deep learning approaches (Qian et al., 2016; Ren et al., 2018; Lazib et al., 2018). Several shared tasks have addressed negation processing: the BioNLP'09 Shared Task 3 (Kim et al., 2009), the CoNLL-2010 shared task (Farkas et al., 2010), the i2b2 NLP Challenge (Uzuner et al., 2011), the *SEM 2012 Shared Task (Morante and Blanco, 2012) and the ShARe/CLEF eHealth Evaluation Lab 2014 Task 2 (Mowery et al., 2014).

Previous work has incorporated negation processing in sentiment analysis systems. Some systems use rules, but do not evaluate the processing of negation (Das and Chen, 2001; Polanyi and Zaenen, 2006; Kennedy and Inkpen, 2006; Jia et al., 2009). Other systems employ a lexicon of negation cues and predict the scope with CRFs using as features lowercased token strings, token PoS tags, token-wise distances from explicit negation cues and dependency syntax information (Councill et al., 2010), or a rich set of lexical and syntactic features, together with cue-dependant information (Lapponi et al., 2012a). Cruz et al. (2016) use SVM and lexical, syntactic and dependency features. They test the system in the SFU Review corpus (Konstantinova et al., 2012).

Work on processing negation in Spanish is relatively recent. Costumero et al. (2014), Stricker et al. (2015) and Cotik et al. (2016) develop systems for the identification of negation in clinical texts by adapting the NegEx algorithm (Chapman et al., 2001). Regarding product reviews, there are some works that treat negation as a subtask of sentiment analysis (Taboada et al., 2011; Vilares et al., 2013; Vilares et al., 2015; Jiménez-Zafra et al., 2015; Miranda et al., 2016; Amores et al., 2016; Jiménez-Zafra et al., 2019c).

The first systems that detect negation cues were developed in the framework of the 2018 and 2019 editions of NEGES, the Workshop on Negation in Spanish (Jiménez-Zafra et al., 2019a; Jiménez-Zafra et al., 2019b) and were trained on the SFU Review$_{SP}$-NEG corpus (Jiménez-Zafra et al., 2018). Participants (Fabregat et al., 2018; Loharja et al., 2018; Giudice, 2019; Beltrán and González, 2019; Domínguez-Mas et al., 2019; Fabregat et al., 2019) addressed the problem as a sequence labeling task using machine learning approaches. Deep learning algorithms and CRF algorithm were the predominant, being CRF the one that performed the best.

The negation processing system that we present is trained also on the SFU Review$_{SP}$-NEG corpus, but it is novel in that it performs scope detection, which no other system does for Spanish. Existing systems that process clinical texts identify negated entities and clinical findings, and those that process reviews detect only negation cues.

3. Methods

As in previous work, we take a machine learning approach and tackle the task in two steps: cue detection and scope identification. The system is trained on the only corpus that contains scope annotations for Spanish, SFU Review$_{SP}$-NEG.

3.1. Corpus

SFU Review$_{SP}$-NEG (Jiménez-Zafra et al., 2018) is a Spanish corpus of user-generated product reviews (9,446 sentences). The reviews belong to eight domains (cars, hotels, washing machines, books, cell phones, music, computers and movies) and are annotated with negation cues and their scopes. Negation cues that do not negate are also annotated. The corpus has a total of 4327 cues, out of which 3941 negate and 386 do not negate.

	Sentences		
	#	%	Avg. #tokens
0 negations	6,621	70.09	19.47
1 negation	2,028	21.47	28.78
2 negations	578	6.12	39.10
$\geq$3 negations	219	2.32	54.72
All	9,446	100.00	23.49

Table 1: Statistics about the sentences that contain negations in the SFU Review$_{SP}$-NEG corpus

As Table 1 shows, 21.5% of the sentences in SFU Review$_{SP}$-NEG have one negation, 6.12% have two negations, and 2.32% have three or more negations. Negation tends to occur in longer sentences: the average length of all sentences is 23.49 tokens, but the average length of sentences increases with the number of negations (28.78 tokens for 1 negation, 39.10 tokens for 2, and 54.72 for more than 2).

Table 3.1. provides counts of the most common negation cues in the corpus. While *no* accounts for 58.79% of negation cues, we note that there is a long tail of infrequent negation cues, making negation cue detection a challenging task.

Simple **negation cues**, which are expressed by one token such as *no*, in Ex. (3), are more frequent (79.85%). 4.72% of cues are a sequence of two or more contiguous tokens (*continuous cues*, e.g. *Ni nunca* in Ex. (4)) and 15.43% are expressed by two or more non-contiguous tokens (*discontinuous cues*, e.g. *No-nada* in Ex. (5)).

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

		#	%
Type	Simple (1 token)	3,147	79.85
	Continuous (> 1 token)	186	4.72
	Discontinuous (> 1 token)	608	15.43
Cue Tokens	no	2,317	58.79
	sin	282	7.16
	ni	151	3.83
	nada	125	3.17
	no-nadas	120	3.04
	nunca	76	1.93
	nadie	57	1.45
	tampoco	50	1.27
	no-ni	38	0.96
	Others	725	18.40

Table 2: Basic statistics for negation cues in the SFU Review$_{SP}$-NEG corpus

4. "[**Ni nunca**[1] quiso ser de [**nadie**[2]][2]][1]."
   He never wanted to depend on anybody.
5. "[**No** tengo **nada** en contra de Opel]."
   I don't have nothing against Opel.[3]

The **negation scope** always includes the corresponding negation cue, and may span only tokens before the cue (5.84%, see Table 3), only after the cue (68.56%), or both tokens before and after the cue (25.60%).

		#	%
Type	before cue	230	5.84
	after cue	2,702	68.56
	before and after cue	1,009	25.60
#Toks.	<3	564	14.31
	≥3 and <5	1,076	27.31
	≥5 and <8	1,332	33.81
	≥8	968	24.57
%sent.	<10%	1,081	27.43
	≥10% and <17%	928	23.55
	≥17% and <30%	965	24.49
	≥30%	935	23.72

Table 3: Basic statistics for negation scopes in the SFU Review$_{SP}$-NEG corpus

Scopes in SFU Review$_{SP}$-NEG span up to 43 tokens. Most scopes span between 3 and 7 tokens (61.12%), but almost 25% span more than 7 tokens. Finally, negation scopes almost always cover a small percentage of the sentence they belong to. Only 23.72% of negation scopes cover over 30% of the tokens in their sentence, and almost 51% cover less than 16% of the sentence tokens.

Each review of this corpus was automatically annotated at the token level with fine and coarse PoS-tags, PoS-types and lemmas using Freeling[4] (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. Moreover, we pre-processed it to add dependency relations using also Freeling.

In our experiments we use the corpus splits (train, development and test) provided for the shared task of the Workshop NEGES 2018 (Jiménez-Zafra et al., 2019a). The train, development and test splits consist of 264, 56 and 80 reviews respectively (33 reviews per domain in training, 7 reviews per domain in development and 10 reviews per domain in test). The distribution of negation cues and scopes present in them is the following: 2,511 for training, 594 for development and 836 for test.

3.2. Experiments

As in previous work, we model negation processing as a sequence labelling task. We choose a CRF algorithm (Lafferty et al., 2001) because it has been shown to be effective for this type of task (Morante et al., 2008; Councill et al., 2010; Lapponi et al., 2012b; Reitan et al., 2015; Loharja et al., 2018; Beltrán and González, 2019; Domínguez-Mas et al., 2019). CRF is well-suited to sequence modeling tasks because it makes predictions based not only on the current element, but also on other elements in the sequence; and negation cues and scopes are modeled as sequences of tokens.

We use the CRF implementation in CRFsuite (Okazaki, 2007) and scikit-learn (Pedregosa et al., 2011) with the L-BFGS training algorithm (default) and Elastic Net (L1 + L2) regularization.[5] Specifically, we train two classifiers: the first one takes as input a sentence and predicts the negation cue BIO labels, and the second one takes as input a sentence along with information about the predicted cues and predicts the scope BIO labels.

We train with the train split, select features based on results with the development set, and report results using the test set. Evaluation is performed in terms of precision (P), recall (R) and F1-score (F) measures, using the evaluation script released by the *SEM-2012 Shared Task[6] (Morante and Blanco, 2012).

The feature set is inspired by the work of Cruz et al. (2016), who train their system on the SFU Review corpus with negation and speculation annotations (Konstantinova et al., 2012). It is summarized in Table 4 and described below. We decided to use similar features because the SFU Review$_{SP}$-NEG corpus (Spanish) is the comparable version of the SFU Review corpus (English) (Taboada et al., 2006; Konstantinova et al., 2012).

Features for cue detection. We experimented with 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 tokens before and after the token in focus. We found that the most useful features were lemmas and part-of-speech tags. Therefore, we discarded the rest of features and conducted experiments to find out the optimal window for which lemma and PoS tags features should be added. We decided to use as features the lemma and PoS

[3]Literally, *I don't have nothing against Opel.* Replacing *nothing* (*nada*) with *anything* (*algo*) is incorrect in Spanish.

[4]http://nlp.lsi.upc.edu/freeling/index.php/node/1

[5]Parameters: algorithm='lbfgs', c1=0.1, c2=0.1, max_iterations=100, all_possible_transitions=True

[6]https://www.clips.uantwerpen.be/sem2012-st-neg/data.html

	Name	Description	Neg. cue?	Scope?
1, 2	current	Lemma and part-of-speech tag of t	✓	✓
3–30	token_window	Lemmas and part-of-speech tags of 7 tokens before and after than t	✓	✗
31	known_cue	Whether t was seen as a cue during training (B, I , B_I, or O)	✓	✗
32, 33	cue	Lemma and part-of-speech tag of nc	✗	✓
34	location	Location of current t with respect to nc (before, inside or after)	✗	✓
35	distance	Number of tokens between t and nc	✗	✓
36	chain_pos_f	Sequence of fine part-of-speech tags between t and nc	✗	✓
37	chain_pos_c	Same than *chain_pos_fine* but with coarse tags	✗	✓
38–41	{l,r}_tokens	Lemma and part-of-speech tags of the tokens to the left and right of t	✗	✓
42,43	rel_positions	Position of nc and t in the sentence over number of tokens in the sentence	✗	✓
44,45	dep_rel	Dependency type and direction (head or dependent) between t and nc	✗	✓
46, 47	heads	Part-of-speech tags of the first and second order syntactic heads of t	✗	✓
48, 49	is_ancestor	Whether t is an ancestor of nc and vice versa	✗	✓
50, 51	path_types	Dependency types in the syntactic path from t to nc and vice versa	✗	✓
52	path_types_dir	Same than path_types but including direction (up or down) and only for t	✗	✓
53	path_length	Length of path_types	✗	✓

Table 4: Features used to train the CRF classifiers to detect negation cues and scopes. We use t to refer to the token to be predicted, and nc to the negation cue

tags of the current token (features 1,2) as well as 7 tokens before and after (features 3-30). These features are positional, we do not use a bag-of-words representation. Additionally, we used a binary flag (known_cue) to indicate whether the token was seen as part of a negation cue in the training instances (feature 31)). This feature has four possible values: seen only as the first token of a cue (B), seen only as any token of a cue except the first (I), seen as both the first token of a cue and other positions (B_I)[7], and not seen (O). The rationale is that, while negation cues are ambiguous, they constitute a closed set (96.41% of cues in the test split are present in the training or development splits).

Features for detecting scopes. This feature set is more sophisticated and is the one used by Cruz et al. (2016) for detecting scopes in English: lemma and PoS tag of the current token and the cue in focus (features 1, 2, 32, 33), location of the token respect the cue (feature 34) (before, inside or after), distance in number of tokens between the cue and the current token (feature 35), chain of PoS tags and chain of types between the the cue and the token (features 36 and 37), lemma and PoS tags of the token to the left and right of the token in focus (features 38-41), relative position of the cue and the token in the sentence (features 42, 43), dependency relation and direction (head or dependent) between the token and the cue (features 44, 45), PoS tags of the first and second order syntactic heads of the token (features 46, 47), whether the token is ancestor of the token and vice versa (features 48, 49), dependency shortest path from the token in focus to the cue and vice versa (features 50, 51), dependency shortest path from the token in focus to the cue but including direction (up or down) (feature 52), length of the short path between the token and the cue (feature 53). During the feature tuning process, we discovered that the least informative features were *dep_rel* (44, 45), *is_ancestor* (48, 49), *heads* second order (46, 47) and *path_length* (53). Therefore, we conducted experiments removing all these features and the two least informative[8] (45, 49), but the results did not improve the initial experiment. Consequently, we decided to select the initial set (24 features in total) as features for reporting results with the test set.

4. Results

We take as baselines for **negation cue detection** the results of existing systems (Table 5). The comparison between the systems is possible and totally reliable because the results have been obtained on the same data set[9] and have been evaluated in the same way, using the evaluation script provided in the *SEM 2012 Shared Task: "Resolving the Scope and Focus of Negation"[10] (Morante and Blanco, 2012). Our results (87.32 F) outperform state-of-the-art results (86.45 F1, 84.09 F1, 82.99 F1, 80.50 F1, 67.97 F1 and 22.58 F1), although the UPC results are very close. Our system is in general accurate: precision is between 83% and 99% in the different domains (Table 6); and F1-score is between 81% and 93%. However, there are domains in which the recall does not exceed 80%. It seems that the most difficult negation cues to identify are present in the *washing machines* and *music* domains, which are the ones with the lowest recall. In Section 5. we provide an error analysis.

For **scope identification**, comparison with other scope detection systems is not possible because ours are the first

[7]The *B_I* value is useful for disambiguating the cues as many of them appear as a single token (e.g., *ni*, 'neither') and are also part of multiword cues (e.g., *ni siquiera*, 'not even').

[8]Their contribution is practically nil according to the chi-squared feature selection method.

[9]The test set used in our experiments is the same as the one we provided for the shared task "Negation cues detection" of the Workshop on Negation in Spanish: NEGES 2018 (Jiménez-Zafra et al., 2019a) and NEGES 2019 (Jiménez-Zafra et al., 2019b).

[10]https://www.clips.uantwerpen.be/sem2012-st-neg/data.html

	Cue		
	P	R	F
Aspie96 (Giudice, 2019)	18.80	28.34	22.58
UNED_2018 (Fabregat et al., 2018)	79.45	59.58	67.97
IBI (Domínguez-Mas et al., 2019)	91.22	72.16	80.50
UNED_2019 (Fabregat et al., 2019)	91.82	75.98	82.99
CLiC (Beltrán and González, 2019)	89.67	79.40	84.09
UPC (Loharja et al., 2018)	91.48	82.18	86.45
Our results	91.99	83.35	87.32

Table 5: System results on cue detection compared to existing results

results. We calculated two baselines: i) scope is identified as all tokens from the cue to the token previous to the end of the sentence; ii) scope is identified as all tokens from the cue to the token previous to the first punctuation mark. Table 7 shows the results for both baselines using predicted cues. Although precision is acceptable for both baseline systems, recall is very low. The first system only covers 20% of the scopes, approximately, and the second one 40%. This shows that scope identification is not an easy task and that the results obtained with our system are promising. We calculate the results of our system with *gold cues* in order to calculate the upper bound of the system, and with *predicted cues* (Table 6). The system is relatively accurate, precision is above 84% in all domains, except in the books domain, that is of 79.38%. However, the recall is not as high, on average 61.91. We will study in Section 5., what types of scopes have been the most difficult to predict.

The results obtained, with an F score of 73.35, suggest that the methods that have been previously proposed for English are transferable to Spanish. However, a question that remains open is whether the methodology used is optimal for Spanish. We perform an error analysis in order to detect where the system fails. It would be interesting to investigate also whether the errors of the English system are similar to the errors of the Spanish system, but we do not have the necessary resources to address this in this paper.

5. Qualitative Analysis

In order to better understand what are the limitations of the system and how can it be improved, we have performed a qualitative error analysis.

5.1. Negation Cues

The test set has a total of 836 negation cues. Specifically, there are 83 different negation cues, of which 15 are simple cues, 19 are continuous cues and 49 are discontinuous cues. Of these, the system has been able to detect 11 different simple cues, 11 different continuous cues and 21 different discontinuous cues, which indicates that the most difficult cues are the discontinuous ones. However, most system errors have been related to simple cues, followed by discontinuous and continuous cues. Errors due to negation cues predicted by the system and not annotated in test set, that is false positives, are distributed as follows: 86.97% correspond to simple cues, 8.51% to discontinuous cues and 5.32% to continuous cues. On the other hand, errors related to negation cues present in the test set and not predicted by the system, that is false negatives, are mainly due to discontinuous cues (56.25%), followed by simple cues (33.75%) and continuous cues (17.5%). It seems that continuous cues have been easier to predict.

The easiest continuous cues to predict have been *sin ningún*, *aún no*, *no tanto*, *todavía no*, *en absoluto*, *ni tan siquiera*, *ni jamás*, *ni nunca*, *sin apenas* and *ni siquiera*, which are cues present in dev+training set (except *ni siquiera*) and composed of two tokens. However, the system has not been able to learn the continuous cue *ya no*. Most of the errors with this cue are due to the system predicting the simple cue *no*, rather than the continuous cue *ya no*. For example, in Ex. (6),[11] the system has identified the negation cue *no* instead of *ya no*.

6. '**Ya** **no** cierra bien la puerta."
   Doesn't close the door well anymore.

Regarding discontinuous cues, some of them are always correctly predicted by the system: *sin-alguna*, *no-nunca*, *no-ningún*, *no-para nada*, *no-en absoluto*, *no-aun*, *no-demasiado*, *no-tampoco*, *ni-ninguna*, and *aun no-ninguna*. These cues have in common that they have between 2 and 5 intermediate tokens, which are covered by the token window used in the experimentation. Most of the errors with these cues are due to (i) negation cues not present in the dev+train set[12] or with a frequency of occurrence between 1 and 2[13], and (ii) the identification of *no* as simple cue instead of as one of the following discontinuous cue: *no-muy*, *no-tan* and *no-del todo*. For example, in Ex. (7), the system predicts *no* as negation cue, rather than the discontinuous cue *no muy*.

7. "Existe un adaptador que **no** sale **muy** caro."
   There is an adapter that is not very expensive.

Simple cues represent most of the cues in the test set. Although the system is able to predict correctly 95.95% of them, 62.07% of the errors affect these cues. The easiest simple cues to predict have been *sin*, *nunca*, *nadie*, *ninguna*, *ninguno* and *ningún*. Regarding errors, most of them are due to the most frequent cue in dev+train and test sets, *no*. Most of the system errors with this cue are related to the prediction of *no* as negation cue instead of the continuous cue *ya no* or the discontinuous cue of which it is part (Ex. 6). Moreover, in some cases it is wrongly identified as negation cue when it is part of a contrasting structure (Ex. 8). A significant part of the errors are also due to the negation cues *ni* and *nada*. Although they are in the dev+train set with a frequency of occurrence of 104 for *nada* and 112 for *ni*, the system sometimes identifies them as simple cues and sometimes as part of a discontinuous cue. Looking at the sentences incorrectly predicted by the system, it seems that there are cases in which the dev+train set is not consistent. We also find errors with the cue *jamás*, that is not

[11] Gold cues are in bold, system cues underlined.

[12] These are: *ya no-más*, *no-a menudo*, *no-ni una pizca*, *no solo-sino que*, *ningún-tampoco*, *nunca-mucha*, *no-casi*, *ni-no*, *npo-nada*, *no-ni de broma*, *no-pero nada de nada*, *no-ni borracho-ni al borde del coma etílico*, *sin-mucho*, *no-siente*, *ningún-nunca*, *no-no-nunca*, *no-casi nunca*, *ni tampoco*, *no-ni una sola palabra* and *sin-una palabra*.

[13] These are: *tampoco-tan*, *no-no*, *ya no-nada*, *no-totalmente*, *no-absolutamente nada*, *no-todavía*, and *no-nada de*

	Cue			Scope (gold cues)			Scope (predicted cues)		
	P	R	F	P	R	F	P	R	F
Cars	93.44	83.82	88.37	100	61.76	76.36	90.48	52.88	69.09
Hotels	98.11	88.14	92.86	100	71.19	83.17	97.50	66.1	78.79
Washing machines	94.44	73.91	82.92	100	72.46	84.03	93.75	65.22	76.92
Books	83.47	80.16	81.78	100	67.06	80.28	79.38	61.11	69.06
Cell phone	90.57	84.21	87.27	100	68.42	81.25	87.50	61.4	72.16
Music	95.83	79.31	86.79	100	66.67	80.00	94.34	57.47	71.43
Computers	89.29	92.59	90.91	100	61.73	76.34	84.75	61.73	71.43
Movies	90.79	84.66	87.62	100	72.39	83.98	88.98	69.33	77.94
All	91.99	83.35	87.32	100	67.71	80.68	89.59	61.91	73.35

Table 6: System results on the test set for cue and scope detection

	From cue to end of sentence			From cue to first punctuation mark		
	P	R	F	P	R	F
Cars	71.43	14.71	24.40	87.10	39.71	54.55
Hotels	88.89	13.56	23.53	94.44	28.81	44.15
Washing machines	83.33	21.74	34.48	90.00	39.16	54.55
Books	55.56	19.84	29.24	70.37	37.70	49.10
Cell phone	64.29	15.79	25.35	83.87	45.61	59.09
Music	84.21	18.39	30.19	91.89	39.08	54.84
Computers	68.96	24.69	36.36	79.07	41.98	54.84
Movies	76.27	27.61	40.54	84.62	47.24	60.63
All	74.12	19.54	30.51	85.17	39.91	53.97

Table 7: Baseline results for scope detection using predicted cues

correctly identified by the system in most cases, not even in simple sentences such as Ex. (9).

8. "No exige sino aquello que se le da."
   He demands only that which is given to him.
9. "**Jamás** compréis un ordenador de marca"
   Never buy a branded computer.
10. "Cuánta es su pequeñez y, sin embargo, qué ansia de perdurar."
    How small he is, and yet how eager he is to endure.

In short, we can say that most of the errors are due to: (i) cues identified as simple instead of as continuous or discontinuous (Ex. 6 and Ex. 7) , (ii) cues wrongly identified as negation cues (Ex. 10) and (iii) cues identified as negation cues when they are part of a contrasting structure (Ex. 8). This suggest that the system is not able to identify low frequency cues and it is not able to disambiguate cues. As future work we would like to experiment with starting the cue detection process with word sense disambiguation.

5.2. Negation Scope

The error analysis of scopes[14] is based on predictions of the scope processing module using gold cues. We can make several general observations. (i) The scopes produced by the system are mostly continuous. We found only 2 cases in which the scope was discontinuous without being correct, since the system predicted more than one beginning of the scope. (ii) The system never includes in the scope punctuation that signals the end of the sentence, while the gold annotations does. We do not consider this to be a limitation. (iii) In the gold data (dev+train), a majority of scopes begin in the negation cue (69.50%). As a consequence, the system tends to take the negation cue as the start of the scope. The number of system scopes beginning in cue is 544. From this 452 are correct and 92 incorrect. (iv) The system includes generally all tokens of a syntactic phrase in the scope, so it does not split phrases. Ex. (11) is an exception, because the system finishes the scope in the middle of the noun phrase. However, some syntactic structures, such as coordination, pose challenges (Ex. 12).

11. {[**No** me lo pensaré dos} veces].
    I won't think twice about it.
12. la batería ... [{**no** dura más de un día} y medio]
    The battery does not last more than a day and a half.

(v) Except for a few cases with the cue *ningún*, as in Ex. (13), the system always predicts a scope for a cue, although in two cases the scope contains only the negation cue, whereas the gold scopes are longer.

13. ... a mi gusto no cuenta [con {**ningún**} temazo]
    To my liking it doesn't have any hit.

Based on these observation we can predict that a scope will be easy to learn if it begins at the negation cue, it is continuous, and ends in the token previous to the final punctuation mark of the sentence, regardless of the type of negation cue and size of the scope.

In order to determine where the difficulty of predicting the scope lies, we have analyzed 170 scopes produced by the system which are different from the gold scopes. In most of the cases either the beginning or the end of the scope are wrong and only in a few cases there are errors both at the beginning and at the end. The errors at the beginning of the scope are due to the system not including the subject, be it

[14]In the examples, gold scopes are between square brackets and system scopes between curly brackets.

nominal or pronominal (Ex. 14),[15] or the adverbial complements of the verb (Ex. 15) when gold does include them, or including them when gold does not (Ex. 16, 17). A cause of these errors could be that the features extracted are based on wrong syntactic information, but the analysis of the automatically generated dependency tree reveals this is not the case. This would indicate that errors are independent of the quality of the syntactic information. Another potential cause of errors can be the observed inconsistency of some gold annotations. In Ex. (14) the gold annotations include the subject in the scope, whereas in Ex. (16) the subject is not included.

14. ... el motor ... [que además {**no** es el que menos gasta}] ...
    The engine that also is not the one that spends less.
15. Vamos, [por 11900 euros {yo **no** me lo compraba}].
    For 11900 euros I didn't buy it.
16. Los plásticos resultan demasiado evidentes y {la tapicería [**no** es nada del otro mundo]}.
    Plastics are too obvious and upholstery is nothing new.
17. {En mi opinión, [no lo compréis]}
    In my opinion, don't buy it.

The errors due to wrong system predictions at the end of the scope are mostly due to the system adding complements when the gold does not. In Ex. (18) the system adds to the scope a clause that acts as causal complement of the verb; in Ex. (19) it adds a relative clause that is a complement of the direct object of the negated verb, and in Ex. (20), a verbal phrase that is not syntactically dependent on the verb included in the gold scope. Everything indicates that the system seems to be extending the scope to the final punctuation mark. However, there are also some errors due to the system shortening the scope, as in Ex. (21), where the second element of the coordinated adjectival phrase is not included, or Ex. (22) where the complement of the noun *cable* is not included.

18. Por cierto, [{no lo probé] porque en ningún sitio lo tenían}.
    By the way, I didn't prove it because nowhere did they have it.
19. ... [{que **no** se adapta a la caja de cambio] que lleva}.
    It doesn't fit the gearbox it carries.
20. La pila de ropa [{**sin** lavar] sigue subiendo}.
    The pile of unwashed clothes keeps coming up.
21. {[**No** me sentí **ni** libre} **ni** poderoso] en aquella suntuosa maãna.
    I felt neither free nor powerful in that sumptuous morning.
22. ... [{**sin** cable} para el pc]
    No cable for the PC.

Errors at the beginning and at the end of the scope are less frequent. In Ex. (23) the system starts the scope at the negation cue and ends it after the closing bracket, not included in the gold scope – which shows another inconsistency in the annotation of the data, since the opening bracket is included in the gold scope. In Ex. (24) the system excludes the subject of the verb affected by the cue and adds the quotation marks at the end. In Ex. (25) the system excludes the subject, but includes a causal complement at the end.

23. ... [(que encima, según Opel, {**no** es un fallo}]) ...
    According to Opel, it's not a failure.
24. "No trates de arreglar [lo que {**no** está descompuesto]"}
    Don't try to fix what's not broken.
25. [Los antiguos PC, {**no** metían casi ruido] debido a la carencia de ventiladores} ...
    The old PCs didn't make any noise due to the lack of fans.

In sum, it seems that a main source of errors might be the inconsistency of annotations in the training corpus, where some scopes include several complements of the verb and others do not. Starting from this, it would be difficult to improve the quality of the system without previously improving the quality of the annotations. Another source of errors are complex syntactic structures such as coordination. An open question for future work is whether adding more complex syntactic information in the features would improve the performance of the system. Finally, discontinuous scopes are challenging. In future work we would like to investigate with classifying syntactic constituents, instead of tokens, using richer syntactic information.

6. Conclusions

We have presented a machine learning negation processing system for Spanish which is based on a system developed for English. For cue identification the system outperforms state-of-the-art results, while for scope detection we provide the first experimental results. The results of the system indicate that the methods used for English are transferable to Spanish.

However, a qualitative error analysis has shown that the methods applied are not optimal. Correctly detecting a frequent simple cue such as *no* remains a challenge (it causes 54.26% of system errors), as well as detecting discontinuous and infrequent cues. The ambiguity of some cues remains a challenge, as well as the cases where a simple cue is part of a discontinuous cue, specially with the cues *no*, *ni* and *nunca*. As for scopes, a scope will be easy to learn if it begins at the negation cue, it is continuous, and ends in the token previous to the final punctuation mark of the sentence, regardless of the type of negation cue and size of the scope. However the system has problems in determining whether the subject and adverbial complements of the verb are included in the scope, as well as the elements of coordination structures. Last, but not least, one of the problems detected are the inconsistent annotations in the training corpus.

For future work we intend to address several issues: i) reviewing the corpus to resolve inconsistent annotations; ii) incorporating word sense disambiguation mechanisms previous to cue detection and experimenting with adding features from word embeddings; iii) experimenting with adding more sophisticated syntactic features for scope detection in order to properly determine the beginning and end of the scopes; iv) experimenting with classifying syntactic constituents instead of tokens in order to better cap-

[15]Gold scopes are marked between square brackets, system scopes between curly brackets.

ture discontinuous scopes. Additionally, we will develop a more complex methodology for error analysis of complex linguistic phenomena such as scope that provides a deeper understanding of a system's output.

7. Acknowledgements

This work was partially supported by a grant from the Ministerio de Educación Cultura y Deporte (MECD Scholarship FPU014/00983), Fondo Europeo de Desarrollo Regional (FEDER) and, REDES project (TIN2015-65136-C2-1-R) and LIVING-LANG project (RTI2018-094653-B-C21) from the Spanish Government. RM was supported by the Netherlands Organization for Scientific Research (NWO) via the Spinoza-prize awarded to Piek Vossen (SPI 30-673, 2014-2019). This material is based upon work supported partly by the National Science Foundation under Grant No. 1845757. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

8. Bibliographical References

Agarwal, S. and Yu, H. (2010). Biomedical negation scope detection with conditional random fields. *Journal of the American medical informatics association*, 17(6):696–701.

Amores, M., Arco, L., and Barrera, A. (2016). Efectos de la negación, modificadores, jergas, abreviaturas y emoticonos en el análisis de sentimiento. In *IWSW*, pages 43–53.

Auerbuch, M., Karson, T., Ben-Ami, B., Maimon, O., and Rokach, L. (2004). Context-sensitive medical information retrieval. *Studies in health technology and informatics*, 107(Pt 1):282–286.

Baker, K., Bloodgood, M., Dorr, B. J., Callison-Burch, C., Filardo, N. W., Piatko, C., Levin, L., and Miller, S. (2012). Modality and negation in simt use of modality and negation in semantically-informed syntactic mt. *Computational Linguistics*, 38(2):411–438.

Beltrán, J. and González, M. (2019). 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. CEUR-WS.

Boytcheva, S., Strupchanska, A., Paskaleva, E., Tcharaktchiev, D., and Str, D. G. (2005). Some aspects of negation processing in electronic health records. In *Proc. of International Workshop Language and Speech Infrastructure for Information Access in the Balkan Countries*, pages 1–8.

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

Costumero, R., López, F., Gonzalo-Martín, C., Millan, M., and Menasalvas, E. (2014). An approach to detect negation on medical documents in Spanish. In *International Conference on Brain Informatics and Health*, pages 366–375. Springer.

Cotik, V., Stricker, V., Vivaldi, J., and Rodríguez Hontoria, H. (2016). 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.

Councill, I., McDonald, R., and Velikovich, L. (2010). 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, July. University of Antwerp.

Cruz, N. P., Taboada, M., and Mitkov, R. (2016). 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.

Cruz Díaz, N. P., Maña López, M. J., Vázquez, J. M., and Álvarez, V. P. (2012). A machine-learning approach to negation and speculation detection in clinical texts. *Journal of the Association for Information Science and Technology*, 63(7):1398–1410.

Das, S. and Chen, M. (2001). Yahoo! for Amazon: Extracting Market Sentiment from Stock Message Boards. In *Proceedings of the Asia Pacific Finance Association Annual Conference (APFA)*.

Domínguez-Mas, L., Ronzano, F., and Furlong, L. I. (2019). 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. CEUR-WS.

Elazhary, H. (2017). Negminer: An automated tool for mining negations from electronic narrative medical documents. *International Journal of Intelligent Systems and Applications*, 9(4):14.

Elkin, P. L., Brown, S. H., Bauer, B. A., Husser, C. S., Carruth, W., Bergstrom, L. R., and Wahner-Roedler, D. L. (2005). A controlled trial of automated classification of negation from clinical notes. *BMC medical informatics and decision making*, 5(1):13.

Fabregat, H., Martínez-Romo, J., and Araujo, L. (2018). 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.

Fabregat, H., Duque, A., Martínez-Romo, J., and Araujo, L. (2019). 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. CEUR-WS.

Farkas, R., Vincze, V., Móra, G., Csirik, J., and Szarvas, G. (2010). The CoNLL-2010 Shared Task: Learning to Detect Hedges and their Scope in Natural Language Text. In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning*, pages 1–12, Uppsala, Sweden, July. Association for Computational Linguistics.

Giudice, V. (2019). 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. CEUR-WS.

Goldin, I. and Chapman, W. W. (2003). Learning to detect negation with 'not'in medical texts. In *Proc Workshop on Text Analysis and Search for Bioinformatics, ACM SIGIR*.

Goryachev, S., Sordo, M., Zeng, Q. T., and Ngo, L.). Implementation and evaluation of four different methods of negation detection. Technical report.

Harkema, H., Dowling, J. N., Thornblade, T., and Chapman, W. W. (2009). Context: an algorithm for determining negation, experiencer, and temporal status from clinical reports. *Journal of biomedical informatics*, 42(5):839–851.

Horn, L. R. and Wansing, H. (2017). Negation. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Spring 2017 edition.

Horn, L. (1989). *A natural history of negation*. Chicago University Press, Chicago.

Horn, L. (2010). *The expression of negation*. De Gruyter Mouton, Berlin.

Huang, Y. and Lowe, H. J. (2007). A novel hybrid approach to automated negation detection in clinical radiology reports. *Journal of the American medical informatics association*, 14(3):304–311.

Jia, L., Yu, C., and Meng, W. (2009). 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.

Jiménez-Zafra, S. M., Martínez-Cámara, E., Martín-Valdivia, M. T., and Molina-González, M. D. (2015). Tratamiento de la negación en el análisis de opiniones en español. *Procesamiento del Lenguaje Natural*, 54:367–44.

Jiménez-Zafra, S. M., Taulé, M., Martín-Valdivia, M. T., Ureña-López, L. A., and Martí, M. A. (2018). 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, Jun.

Jiménez-Zafra, S. M., Cruz Díaz, N. P., Morante, R., and Martín-Valdivia, M. T. (2019a). NEGES 2018: Workshop on Negation in Spanish. *Procesamiento del Lenguaje Natural*, 62:21–28.

Jiménez-Zafra, S. M., Cruz Dıaz, N. P., Morante, R., and Martın-Valdivia, M. T. (2019b). NEGES 2019 Task: Negation in Spanish. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2019). CEUR Workshop Proceedings, CEUR-WS, Bilbao, Spain*.

Jiménez-Zafra, S., Martín-Valdivia, M., Martínez-Cámara, E., and Ureña-López, L. (2019c). Studying the scope of negation for spanish sentiment analysis on twitter. *IEEE Transactions on Affective Computing*, 10(1):129–141, Jan. First published online on April 12, 2017.

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

Kim, J.-D., Ohta, T., Pyysalo, S., Kano, Y., and Tsujii, J. (2009). 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.

Konstantinova, N., De Sousa, S. C., Díaz, N. P. C., López, M. J. M., Taboada, M., and Mitkov, R. (2012). A review corpus annotated for negation, speculation and their scope. In *Proceedings of LREC 2012*, pages 3190–3195.

Ladusaw, W. A. (1992). Expressing negation. In *Semantics and linguistic theory*, volume 2, pages 237–260.

Lafferty, J. D., McCallum, A., and Pereira, F. C. N. (2001). Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In *Proceedings of the Eighteenth International Conference on Machine Learning*, ICML '01, pages 282–289, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Lapponi, E., Read, J., and Øvrelid, L. (2012a). Representing and resolving negation for sentiment analysis. In *2012 IEEE 12th International Conference on Data Mining Workshops*, pages 687–692. IEEE.

Lapponi, E., Velldal, E., Ovrelid, L., and Read, J. (2012b). Uio2: Sequence-labeling negation using dependency features. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics - Volume 1: Proceedings of the Main Conference and the Shared Task*, SemEval '12, pages 319–327, Stroudsburg, PA, USA. Association for Computational Linguistics.

Lazib, L., Qin, B., Zhao, Y., Zhang, W., and Liu, T. (2018). A syntactic path-based hybrid neural network for negation scope detection. *Frontiers of Computer Science*, pages 1–11.

Lewis, M. P. and Gary, F. (2015). Simons, and charles d. fennig (eds.). *Ethnologue: Languages of the world,*, pages 233–62.

Li, J., Zhou, G., Wang, H., and Zhu, Q. (2010). Learning the scope of negation via shallow semantic parsing. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 671–679. Association for Computational Linguistics.

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

Loharja, H., Padró, L., and Turmo, J. (2018). Negation Cues Detection Using CRF on Spanish Product Review Text at NEGES 2018. In *Proceedings of NEGES 2018: Workshop on Negation in Spanish, CEUR Workshop Proceedings*, volume 2174, pages 49–54.

Mehrabi, S., Krishnan, A., Sohn, S., Roch, A. M., Schmidt, H., Kesterson, J., Beesley, C., Dexter, P., Schmidt, C. M., Liu, H., et al. (2015). Deepen: A negation detection system for clinical text incorporating dependency relation into negex. *Journal of biomedical informatics*, 54:213–219.

Miranda, C. H., Guzmán, J., and Salcedo, D. (2016).

Minería de opiniones basado en la adaptación al español de anew sobre opiniones acerca de hoteles. *Procesamiento del Lenguaje Natural*, 56:25–32.

Morante, R. and Blanco, E. (2012). *SEM 2012 Shared Task: Resolving the Scope and Focus of Negation. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics (*SEM 2012)*, pages 265–274, Montréal, Canada, June.

Morante, R., Liekens, A., and Daelemans, W. (2008). Learning the scope of negation in biomedical texts. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 715–724. Association for Computational Linguistics.

Mowery, D. L., Velupillai, S., South, B. R., Christensen, L., Martinez, D., Kelly, L., Goeuriot, L., Elhadad, N., Pradhan, S., Savova, G., et al. (2014). Task 2: Share/clef ehealth evaluation lab 2014. In *Proceedings of CLEF 2014*.

Mutalik, P. G., Deshpande, A., and Nadkarni, P. M. (2001). Use of general-purpose negation detection to augment concept indexing of medical documents: a quantitative study using the umls. *Journal of the American Medical Informatics Association*, 8(6):598–609.

Okazaki, N. (2007). CRFsuite: a fast implementation of Conditional Random Fields (CRFs).

Padró, L. and Stanilovsky, E. (2012). Freeling 3.0: Towards Wider Multilinguality. In Nicoletta Calzolari, et al., editors, *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012)*, pages 2473–2479, Istanbul, Turkey, May. European Language Resources Association (ELRA). ACL Anthology Identifier: L12-1224.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Polanyi, L. and Zaenen, A. (2006). Contextual valence shifters. In James G. Shanahan, et al., editors, *Computing Attitude and Affect in Text: Theory and Applications*, pages 1–10. Springer, Dordrecht.

Qian, Z., Li, P., Zhu, Q., Zhou, G., Luo, Z., and Luo, W. (2016). 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.

Reitan, J., Faret, J., Gambäck, B., and Bungum, L. (2015). Negation scope detection for twitter sentiment analysis. In *Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 99–108.

Ren, Y., Fei, H., and Peng, Q. (2018). Detecting the scope of negation and speculation in biomedical texts by using recursive neural network. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 739–742. IEEE.

Rokach, L., Romano, R., and Maimon, O. (2008). Negation recognition in medical narrative reports. *Information Retrieval*, 11(6):499–538.

Sanchez-Graillet, O. and Poesio, M. (2007). Negation of protein–protein interactions: analysis and extraction. *Bioinformatics*, 23(13):i424–i432.

Savova, G. K., Masanz, J. J., Ogren, P. V., Zheng, J., Sohn, S., Kipper-Schuler, K. C., and Chute, C. G. (2010). 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.

Stricker, V., Iacobacci, I., and Cotik, V. (2015). 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*.

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

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

Vilares, D., Alonso, M. A., and Gómez-Rodríguez, C. (2013). Clasificación de polaridad en textos con opiniones en espanol mediante análisis sintáctico de dependencias. *Procesamiento del lenguaje natural*, 50:13–20.

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

9. Language Resource References

Jiménez-Zafra, S. M., Taulé, M., Martín-Valdivia, M. T., Ureña-López, L. A., and Martí, M. A. (2018). 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, Jun.

Konstantinova, N., De Sousa, S. C., Díaz, N. P. C., López, M. J. M., Taboada, M., and Mitkov, R. (2012). A review corpus annotated for negation, speculation and their scope. In *Proceedings of LREC 2012*, pages 3190–3195.

Taboada, M., Anthony, C., and Voll, K. D. (2006). Methods for Creating Semantic Orientation Dictionaries. In *Proceedings of LREC 2016*, pages 427–432.