Universidad de Jaén

Escuela de Doctorado

DOCTORAL THESIS

●

Negation Processing in Spanish and its Application to Sentiment Analysis

PRESENTED BY:

Salud María Jiménez Zafra

SUPERVISED BY:

Ph.D. María Teresa Martín Valdivia

Ph.D. L. Alfonso Ureña López

JAÉN, July 10, 2019

To my family, the pillar of my life.
– Salud María Jiménez Zafra

Acknowledgements

This doctoral thesis[1] is the result of a long and grateful journey in which I have been supported by many people to whom I would like to show my gratitude.

My deepest thanks to my supervisors, Maite and Alfonso. Thank you for introducing me to the wonderful world of Natural Language Processing, for sharing your knowledge with me and for lovingly guiding me along this path. Thank you for helping me to get here.

I am also very grateful to all the other members of the SINAI group (Miguel Ángel, Loles, Arturo, Manuel Carlos, Manolo, Fernando, Flor and Pilar). Thank you for helping me to grow professionally and for making me feel at home. A special mention is reserved to Miguel Ángel for all the advice he has given me and for always being willing to help me.

I cannot forget Andrea Esuli and Roser Morante, great researchers with whom I have had the opportunity to collaborate thanks to the research stays I have made. Thank you very much for your hospitality and for everything I have learned from you.

On a personal level, the greatest thanks of this doctoral thesis is for my loved ones. Thanks to my parents, my brother and my grandmother, for their continuous and incomparable love, for their help, support, and for always believing in me. I have learned from you that dreams can come true with constancy and illusion. Thanks to my husband Juan Luis and to my little Paula, who will soon be born, who have accompanied me along this journey giving me their affection, support and understanding and encouraging me in the moments that I have needed them most. You are the reason for my daily happiness. You are the motor that moves my life.

[1]This research has been possible thanks to the financial support of the Ministerio de Eduación, Cultura y Deporte (MECD - scholarship FPU014/00983).

Abstract

Natural Language Processing is the area of Artificial Intelligence that aims to develop computationally efficient mechanisms to facilitate communication between people and machines through natural language. To ensure that machines are capable of processing, understanding and generating human language, a wide range of linguistic phenomena must be taken into account, such as negation, irony or sarcasm, which are used to give words a different meaning.

This doctoral thesis focuses on the study of negation, a complex linguistic phenomenon that we use in our daily communication. In contrast to most of the existing studies to date, it is carried out on Spanish texts, because i) it is the second language with most native speakers, ii) it is the third language most used on the Internet, and iii) there are no negation processing systems available on this language.

Negation has been widely studied from a theoretical perspective, and less from an applied point of view. However, the computational treatment of this phenomenon is of growing interest because it is relevant for a wide range of Natural Language Processing applications such as sentiment analysis, information retrieval, information extraction or machine translation, where it is crucial to know when the meaning of a part of the text changes due to the presence of negation.

The objective of this doctoral thesis is to advance in the processing of negation in Spanish and to show the importance of the computational treatment of negation for Natural Language Processing systems. For this purpose, an exhaustive study of negation is carried out, incorporating negation processing systems, corpora and sentiment analysis systems in which negation has been taken into account. In addition, a typology of negation patterns in Spanish is defined, which is applied for the annotation of a corpus with negation, the SFU Review$_{\text{SP}}$-NEG corpus. This corpus is used to develop a Spanish negation processing system which is applied to sentiment analysis in order to improve the predictive capacity of opinion classification systems that are so in demand today. Finally, NEGES has been launched, the first initiative promoting negation research in Spanish for which three editions have already been held in the context of the International Conference of the Spanish Society for Natural Language Processing (SEPLN).

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Speaking is one of humanity's greatest cultural achievements. Thinking, speaking and writing are actions that require a common element: language. Depending on our birthplace and other factors we know one or several languages, but all have the same goal: to enable communication. Since the first machines were devised and built, surely there was someone who dreamed of one that would speak and answer our questions, in short, that would facilitate our life, and with that end Natural Language Processing arises.

What is Natural Language Processing? I am absolutely sure that all of us use the *Google* search engine and that most of us use *Ok Google* or *Siri* and that, more than once, we have used *Google translator* or *DeepL* to translate the text of an indicator or a menu. All of these application have been developed thanks to Natural Language Processing. More formally, Natural Language Processing is the area of Artificial Intelligence that aims to develop computationally efficient mechanisms to facilitate communication between people and machines through natural language.

Natural Language Processing is the great challenge of Artificial Intelligence because computers must be able to process, understand and generate natural language. If we want to develop systems that approach human understanding, we must incorporate in them the treatment of a diversity of linguistic phenomena such as negation, speculation, uncertainty, irony or sarcasm.

This doctoral thesis focuses on the study of one of the main linguistic phenomena used by people in their daily communication: *negation*. Furthermore, in contrast to most of the existing studies to date, it is carried out on *Spanish* texts, because i) it is the second language with most native speakers, ii) it is the third language most used on the Internet, and iii) there are no negation processing systems available on this language.

1.1 Motivation

The idea of focusing this doctoral thesis on the study of negation in Spanish texts arises as a consequence of reading the work of Bing Liu, *"Sentiment analysis: Mining opinions, sentiments, and emotions"* (Liu, 2015). In this book, Bing Liu expounds that there are several open challenges for the classification of opinionated texts that are related with the treatment of some linguistic phenomena, such as negation. This fact provokes an enormous interest in the study of this phenomenon because our findings could be a breakthrough for Natural Language Processing systems.

This doctoral thesis thus begins with the study of the role of negation in the task of sentiment analysis, but it is extended to the study of negation processing in Spanish texts because there are no negation processing systems available for this language and it is a universal linguistic phenomenon with a great qualitative impact in a wide range of applications of Natural Language Processing, including sentiment analysis.

The large amount of content that is published every day on the Internet has generated great interest in the opinions and emotions that are shared in this environment. This user-generated content is useful for marketing strategies because it can be used to measure and monitor customer satisfaction. It is a quick way to find out what customers liked and what they did not like. Moreover, micro-blogging sites such as Twitter are being used to measure voting intention, people's moods and even to predict the success of a film. The study of negation in sentiment analysis is very important because if negation is present in a sentence and it is not taken into account, a system can extract a completely different opinion than the one published by the user. In Example (1) we can find a positive opinion that changes to negative if negation is present as in Example (2), or by contrast, in Example (3) there is a positive opinion in which negation is present whose meaning changes if it does not have negation as in Example (4).

1. El teléfono móvil funciona bien.

   The mobile phone works well.

2. El teléfono móvil **no** funciona bien.

   The mobile phone does not work well.

3. **No** he encontrado un teléfono móvil que funcione mejor que el mío.

   I have not found a mobile phone that works better than mine.

4. He encontrado un teléfono móvil que funciona mejor que el mío.

   I have found a mobile phone that works better than mine.

Moreover, the presence of negation in a sentence can have enormous consequences in many real situations. For example, detecting negated concepts in clinical texts is crucial because they often refer to concepts that are explicitly not present in the patient, for example, to document the process of ruling out a diagnosis:

> *"In clinical reports the presence of a term does not necessarily indicate the presence of the clinical condition represented by that term. In fact, many of the most frequently described findings and diseases in discharge summaries, radiology reports, history and physical exams, and other transcribed reports are denied in the patient."*
>
> (Chapman et al., 2001, p. 301)

Not recognizing these negated concepts can cause problems because the diagnosis of a patient will be totally different if the concept "tumor" is recognized in the sentence "There is no evidence of a tumor." and negation is not detected.

Another example that shows the relevance of processing negation are warning systems. We just have to think in a warning system created from information published on social networks and the impact generated if the system recognizes "plane" and "crash" in the tweet "I can not believe the airplane finally did not crash, it was so close", but it does not detect negation. It would create a false alert.

We might think that, given the fact that negation is so crucial in language, most Natural Language Processing pipelines incorporate negation modules and that the computational linguistics community has already addressed this phenomenon. However, this is not the case. Not even *Google* deals properly with negation in Spanish. For example, the search "películas que no sean de aventuras", returns adventure movie, and the search "recetas que no tengan tomate", returns recipes with tomato, whereas they should return non-adventure movies and recipes without tomato, respectively.

1.2 Objective

The objective of this doctoral thesis is to advance in the processing of negation in Spanish and to show the importance of the computational treatment of negation for Natural Language Processing systems.

1.3 Difficulty of the task

Negation has been widely studied from a theoretical perspective (L. R. Horn, 1989; L. Horn, 2010; Morante & Sporleder, 2012a), and less from an applied point of view. Its computational treatment has not been solved yet, due to its complexity, the multiple linguistic forms in which it can appear (syntactic, lexical, morphological) and the different ways it can act on the words within its scope. Negation can be expressed fundamentally with the use of syntactically independent negation words (Example 5), with words whose meaning has a negative component (Example 6) or by means of prefixes (Example 7). The elements used to represent negation explicitly in a text are called negation cues, and they can be *simple*, if they are expressed by a single token (e.g., "jamás" *[never]* in Example 5), *continuous*, if they are composed of a sequence of two or more contiguous tokens (e.g., "ni siquiera" *[not even]* in Example 8), or *discontinuous*, if they consist of a sequence of two or more non-contiguous tokens (e.g., "no...nada" *[nothing]* in Example 9).

5. **Jamás** recomendaría comprar este producto.

   I would never recommend buying this product.

6. Ella **ignoraba** que habíamos estado en Jaén.

   She did not know we had been in Jaén.

7. Estaba **des**contento con el trabajo que había realizado.

   He was unhappy with the work he had done.

8. **Ni siquiera** las vistas son buenas.

   Not even the wiews are good.

9. **No** tengo **nada** en contra del servicio del hotel.

   I have nothing against the service of the hotel.

Most applications treat negation in an ad hoc manner by processing main negation constructions, but processing negation is not as easy as using a list of negation markers and applying look-up methods because negation cues do not always act as negators. For example, in the sentence "You bought the car to use it, didn't you?" the cue "not" is not used as a negation, but it is used to reinforce the first part of the sentence. In addition, processing negation does not only consist of identifying negation cues but it is also necessary to identify the scope or part of the sentence affected by the negation, the event that is directly negated by the negation cue, and its focus, the part most prominently negated. All the components of negation are described in detail along with examples in Subsection 2.1.1: "*Definition of negation*" of Chapter 2: "*Background*".

1.4 Structure of the doctoral thesis

This doctoral thesis has been organized in eight chapters and one appendix, including this first introductory chapter in which it has been presented what motivated the development of the thesis, the objective of the same and the difficulty of solving the posed problem. The content of the successive chapters and the appendix is detailed below.

Chapter 2 introduces the two concepts that constitute the basis of this doctoral thesis, negation and sentiment analysis, and presents the state of the art for negation processing systems, the corpora annotated with negation, and sentiment analysis systems that take into account negation.

Chapter 3 shows the preliminary research, in which we carry out a study in order to check whether the detection and integration of negation in a Spanish polarity classification system can improve the accuracy of the final system. This preliminary study allows us to detect the importance of a correct processing of negation and the need to annotate a corpus with sentiment and negation. In order to determine the strengths and weaknesses of sentiment analysis systems that incorporate a module for negation processing, it is necessary a corpus annotated at both levels. In this way, an error analysis could be carried out to check whether the system correctly determines the negation cues and their corresponding scopes or if some of the errors are caused by the polarity classifier used. The approaches proposed so far could not be properly evaluated due to the non-existence of a corpus annotated with such information.

Chapter 4 presents the SFU Review$_{\mathrm{SP}}$-NEG corpus and the process followed for its annotation. In this chapter the components of negation are defined and delimited and it is proposed a typology of negation patterns in Spanish, which is applied for the annotation of the corpus. Moreover, it includes the annotation scheme used, the annotation process followed, the main sources of disagreement and the statistics and description of the corpus.

Chapter 5 includes all the details of the negation processing system developed for Spanish. It contains an exhaustive analysis of the existing corpora to select the set of data to train the system. In addition, it presents the architecture of the proposed system, the experiments carried out, the results obtained and an analysis of errors aimed at understanding the limitations of the system.

Chapter 6 corresponds to the integration of the Spanish negation processing system developed in Chapter 5 into a sentiment analysis system. It presents the methodology followed to study the effect of negation, the experiments carried out and the results obtained, as well as an error analysis. It shows the importance of the development of accurate negation processing systems for Natural Language Processing tasks.

Chapter 7 presents NEGES: Workshop on Negation in Spanish, the first initiative promoting negation research in Spanish. It contains the details of the origin of the workshop, its objective, the editions held, the tasks proposed, the data sets provided and the participants and results obtained.

Chapter 8 summarizes the conclusions of the doctoral thesis, the main contributions, the research awards and distinctions obtained and the future lines of work.

Appendix A contains the tables summarizing the corpora analysis carried out in Chapter 5.

Chapter 2

Background

The two concepts that constitute the basis of this doctoral thesis are *negation* and *sentiment analysis*. We approach negation from a computational point of view, with the objective of developing a Spanish negation processing system that can be incorporated into NLP systems to improve its accuracy. Moreover, we analyze its impact on sentiment analysis. Therefore, this chapter defines both concepts and presents the state of the art related to negation processing systems, the corpora annotated with negation, which are essential to the development of them, and systems that incorporate negation to improve sentiment analysis.

2.1 Negation

2.1.1 Definition of negation

The definition of negation according to the Spanish grammar of the Real Academia Española (RAE) says that "negation, in its multiple grammatical expressions, is considered to be a syntactic operator that is similar to quantifiers and certain adverbs, that is, it is an element that conditions the reference of the units within its scope of influence.".

> *"En sus múltiples manifestaciones gramaticales, la negación se considera un operador sintáctico en un sentido similar al de los cuantificadores y determinados adverbios, es decir, un elemento que condiciona (...) la referencia de otras unidades que se hallan en su ámbito de influencia."*
>
> (Española, 2009, p. 3631)

We can also find another relevant definitions in the literature:

> *"Negation is a grammatical category that allows the changing of the truth value of a proposition. In natural language, negation functions as an operator, like quantifiers and modals. A main characteristic of operators is that they have a scope, which means that their meaning affects other elements in the text."*
>
> (Morante & Sporleder, 2012a, p. 224, p. 229)

> *"Negation is in the first place a phenomenon of semantical opposition. As such, negation relates an expression e to another expression with a meaning that is in some way opposed to the meaning of e."*
>
> (L. R. Horn & Wansing, 2017, p. 1)

Negation can be represented in a *logical form*, using quantifiers, predicates and relations, or in a *string-level*, where its elements are defined as spans of text. Bearing in mind that this doctoral thesis focused on the analysis of negation from a computational point of view, we refer to it according to the *string-level*. From the previous definitions we can extract several ideas:

- Negation functions as an operator and, like any operator, has a form of representation, known as negation cues. Just as the sum operator is represented by the + symbol, the negation operator is represented by spans of text, like the words *no* or *never*, that are grouped under the name of negation cues.

- Negation is an operator and as any operator has a scope of influence, that is, it affects other words in the text.
- Negation can change the meaning of the words within its scope.

Therefore, negation cues are the explicit way of representing negation in a text. Depending on the negation cue used we can find different types of negation:

- **Syntactic negation**, if a syntactically independent negation cue is used to express negation (e.g. no *[no/not]*, nunca *[never]*).
- **Lexical negation**, if the cue is a word whose meaning has a negative component (e.g. negar *[deny]*, desistir *[desist]*).
- **Morphological negation**, if a morpheme is used to express negation (e.g. *i-* in ilegal *[illegal]*, *in* in incoherente *[incoherent]*). It is also known as affixal negation.

However, negation cues are not the unique elements that compose negation. In the definitions of negation, reference is made to two elements: the negation cues and the scope. Moreover, there are two elements that are part of the scope: the negated event and the focus. Therefore, the components of negation are:

- **Cues**: lexical items that modify the truth value of the propositions that are within their scope (Morante, 2010), that is, they are words that express negation. Negation cues can be adverbs (e.g., **Nunca** he estado en Los Ángeles *[I have **never** been to Los Angeles]*), pronouns (e.g., Sus decisiones **nada** tienen que ver conmigo *[His decisions have **nothing** to do with me]*), verbs (e.g., La revista desistió de publicar noticias falsas sobre la estrella *[The magazine **desisted** from published false stories about the celebrity]*) and words with negative prefixes (e.g., Lo que has hecho es **ilegal** *[What you've done is **illegal**]*). They may consist of a single token (e.g. **No** me gusta la comida de este restaurante *[I do **not** like the food of this restaurant]*), a sequence of two or more contiguous tokens (e.g. **Ni siquiera** lo ha intentado *[He has **not even** tried it]*) or two or more non-contiguous tokens (e.g. **No** voy a volver **en absoluto** *[I am **not** going back **at all**]*). The annotation of cues in corpora is very important because they are the elements that act as triggers of negation. The identification of negation cues is usually the first task that a negation processing system needs to perform, hence the importance of the annotation of corpora with this information.

- **Scope**: part of the sentence affected by the negation cue (Vincze et al., 2008), that is, all elements whose individual falsity would make the negated statement strictly true (Blanco & Moldovan, 2011b). For example, consider the sentence (a) "A mis hijos no les gusta la carne" *[My children do not like meat]* and its positive counterpart (b) "A mis hijos les gusta la carne" *[My children like meat]*. In order for (b) be true the following conditions must be satisfied: i) somebody likes something, ii) my children are the one who like and iii) meat is what it is liked. The falsity of any of them would make (a) true. Therefore, all these elements are the scope of negation: "A mis hijos no les gusta la carne" *[My children do not like meat]*. The words identified as scope are those on which the negation acts and on which it will be necessary to make certain decisions based on the objective of the final system. For example, in a sentiment analysis system, these words could see their polarity modified.

- **Negated event**: the event that is directly negated by the negation cue, usually a verb, a noun or an adjective (Kim et al., 2008). The negated event or property is always within the scope of a cue, and it is usually the head of the phrase in which the negation cue appears. For example, in the sentence "La asistencia técnica no llegó a tiempo" *[Technical assistance did not arrive on time]*, the event is the verbal form "llegó" *[arrive]*, which is the head of the sentence. There are some domains in which the identification of the negated events is crucial. For example, in the clinical domain it is relevant for the correct processing of diagnoses and for the analysis of clinical records.

- **Focus**: part of the scope that is most prominently or explicitly negated (Blanco & Moldovan, 2011a). It can also be defined as the part of the scope that is intended to be interpreted as false or whose intensity is modified. It is one of the most difficult aspects of negation to identify, especially without knowing the stress or intonation. For example, in the sentence "No voy a ir al concierto contigo" *[I'm not going to the concert with you]*, the focus is "contigo" *[with you]* because what is false is not the fact of going to the concert, but the fact of going with a specific person (*with you*). Detecting the focus of negation is useful for retrieving the numerous words that contribute to implicit positive meanings within a negation (Morante & Blanco, 2012b).

Below, Example (10) shows a sentence with the components of negation. The negation cue appears in bold, the event in italics, the focus underlined and the scope between brackets. The adverb "no" *[no]* is the negation cue because it is used to change the meaning of the words that are within its scope. The negated event is the verbal form "tiene" *[has]* and the focus is the noun "límites" *[limits]*, because it is the part that is intended to be false, it is equivalent to saying "cero límites" *[zero limits]*. The scope goes from the negation cue[1] to the end of the verb phrase, although this is not always the case, or else it would be very easy to detect the words affected by the negation. In Example (11) it is shown a sentence in which the scope of negation is the whole sentence and, in Example (12), it is presented a sentence with two coordinated structures with independent negation cues and predicates in which a scope is annotated for each coordinated negation cue.

10. Es una persona que [**no** *tiene* <u>límites</u>], aunque a veces puede controlarse.
    He is a person who has no limits, although sometimes he can control himself.

11. [El objetivo de la cámara **nunca** *ha funcionado* <u>bien</u>].
    The camera lens has never worked well.

12. [**No** *soy* <u>alta</u>] aunque [**tampoco** *soy* <u>un pitufo</u>].
    I'm not tall, but I'm not a smurf either.

An interesting overview on how modality and negation have been modeled in computational linguistics was presented by Morante & Sporleder (2012a). The authors emphasize that most

[1]There are authors that do not include the negation cue within the scope.

research in Natural Language Processing has focused on propositional aspects of meaning, but extra-propositional aspects, such as negation and modality, are also important to understanding language. They also observe a growing interest in the computational treatment of these phenomena, evidenced by several annotations projects. In this overview, modality and negation are defined in detail with some examples. Moreover, details on the linguistic resources annotated with modality and negation until then are provided as well as an overview of automated methods for dealing with these phenomena. In addition, a summary of studies in the field of sentiment analysis that have modeled negation and modality are shown. Some of the conclusions drawn by Morante and Sporleder are that although work on the treatment of negation and modality has been carried out in recent years, there is still much to do. Most research has been carried out on the English language and on specific domains and genres (biomedical, reviews, newswire, etc.). At the time of this overview only corpora annotated with negation for English had been developed, with the exception of one Swedish corpus (Dalianis & Velupillai, 2010). Therefore, the authors indicate that it would be interesting to look at different languages and also distinct domains and genres, due to the fact that extra-propositional meaning is susceptible to domain and genre effects. Another interesting conclusion drawn from this study is that it would be a good idea to study which aspects of extra-propositional meaning need to be modeled for which applications, and the appropriate modeling of modality and negation.

2.1.2 Negation processing

Negation processing is relevant for a wide range of NLP applications, such as information retrieval (Liddy et al., 2000), information extraction (Savova et al., 2010), machine translation (Baker et al., 2012) or sentiment analysis (Kennedy & Inkpen, 2006; Wiegand et al., 2010; Benamara et al., 2012; Liu, 2015).

Information retrieval systems aim to provide relevant documents from a collection, given a user query. Negation has an important role because it is not the same to make a search ("recetas con queso y leche" *[recipes with milk and cheese]*) than to make the negated version of the search ("recetas sin leche y queso" *[recipes without milk and cheese]*). The information retrieval system must return completely different documents for both queries. In other tasks, such as information extraction, negation analysis is also beneficial. Clinical texts often refer to negative findings, that is, conditions that are not present in the patient. Processing negation in these documents is crucial because the health of patients is at stake. For example, a diagnosis of a patient will be totally different if negation is not detected in the sentence “No hay signos de TVP” *[No signs of DVT]*. Translating a negative sentence from one language into another is also challenging because negation is not used in the same way. For example, the Spanish sentence *“No tiene ninguna pretensión en la vida”* is equivalent to the English sentence *“He has no pretense in life”*, but in the first case two negation cues are used while in the second only one is used. Sentiment analysis is also another task in which the presence of negation has a great impact. A sentiment analysis system that does not process negation can extract a completely different opinion than the one expressed by the opinion holder. For example, the polarity of the sentence “Una película fascinante, repetiría” *[A fascinating film, I would repeat]* should be the opposite of its negation “Una película nada fascinante, no repetiría” *[A film nothing fascinating, I would not repeat]*. Notwithstanding, negation does not always imply polarity reversal, it can also increment, reduce or have no effect on sentiment expressions, which makes the task even more difficult.

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, in order 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, in order to find the part of the scope that is most prominently negated.

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. A great deal of the research on negation, whether in and of itself or for various applications, has focused on English. However, the study of this problem in other languages than English is a necessity, since negation is a language-dependent phenomenon. Below, it is presented the state of the art related to the processing of negation in English, the reference language, and in Spanish, the language of study in this doctoral thesis.

2.1.2.1 Negation processing in English

Negation detection in English has been an active research area during recent years in the NLP community. In fact, several challenges and shared tasks have included the extraction of this language form, such as the BioNLP'09 Shared Task 3 (Kim et al., 2009), the i2b2 NLP Challenge (Uzuner et al., 2011), the *SEM 2012 Shared Task (Morante & Blanco, 2012a) and the ShARe/CLEF eHealth Evaluation Lab 2014 Task 2 (Mowery et al., 2014). Most of these challenges and workshops are related to the biomedical domain, which means that negation detection has been focused mainly on this area. However, other areas such as literature or reviews, where some corpora have recently published, have been investigated. Díaz & López (2019) provide a good overview of the most relevant works on the recognition of negation, from ruled-based systems that make use of linguistic information directly integrated into the work-flow to statistical machine learning systems that rely on textual data from which the algorithm learns generalizations on its own. They also include recent studies which are trying to explore how efficient the deep-learning algorithms are when applied to negation recognition in English. As Díaz & López (2019) state, the work of Chapman et al. (2001) stands out above all others in the biomedical domain. Their algorithm, NegEx, which is based on regular expressions, determines whether a finding or disease mentioned in narrative medical reports is present or absent. Although the algorithm has proven to be powerful in negation identification in discharge summaries, NegEx's overall precision lowers when it is applied to documents from a different domain than that for which it was conceived (Mitchell et al., 2004). In an attempt to improve NegEx's performance, other rule-based systems are developed, such as ConText (Harkema et al., 2009), DEEPEN (Mehrabi et al., 2015), and NegMiner (Elazhary, 2017).

The task of resolving the cues and scope of negation was first introduced in (Morante et al., 2008). Their machine learning system consists of two classifiers. The first decides if the tokens in a sentence are negation cues. The second determines which words in the sentence are affected by the negation. Other examples of the detection of negation cues and their scope in the biomedical domain using machine learning techniques are to be found in studies by Agarwal & Yu (2010); J. Li et al. (2010); Cruz Díaz et al. (2012); Cotik, Stricker, et al. (2016). All these systems use the

BioScope corpus (Vincze et al., 2008), a collection of clinical documents, scientific papers and abstracts annotated with negation cues and their scope, for experimentation. Recently, authors such as Qian et al. (2016), Ren et al. (2018) and Lazib et al. (2018) have investigated whether deep learning approaches are a valid alternative when they come to recognizing negation in NLP, showing that these kinds of models achieve competitive performance.

In the review domain, some works incorporate negation processing in sentiment analysis systems using rules, but do not evaluate the processing of negation (Das & Chen, 2001; Polanyi & Zaenen, 2006; Kennedy & Inkpen, 2006; Jia et al., 2009). The results show that the identification of the scope of negation improves both the accuracy of sentiment analysis and the retrieval effectiveness of opinion retrieval. In contrast to the biomedical domain, the impact of negation identification on sentiment analysis using machine learning techniques has not been sufficiently investigated. As Díaz & López (2019) point out, this is perhaps because reasonably sized standard corpora annotated with this kind of information have only recently become available. 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 cues), and a Conditional Random Field (CRF) algorithm is used to predict the scope. This classifier incorporates as features the lower-cased token string, the token PoS, the token-wise distance from explicit negation cues and dependency syntax information. 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 SFU Review corpus (Konstantinova et al., 2012), showing results in line with the results of other authors in the same task and domain. For Twitter sentiment analysis, Reitan et al. (2015) define an approach to negation scope detection which consists of a negation cue detector which uses a lexicon lookup and a CRF-based scope classifier. The system is evaluated on the Twitter

Negation corpus, a set of 4,000 tweets annotated for the task by two of the authors. Finally, Pröllochs et al. (2016) propose a novel learning strategy to detect negations in financial news. They apply reinforcement learning to develop a system that replicates the human perception of negations based on an exogenous response, such as a user rating for reviews. The results show that reinforcement learning outperforms common approaches from the related literature.

2.1.2.2 Negation processing in 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, Stricker, 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 trigger negation cues.

In the review domain, negation has also been taken into account for Spanish sentiment analysis. Until 2018,[2] existing 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., 2019) apply negation for a better classification of opinions, without assessing the processing of negation, probably due to the lack of an annotated corpus for negation in the review domain.

However, after we annotated the SFU Review$_{\mathrm{SP}}$-NEG corpus (Jiménez-Zafra et al., 2018) and organized the 2018 and 2019 editions of NEGES (Jiménez-Zafra et al., 2019a,b), the Workshop on Negation in Spanish, we find some systems for the processing of 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. It will be described on Chapter 7: "*NEGES: Workshop on Negation in Spanish*". For negation cues detection task, six systems have been developed. Two systems are presented in the 2018 edition and four in the 2019 edition. Fabregat et al. (2018) address the problem as a sequence labeling task. They use words, lemmas, PoS-tagging and case-tagging as embedded features and apply a deep learning model based on the combination of some dense neural networks and one Bidirectional

[2]The work of Jiménez-Zafra et al. (2019) was first published online on April 12, 2017.

Long Short-Term Memory network (Bi-LSTM). Loharja et al. (2018) also address the problem as a sequence labeling task, but using a Conditional Random Field (CRF) model with a set of features consisting of the part-of-speech of the words, information about how the words are written (capitalization, affixes, etc.) and neighboring words in the window [-6,1] using bigrams. Giudice (2019b) presents a model based in a convolutional Recurrent Neural Network (RNN) previously used for irony detection in Italian tweets (Giudice, 2018), but for negation it does not work well. Beltrán & González (2019) develop a CRF system based on the work of (Loharja et al., 2018), but using as features the word forms and PoS-tags of the actual word, the posterior word and the previous six words. Domínguez-Mas et al. (2019) experiments with four supervised learning approaches (CRF, Random Forest, Support Vector Machine with linear kernel and XGBoost), but the highest performance is achieved with the CRF algorithm with shallow textual, lemma, PoS-tags and dependency tree features. Finally, Fabregat et al. (2019) propose a BILSTM-based model that is an evolution of the system presented in the 2018 edition (Fabregat et al., 2018). They use words, PoS-tags, characters embedding features and a one-hot vector to represent casing information, along with a post-processing phase with some rules to correct frequent errors.

The negation processing system that we present in Chapter 5: "*A system to process negation in Spanish*" is trained also on the SFU Review$_{\mathrm{SP}}$-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. Moreover, it overcomes state-of-the-art results for negation cues detection task.

2.1.3 Corpora annotated with negation

Although this doctoral thesis focuses on the processing of negation in Spanish, we present a compilation of the corpora existing so far, as it may be useful for the scientific community to advance in the study of this phenomenon in other languages. To the best of our knowledge, there are corpora annotated for English, Spanish, Swedish, Chinese, Dutch, German and Italian.

We will start by reviewing corpora in English, the language for which most corpora annotated with negation exist. We will continue with the language object of study in this thesis, Spanish and, finally, we will show the corpora annotated in other languages.

2.1.3.1 English corpora

As indicated above, we focuses on corpora with string-level annotations. We are aware of two corpora that do not follow this annotation approach: Groningen Meaning Bank (Basile et al., 2012) and DeepBank (Flickinger et al., 2012)). The Groningen Meaning Bank[3] corpus is a collection of semantically annotated English texts with formal meaning representations rather than shallow semantics. It is composed of newswire texts from Voice of America, country descriptions from the CIA Factbook, a collection of texts from the open ANC (Ide et al., 2010) and Aesop's fables. It was automatically annotated using C&C tools and Boxer (Curran et al., 2007) and then manually corrected. The DeepBank corpus[4] contains rich syntactic and semantic annotations for the 25 Wall Street Journal sections included in the Penn Treebank (Taylor et al., 2003). The annotations are for the most part produced by manual disambiguation of parses licensed by the English Resource Grammar (Flickinger, 2000). It is available in a variety of representation formats.

To the best of our knowledge, the following are corpora that contain texts in English and string-level annotations.

BioInfer

The first corpus annotated with negation was BioInfer (Pyysalo et al., 2007). It focuses on the development of Information Extraction systems for extracting relationships between genes, proteins, and RNAs. Therefore, only entities relevant to this focus were annotated. It consists of 1,100 sentences extracted from the abstracts of biomedical research articles that were annotated with named entities and their relationships, and with syntactic dependencies including negation predicates. Out of 2,662 relationships, 163 (6%) are negated using the predicate NOT. The predicate NOT was used to annotate any explicit statements of the non-existence

[3]The Groningen Meaning Bank is available at: `http://gmb.let.rug.nl`.
[4]DeepBank is available at `http://moin.delph-in.net/DeepBank`.

of a relationship. For this purpose, the three types of negation were considered: syntactic, morphological and lexical. The scope of negation was not annotated as such, but the absence of a relationship between entities, such as *not affected by* (Example 13, Figure 2.1) or *unable to* (Example 14, Figure 2.2), was annotated with the predicate NOT.

13. Abundance of actin is not affected by calreticulin expression.

    NOT(affected by:AFFECT(abundance of actin, calreticulin expression))

```
<formula>
  <relnode entity="e.8.1" predicate="NOT">
    <relnode entity="e.8.0" predicate="AFFECT">
        <entitynode entity="e.8.8" />
        <entitynode entity="e.8.10" />
    </relnode>
  </relnode>
</formula>
```

Figure 2.1: Annotated example from the BioInfer corpus (not affected by).

14. N-WASP mutant unable to interact with profilin.

    NOT(interact with:BIND(N-WASP mutant, profilin))

```
<formula>
  <relnode entity="e.749.4" predicate="NOT">
    <relnode entity="e.749.1" predicate="BIND">
        <entitynode entity="e.749.0" />
        <entitynode entity="e.749.3" />
    </relnode>
  </relnode>
</formula>
```

Figure 2.2: Annotated example from the BioInfer corpus (unable to).

In relation to the annotation process, this was divided into two parts. On the one hand, the dependency annotations were created by six annotators who worked in rotating pairs to reduce variation and avoid systematic errors. Two of the annotators were biology experts and the other four had the possibility of consulting with an expert. On the other hand, the entity and relationship annotations were created based on a previously unpublished annotation of the corpus and were carried out by a biology expert, with difficult cases and annotation rules being discussed with two Information Extraction researchers. The inter-annotator agreement was not measured in this corpus because the authors considered that there were some difficulties in calculating the kappa statistic for many of the annotation types. They said that they intended to measure agreement separately for the different annotation types, applying the most informative measures for each type but, to the best of our knowledge, this information was not published. The annotation manual used for producing the annotation can be found at http://tucs.fi/publications/view/?pub_id=tGiPyBjHeSa07a.

The BioInfer corpus is in XML format, licensed under a Creative Commons Attribution-ShareAlike 3.0 Unported License and can be downloaded at http://mars.cs.utu.fi/BioInfer/.

GENIA Event

The GENIA Event corpus (Kim et al., 2008) is composed of 9,372 sentences from Medline abstracts that were annotated with biological events and with negation and uncertainty. It is an extension of the GENIA corpus (Ohta et al., 2002; Kim et al., 2003), which was annotated with the Part Of Speech (POS), syntactic trees and terms (biological entities).

As for negation, it was annotated whether events were explicitly negated or not (Example 15, Figure 2.3), using the label *non-exists* or *exists*, respectively. The three types of negation were considered, but linguistic cues were not annotated. Out of a total of 36,858 tagged events, 2,351 events were annotated as explicitly negated. The annotation process was carried out by a biologist and three graduate students in molecular biology following the annotation guidelines defined[5]. However, there is no information about inter-annotator agreement.

The corpus is provided as a set of XML files, and it can be downloaded at http://www.geniaproject.org/genia-corpus/event-corpus under the terms of the Creative Commons Public License.

[5]http://www.nactem.ac.uk/meta-knowledge/Annotation_Guidelines.pdf

15. This pathway involves the Rac1 and Cdc42 GTPases, two enzymes which are not required for NF-kappaB activation by IL-1beta in epithelial cells.

```
<event assertion="non-exist" id="E40">
    <type class="Positive\_regulation" />
    <theme idref="E39" />
    <cause idref1="T56" idref="T54" />
    <clue>
        This pathway involves the Rac1 and Cdc42 GTPases,
        two enzymes which are not
        <clueType>required</clueType>
        <linkTheme>for</linkTheme>
        NF-kappaB activation by IL-1beta
        <clueLoc>in epithelial cells</clueLoc>.
    </clue>
</event>
```

Figure 2.3: Annotated example from the Genia Event corpus.

BioScope

The BioScope corpus (Vincze et al., 2008) is one of the largest corpus and is the first in which negation and speculation cues have been annotated with their scopes. It contains 6,383 sentences from clinical free-texts (radiology reports), 11,871 sentences from full biological papers and 2,670 sentences from biological paper abstracts from the GENIA corpus (Ohta et al., 2002; Kim et al., 2003). In total, it has 20,924 sentences, out of which 2,720 contains negations.

Negation is understood here as the implication of the non-existence of something. The strategy for annotating keywords was to mark the minimal unit possible (only lexical and syntactic negations were considered), and the largest syntactic unit for scopes. Moreover, negation cues were also included within the scope (Example 16, Figure 2.4).

The corpus was annotated by two independent linguist annotators and a chief linguist following annotation guidelines.[6] The consistency level of the annotation was measured using the inter-annotator agreement rate defined as the $F_\beta - 1$ measure of one annotation considering the second one as the gold standard. The average agreement of negation keywords annotation was 93.69, 93.74 and 85.97 for clinical records, abstracts and full articles respectively and the average agreement of scope identification for the three corpora was 83.65, 94.98 and 78.47 respectively.

The BioScope corpus is in XML format and is freely available for academic purposes at http://rgai.inf.u-szeged.hu/index.php?lang=en&page=bioscope. This corpus was also employed in the CoNLL-2010 Shared Task: *Learning to detect hedges and their scope in natural language text* (Farkas et al., 2010).

16. PMA treatment, and not retinoic acid treatment of the U937 cells acts in inducing NF-KB expression in the nuclei.

```
<sentence id="S1.4">
  PMA treatment, and
    <xcope id="X1.4.1">
        <cue type="negation" ref="X1.4.1">not</cue>
        retinoic acid treatment of the U937 cells
    </xcope>
  acts in inducing NF-KB expression in the nuclei.
</sentence>
```

Figure 2.4: Annotated example from the BioScope corpus.

Product Review corpus

In 2010, the Product Review corpus was presented (Councill et al., 2010). It is composed of 2,111 sentences from 268 product reviews extracted from Google Product Search. This corpus was annotated with the scope of syntactic negation cues and 679 sentences were found to contain negation. Each review was manually annotated with the scope of negation by a single person, after achieving inter-annotator agreement of 91% with a second person on a smaller subset

[6] The annotation guidelines can be downloaded at http://rgai.inf.u-szeged.hu/project/nlp/bioscope/Annotation%20guidelines2.1.pdf and a discussion of them can be found in Vincze (2010)

of 20 reviews containing negation. Inter-annotator agreement was calculated using a strict exact span criteria where both the existence and the left/right boundaries of a negation span were required to match. In this case, negation cues were not included within the scope. The guidelines used for the annotation are described in the work in which the corpus was presented.

The format of the corpus is not mentioned by the authors and is not publicly available. However, we contacted the authors and they sent us the corpus. In this way we were able to see that it is in XML format and extract an example of it (Example 17, Figure 2.5).

17. I am a soft seller, If you don't want or need the services offered that's cool with me.

```
<sentence>
    I am a soft seller, If you don't
    <negation_span>
        want or need the services offered
    </negation_span>
    that's cool with me.
</sentence>
```

Figure 2.5: Annotated example from the Product Review corpus.

PropBank Focus (PB-FOC)

In 2011, the PropBank Focus (PB-FOC) corpus was presented. It introduced a new element for the annotation of negation, the focus. Blanco & Moldovan (2011a) selected 3,993 verbal negations contained in 3,779 sentences from the WSJ section of the Penn TreeBank marked with MNEG in the PropBank corpus (Palmer et al., 2005), and performed annotations of negation focus. They reduced the task to selecting the semantic role most likely to be the focus.

50% of the instances were annotated twice by two graduate students in computational linguistics and an inter-annotator agreement of 72% percent was obtained (it was calculated as the percentage of annotations that were a perfect match). Later, disagreements were examined and resolved by giving annotators clearer instructions. Finally, the remaining instances were annotated once. The annotation guidelines defined are described in the paper in which the corpus was presented.

This corpus was used in Task 2, focus detection, at the *SEM 2012 Shared Task (Resolving the scope and focus of negation) (Morante & Blanco, 2012b). It is in CoNLL format (Farkas et al., 2010) and can be downloaded at `http://www.clips.ua.ac.be/sem2012-st-neg/data.html`. Figure 2.6 shows the annotations for Example (18). The columns provide the following information: token (1), token number (2), POS tag (3), named entities (4), chunk (5), parse tree (6), syntactic head (7), dependency relation (8), semantic roles (9 to previous to last, with one column per verb), negated predicates (previous to last), focus (last).

PB-FOC is distributed as standalone annotations on top of the Penn TreeBank. The distribution must be completed with the actual words from the the Penn TreeBank, which is subject to an LDC license.

18. Marketers believe most Americans won't make the convenience trade-off.

Marketers	1	NNS	O	B-NP	(S1(S(NP*)	2	nsubj	(A0*)	*	-	*
believe	2	VBP	O	B-VP	(VP*	0	root	(V*)	*	-	*
most	3	RBS	O	B-NP	(SBAR(S(NP*	4	amod	(A1*	(A0*	-	FOCUS
Americans	4	NNPS	O	I-NP	*)	7	nsubj	*	*)	-	FOCUS
wo	5	MD	O	B-VP	(VP*	7	aux	*	(AM-MOD*)	-	*
n't	6	RB	O	I-VP	*	7	neg	*	(AM-NEG*)	-	*
make	7	VB	O	I-VP	(VP*	2	ccomp	*	(V*)	N	*
the	8	DT	O	B-NP	(NP*	10	det	*	(A1*	-	*
convenience	9	NN	O	I-NP	*	10	nn	*	*	-	*
trade-off	10	NN	O	I-NP	*)))))	7	dobj	*)	*)	-	*
...	11	:	O	O	*	2	punct	*	*	-	*
.	12	.	O	O	*))	2	punct	*	*	-	*

Figure 2.6: Annotated example from the PropBank Focus (PB-FOC) corpus.

ConanDoyle-neg

The ConanDoyle-neg (Morante & Daelemans, 2012) is a corpus of Conan Doyle stories annotated with negation cues and their scopes, as well as the event or property that is negated. It is composed of 3,640 sentences from *The Hound of the Baskervilles* story, out of which 850 contain negations and, 783 sentences from *The Adventure of Wisteria Lodge* story out of which 145 contain negations. In this case, the three types of negation cues (lexical, syntactic and morphological) were taken into account.

The corpus was annotated by two annotators, a master's student and a researcher, both with a background in linguistics. The inter-annotator agreement in terms of F1 was of 94.88% and 92.77% for negation cues in *The Hound of the Baskervilles* story and *The Adventure of Wisteria Lodge* story, respectively, and of 85.04% and 77.31% for scopes. The annotation guidelines[7] are based on those of the BioScope corpus, but there are some differences. The most important differences are that in the ConanDoyle-neg corpus the cue is not considered to be part of the scope, the scope can be discontinuous and all the arguments of the event being negated are considered to be within the scope, including the subject, which is kept out of the scope in the BioScope corpus.

19. After his habit he said nothing, and after mine I asked no questions.

WL2	108	0	After	After	IN	(S(S(PP*	_	After	_	_	_	_
WL2	108	1	his	his	PRP$	(NP*	_	his	_	_	_	_
WL2	108	2	habit	habit	NN	*))	_	habit	_	_	_	_
WL2	108	3	he	he	PRP	(NP*)	_	he	_	_	_	_
WL2	108	4	said	say	VBD	(VP*	_	said	said	_	_	_
WL2	108	5	nothing	nothing	NN	(NP*)))	nothing	_	_	_	_	_
WL2	108	6	,	,	,	*	_	_	_	_	_	_
WL2	108	7	and	and	CC	*	_	_	_	_	_	_
WL2	108	8	after	after	IN	(S(PP*	_	_	_	_	after	_
WL2	108	9	mine	mine	NN	(NP*))	_	_	_	_	mine	_
WL2	108	10	I	I	PRP	(NP*)	_	_	_	_	I	_
WL2	108	11	asked	ask	VBD	(VP*	_	_	_	_	asked	asked
WL2	108	12	no	no	DT	(NP*	_	_	_	no	_	_
WL2	108	13	questions	question	NNS	*)))	_	_	_	_	questions	_
WL2	108	14	.	.	.	*)	_	_	_	_	_	_

Figure 2.7: Annotated example from the ConanDoyle-neg corpus.

The ConanDoyle-neg corpus was prepared with the aim of using it at the *SEM 2012 Shared Task[8] (Morante & Blanco, 2012b), which was dedicated to resolving the scope and focus of negation. It is in CoNLL format (Farkas et al., 2010) and can be downloaded at `http://www.clips.ua.ac.be/sem2012-st-neg/data.html`. In Figure 2.7 it can be seen how Example (19) is represented in the corpus. The content of the columns is as follows: chapter name (1), sentence number within chapter (2), token number within sentence (3), token (4), lemma (5), POS tag (6), parse tree information (7). If the sentence has no negations, column 8 has a "***" value and there are no more columns, but if the sentence has negations, the annotation for each

[7]The annotation guidelines are described in Morante et al. (2011)

[8]`www.clips.ua.ac.be/sem2012-st-neg/`

negation is provided in three columns. The first column contains the word that belongs to the negation cue, the second the word that belongs to the scope of the negation cue and the third the word that is the negated event or property.

No license is needed to download the corpus.

SFU Review$_{EN}$

Konstantinova et al. (2012) annotated the SFU Review$_{EN}$ corpus (Taboada et al., 2006) with information about negation and speculation. This corpus is composed of 400 reviews extracted from the website *Epinions.com* that belong to 8 different domains: books, cars, computers, cookware, hotels, films, music and phones. It was annotated with negation and speculation markers and their scopes. Out of the total amount of 17,263 sentences, 18% contain negation cues (3,017 sentences). In this corpus syntactic negation was annotated, but not lexical nor morphological negation. Below, Figure 2.8 shows how Example (20) is annotated in the corpus.

The annotation process was carried out by two linguists. The entire corpus was annotated by one of them and 10% of the documents (randomly selected in a stratified way) were annotated by the second one in order to measure inter-annotator agreement. The kappa agreement was of 0.927 for negation cues and 0.872 for the scope. The guidelines of the BioScope corpus were taken into consideration with some modifications. The min-max strategy of BioScope corpus was used but negation cues were not included within the scope. A complete description of the annotation guidelines can be found in Konstantinova et al. (2011).

This corpus is in XML format and publicly available at https://www.sfu.ca/~mtaboada/SFU_Review_Corpus.html, under the terms of the GNU General Public License as published by the Free Software Foundation.

20. I have never liked the much taller instrument panel found in BMWs and Audis.

```xml
<SENTENCE>
  <W>I</W>
  <W>have</W>
  <C>
    <cue ID="15" type="negation">
        <W>never</W>
    </cue>
  </C>
  <xcope ID="17">
    <ref ID="19" SRC="15"/>
    <W>liked</W>
    <W>the</W>
    <W>much</W>
    <W>taller</W>
    <W>instrument</W>
    <W>panel</W>
    <W>found</W>
    <W>in</W>
    <W>BMWs</W>
    <C>
        <W>and</W>
    </C>
    <W>Audis</W>
  </xcope>
  <W>.</W>
</SENTENCE>
```

Figure 2.8: Annotated example from the SFU Review$_{EN}$ corpus.

NEG-DrugDDI

In the biomedical domain, the DrugDDI 2011 corpus (Segura Bedmar et al., 2011) was also tagged with negation cues and their scopes, producing the NEG-DrugDDI corpus (Bokharaeian et al., 2013). It contains 579 documents extracted from the DrugBank database and it is composed of 5,806 sentences, out of which 1,399 sentences (24%) contain negation.

This corpus was automatically annotated with a subsequent manual revision. The first annotation was performed using a rule-based system (Ballesteros et al., 2012), which is publicly available and works on biomedical literature following the BioScope guidelines to annotate sentences with negation. After applying the system, a set of 1,340 sentences were annotated with negation. Then, the outcome was manually checked correcting annotations when needed. In order to do so, the annotated corpus was divided into 3 different sets that were assigned to 3 different evaluators. The evaluators checked all the sentences contained in each set and corrected the annotation errors. After this revision, a different evaluator revised all the annotations produced by the other 3 evaluators. Next, sentences were explored in order to annotate some negation cues that were not detected by the system, such as *unaffected*, *unchanged* or *non-significant*. Finally, 1,399 sentences of the corpus were annotated with the scope of negation.

21. Repeating the study with 6 healthy male volunteers in the absence of glibenclamide did not detect an effect of acitretin on glucose tolerance.

```
<sentence origId="s2" id="DrugDDI.d393.s2" text="Repeating
the study with 6 healthy male volunteers in the absence of
glibenclamide did not detect an effect of acitretin on
glucose tolerance.">
  <entity origId="s2.p31" id="DrugDDI.d393.s2.e0"
  text="glibenclamide" type="drug" charOffset="69-82"/>
  <entity origId="s2.p36" id="DrugDDI.d393.s2.e1"
  text="acitretin" type="drug" charOffset="111-120"/>
  <pair id="DrugDDI.d393.s2.p0" interaction="false"
  e2="DrugDDI.d393.s2.e1" e1="DrugDDI.d393.s2.e0"/>
  <negationtags>Repeating the study with 6 healthy male
  volunteers in the <xcope><cue>absence</cue> of
  glibenclamide </xcope>did <xcope><cue>not</cue> detect
  an effect of acitretin on glucose tolerance</xcope>.
  </negationtags>
</sentence>
```

Figure 2.9: Annotated example from the NEG-Drug DDI corpus.

The NEG-DrugDDI corpus is in XML format and can be downloaded at http://nil.fdi.ucm.es/sites/default/files/NegDrugDDI.zip. Figure 2.9 shows a corpus sentence containing two negations (Example 21).

NegDDI-DrugBank

A new corpus which included the DrugDDI 2011 corpus as well as Medline abstracts was developed and it was named the DDI-DrugBank 2013 corpus (Herrero Zazo et al., 2013). This corpus was also annotated with negation markers and their scopes and it is known as the NegDDI-DrugBank corpus (Bokharaeian et al., 2014). It consists of 6,648 sentences from 730 files and it has 1,448 sentences with at least one negation scope, which corresponds to 21.78% of the sentences. The same approach as the one used for the annotation of the NEG-DrugDDI corpus was followed.

22. Drug-Drug Interactions: The pharmacokinetic and pharmacodynamic interactions between UROXATRAL and other alpha-blockers have not been determined.


```
<sentence id="DDI-DrugBank.d273.s0" text="Drug-Drug
Interactions: The pharmacokinetic and pharmacodynamic
interactions between UROXATRAL and other alpha-blockers
have not been determined.">
  <entity id="DDI-DrugBank.d273.s0.e0" text="UROXATRAL"
  type="brand" charOffset="85-93"/>
  <entity id="DDI-DrugBank.d273.s0.e1" text="alpha-blockers"
  type="group" charOffset="105-118"/>
  <pair id="DDI-DrugBank.d273.s0.p0"
  e2="DDI-DrugBank.d273.s0.e1" e1="DDI-DrugBank.d273.s0.e0"
  ddi="false"/>
  <negationtags><xcope>Drug-Drug Interactions: The
  pharmacokinetic and pharmacodynamic interactions
  between UROXATRAL and other alpha-blockers have
  <cue>not</cue> been determined</xcope>.</negationtags>
</sentence>
```


Figure 2.10: Annotated example from the NEGDDI-DrugBank corpus.

This corpus is in XML format and is freely available at http://nil.fdi.ucm.es/sites/default/files/NegDDI_DrugBank.zip. Below, Figure 2.10 show the annotations from Example (22). It can be seen that the annotation scheme is the same as the one used in the corpus NEG-DrugDDI.

Deep Tutor Negation

The Deep Tutor Negation corpus (DT-Neg) (Banjade & Rus, 2016) consists of texts extracted from tutorial dialogues where students interacted with an Intelligent Tutoring System to solve conceptual physics problems. It contains annotations about negation cues, and the scope and focus of negation. From a total of 27,785 student responses, 2,603 responses (9.36%) contain at least one explicit negation cue. In this corpus, syntactic and lexical negation were taken into account but not morphological negation. Figure 2.11 presents how the response of Example (23) is annotated in the corpus.

23. They will not hit the water at the same time.

```
ID: APR2639A
METAINFO: SpeechAct:
Contribution Corpus: April2013CollegeStudents
AnswerId: 2639 Strand: VM_LV02_PR00.FCI-38.vMHK
QUESTION: If initial velocity and the rate of change in velocity,
which the acceleration, are the same vertically what can you say
about the time it takes for the two girls to travel the same
distance vertically?
ANSWER: They will not hit the water at the same time.
CUE: not
ANNOTATEDANSWER: [They will] <<not>>
[hit the water {at the same time}] .
TAG: 0
WATCH: 0
```

Figure 2.11: Annotated example from the Deep Tutor Negation corpus.

In relation to the annotation process, the corpus was first automatically annotated based on a list of cue words which the authors compiled from different research reports (Morante et al., 2011; Vincze et al., 2008). After this, annotators validated the automatically detected negation cues and annotated the corresponding negation scope and focus. The annotation was carried out by a total of 5 graduate students and researchers following an annotation manual that was inspired by the guidelines of (Morante et al., 2011). In order to measure inter-annotator agreement, a subset of 500 instances was randomly selected. It was equally divided into five subsets and each of them was annotated by two annotators. The averaged agreement for scope and focus detection was 89.43% and 94.20%, respectively (the agreement for negation cue detection was not reported).

This corpus is in TXT format and it is available for research-only, non-commercial, and internal use at http://deeptutor.memphis.edu/resources.htm.

SOCC

Finally, the last English corpus we are aware of is the SFU Opinion and Comments Corpus (SOCC) (Kolhatkar et al., 2018) that was presented at the beginning of 2018. The original corpus contains 10,339 opinion articles (editorials, columns, and op-eds) together with their 663,173 comments from 303,665 comment threads, from the main Canadian daily newspaper in English, *The Globe and Mail*, for a five-year period (from January 2012 to December 2016). The corpus is organized into three subcorpora: the articles corpus, the comments corpus, and the comment-threads corpus. The corpus description and download links are publicly available.[9]

SOCC was recollected to study different aspects of on-line comments such as the connections between articles and comments; the connections of comments to each other; the types of topics discussed in comments; the nice (constructive) or mean (toxic) ways in which commenters respond to each other; and how language is used to convey very specific types of evaluation. However, the main focus of the annotation is oriented towards the study of the constructiveness and evaluation in the comments. Thus, a subset of SOCC with 1,043 comments was selected to be annotated with three different layers: constructiveness, appraisal and negation.

[9]https://github.com/sfu-discourse-lab/SOCC

The primary intention of the research and annotation was to examine the relationship between negation, negativity, and appraisal. In the annotation process up to two individuals participated. Specific guidelines were developed to assist the annotators throughout the annotation process, and to ensure that annotations were standardized. These guidelines are publicly available through the GitHub page for the corpus.[10] The 1,043 comments were annotated for negation using Webanno (de Castilho et al., 2016) and the elements to consider were the negation cue or keyword, focus and scope. Syntactic negation was taken into account, as well as some verbs and adjectives that indicate negation. The negation cue is excluded from the scope. In cases of elision or question and response, a special annotation label, *xscope*, was created to indicate the implied content of a non explicit scope. For the 1,043 comments there were 1,397 negation cues, 1,349 instances of scope, 34 instances of xscope, and 1,480 instances of focus. Next, Figure 2.12 shows how Example (24) is annotated in the corpus:

24. Because if nobody is suggesting that then this is just another murder where someone was at the WRONG PLACE at the WRONG TIME.

```
2-1 186-193 Because _
2-2 194-196 if _
2-3 197-203 nobody NEG
2-4 204-206 is SCOPE[2]
2-5 207-217 suggesting SCOPE[2]
2-6 218-222 that SCOPE[2]|FOCUS[3]
2-7 223-227 then _
...
2-20 293-295 at _
2-21 296-299 the _
2-22 300-305 WRONG _
2-23 306-310 TIME _
2-24 310-311 . _
```

Figure 2.12: Annotated example from the SOCC corpus.

Regarding the agreement, two annotators performed the annotation, a graduate student in

[10] https://github.com/sfu-discourse-lab/SOCC/tree/master/guidelines

computer science and an expert in computational linguistics. The expert was in charge of overseeing the process and training the research assistant. The research assistant annotated the entire corpus. The senior annotator then refined and resolved any disagreements. To calculate agreement, 50 comments from the beginning of the annotation process and 50 comments from the conclusion of the annotation process were compared. Agreement between the annotators was calculated individually based on the label and the span for the keyword, scope, and focus. Agreement was calculated using percentage agreement for nominal data, with annotations regarded as either agreeing or disagreeing. A percentage indicating agreement was measured for both label and span, then combined to yield an average agreement for the tag. The agreement for the first 50 comments was 99.0% for keyword, 98.0% for scope and 85.3% for focus. For the last 50 comments the agreement was 96.4% for keyword, 94.2% for scope and 75.8% for focus.

The annotated corpus is in TSV format and it can be downloaded at https://researchdata.sfu.ca/islandora/object/islandora%3A9109 under a Creative Commons Attribution - Non-Commercial - ShareAlike 4.0 International License.

2.1.3.2 Spanish corpora

Here we present the Spanish corpora annotated with negation. To the best of our knowledge, five corpora exist from different domains, although the clinical domain is the predominant one.

UAM Spanish Treebank

The first Spanish corpus annotated with negation that we are aware of is the UAM Spanish Treebank (Moreno et al., 2003), which was enriched with the annotation of negation cues and their scopes (Sandoval & Salazar, 2013).

The initial UAM Spanish Treebank consisted of 1,500 sentences extracted from newspaper articles (*El País Digital and Compra Maestra*) that were annotated syntactically. Trees were encoded in a nested structure, including syntactic category, syntactic and semantic features, and constituent nodes, following the Penn Treebank model. Later, this version of the corpus was extended with the annotation of negation and 10.67% of the sentences were found to contain negations (160 sentences).

25. No juega a ser un magnate.

    He doesn't play at being a tycoon.

```
<Sentence Neg="YES" Id="138">
  <NP Function="SUBJ" Id="1" Gender="SG" P="3" Elided="Yes"/>
  <VP Tense="Tensed" Verbal_temp="PRES" Mode="IND" Number="SG" P="3">
  <ADVP Type="NEG"> <ADV Lemma="no" Type="NEG"> No </ADV> </ADVP>
  <V Lemma="jugar" Tensed="Yes" Form="PRES" Mode="IND" ...> juega </V>
  <PP Type="A" Class="OBL">
    <PREP Lemma="a"> a </PREP>
      <CL Function="INFINITIVE">
        <NP Function="SUBJ" Ref="1" Elided="Yes"/>
        <VP Tense="Untensed" Verbal_temp="INFINITE">
          <V Verbal_temp="ser" Lemma="ser" Tensed="No" ...> ser </V>
          <NP Function="ATTR" Gender="MASC" Number="SG">
            <ART Lemma="un" Type="INDEF" Gender="MASC" Number="SG"> un </ART>
            <N Lemma="magnate" Type="Common" Gender="MASC" ...> magnate </N>
          </NP>
        </VP>
      </CL>
  </PP>
  </VP>
  <PUNCT Type="PERIOD"/>
</Sentence>
```

Figure 2.13: Annotated example from the UAM Spanish Treebank corpus.

In this corpus, syntactic negation was annotated but not lexical nor morphological negation. It was annotated by two experts in corpus linguistics who followed similar guidelines to those of the Bioscope corpus (Szarvas et al., 2008; Vincze, 2010). They included negation cues within the scope as in Bioscope and NegDDI-DrugBank (Bokharaeian et al., 2014). All the arguments of the negated events were also included in the scope of negation, including the subject (as in ConanDoyle-neg corpus (Morante & Daelemans, 2012)), which was excluded from the scope in active sentences in Bioscope. There is no information about inter-annotator agreement.

The UAM Spanish Treebank corpus is freely available for research purposes at `http://www.lllf.uam.es/ESP/Treebank.html`, but it is necessary to accept the license agreement for non-commercial use and send it to the authors. It is in XML format, negation cues are tagged with the label *Type="NEG"* and the scope of negation is tagged with the label *Neg="YES"* in the syntactic constituent on which negation acts. If negation affects the complete sentence, the label is included as an attribute of the tag <*Sentence*> or, by contrast, if negation only affects part of the sentence, for example, an adjectival syntagma represented as <*Adjp*>, the label *Neg="YES"* is included in the corresponding tag. Figure 2.13 presents an example extracted from the corpus (Example 25) in which negation affects the complete sentence.

IxaMed-GS

The IxaMed-GS corpus (Oronoz et al., 2015) is composed of 75 real electronic health records from the outpatient consultations of the Galdakao-Usansolo Hospital in Biscay (Spain). It was annotated by two experts in pharmacology and pharmacovigilance with entities related to diseases and drugs, and with the relationships between entities indicating adverse drug reaction events. They defined their own annotation guidelines taking into consideration the issues that should be considered for the design of a corpus according to Ananiadou & McNaught (2006).

The objective of this corpus was not the annotation of negation but the identification of entities and events in clinical reports. However, negation and speculation were taken into account in the annotation process. In the corpus, four entity types were annotated: diseases, allergies, drugs and procedures. For diseases and allergies, they distinguished between negated entity, speculated entity and entity (for non-speculative and non-negated entities). On the one hand, 2,362 diseases were annotated, out of which 490 (20.75%) were tagged as negated diseases and 40 (1.69%) as speculated diseases. On the other hand, 404 allergy entities were identified, from which 273 (67.57%) were negated allergies and 13 (3.22%) speculated allergies. The quality of the annotation process was assessed by measuring the inter-annotator agreement, which was 90.53% for entities and 82.86% for events.

The corpus might be possible to acquire via the EXTRECM project[11] following a procedure of some conditions that include a confidentiality agreement, and its format it is not specified.

SFU Review$_{SP}$-NEG

This corpus is one of the results of this doctoral thesis. Although Chapter 4: "*SFU Review$_{SP}$-NEG corpus: a Spanish corpus annotated with negation*" presents all its details, here we describe it briefly to situate it in time and to show a complete compilation of the existing Spanish corpora at the time of the presentation of this thesis.

The SFU Review$_{SP}$-NEG[12] (Jiménez-Zafra et al., 2018) is the first Spanish corpus that includes the event in the annotation of negation and that takes into account discontinuous negation markers. Moreover, it is the first corpus in which it is annotated how negation affects the words that are within its scope, that is, whether there is a change in the polarity or an increment or reduction of its value. It is an extension of the Spanish part of the SFU Review corpus (Taboada et al., 2006) and it could be considered the counterpart of the SFU Review Corpus with negation and speculation annotations[13] (Konstantinova et al., 2012).

The Spanish SFU Review corpus consists of 400 reviews extracted from the website *Ciao.es* that belong to 8 different domains: cars, hotels, washing machines, books, cell phones, music, computers, and movies. For each domain there are 50 positive and 50 negative reviews, defined as positive or negative based on the number of stars given by the reviewer (1-2=negative; 4-5=positive; 3-star reviews were not included). Later, it was extended to the SFU Review$_{SP}$-NEG corpus in which each review was automatically annotated at the token level with pos-tags and lemmas using Freeling (Padró & Stanilovsky, 2012), and manually annotated at the sentence level with negation cues and their corresponding scopes and events. It is composed of 9,455 sentences, out of which 3,022 sentences (31.97%) contain at least one negation marker.

26. El 307 es muy bonito, pero no os lo recomiendo.

    The 307 is very nice, but I don't recommend it.

[11]http://ixa.si.ehu.eus/extrecm
[12]First Online: 22 May 2017 https://doi.org/10.1007/s10579-017-9391-x
[13]https://www.sfu.ca/~mtaboada/SFU_Review_Corpus.html

```xml
<sentence complex="no">
  <d wd="El" postype="article" pos="da0ms0" name="d" lem="el" .../>
  <z wd="307" pos="z" name="z" lem="307"/>
  <v wd="es" postype="semiauxiliary" pos="vsip3s0" name="v"
  lem="ser" person="3" num="s" tense="present"
  mood="indicative"/>
  <r wd="muy" pos="rg" name="r" lem="muy"/>
  <a wd="bonito" postype="qualificative" pos="aq0ms0" name="a"
  lem="bonito" num="s" gen="m"/>
  <f wd="," pos="fc" name="f" lem="," punct="comma"/>
  <c wd="pero" postype="coordinating" pos="cc" ... lem="pero"/>
  <neg_structure polarity="negative" value="neg" change="yes">
    <scope>
      <negexp>
        <r wd="no" postype="negative" pos="rn" name="r" lem="no"/>
      </negexp>
      <p wd="os" postype="personal" pos="pp2cp000" name="p"
      lem="os" person="2" num="p" gen="c"/>
      <p wd="lo" postype="personal" pos="pp3cna00" name="p"
      lem="lo" person="3" num="n" gen="c" case="accusative"/>
      <event>
        <v wd="recomiendo" postype="main" pos="vmip1s0" name="v"
        lem="recomendar" person="1" num="s" tense="present" .../>
      </event>
    </scope>
  </neg_structure>
  <f wd="." pos="fp" name="f" lem="." punct="period"/>
</sentence>
```

Figure 2.14: Annotated example from the SFU Review$_{SP}$-NEG corpus.

In this corpus, syntactic negation was annotated but not lexical nor morphological negation, as in the UAM Spanish Treebank corpus. Unlike this one, annotations on the event and on how negation affects the polarity of the words within its scope were included. It was annotated by two senior researchers with in-depth experience in corpus annotation who supervised the whole process and two trained annotators who carried out the annotation task. The Kappa coefficient for inter-annotator agreement was 0.97 for negation cues, 0.95 for negated events and 0.94 for scopes.[14] A detailed discussion of the main sources of disagreements can be found in (Jiménez-Zafra et al., 2016).

The guidelines of the Bioscope corpus were taken into account, but after a thorough analysis of negation in Spanish, a typology of negation patterns in Spanish (Martí et al., 2016) was defined. As in Bioscope, NegDDI-DrugBank and UAM Spanish Treebank, negation markers were included within the scope. Moreover, the subject was also included within the scope when the word directly affected by negation is the verb of the sentence. The event was also included within the scope of negation as in the ConanDoyle-neg corpus.

The SFU Review$_{SP}$-NEG is in XML format. It is publicly available and can be downloaded at `http://sinai.ujaen.es/sfu-review-sp-neg-2/` under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. Figure 2.14 present how Example (26), a sentence containing negation, is annotated in this corpus.

The annotations of this corpus were used in *NEGES 2018: Workshop on Negation in Spanish* (Jiménez-Zafra et al., 2019a) for Task 2: "Negation cues detection" (Jiménez-Zafra et al., 2018b). The corpus was converted to CoNLL format (Farkas et al., 2010) as in the *SEM 2012 Shared Task (Morante & Blanco, 2012b). This format of the corpus can be downloaded from the web of the workshop `http://www.sepln.org/workshops/neges/index.php?lang=en` or by sending an email to the organizers. In Figure 2.15, we show an example of a sentence with two negations. In this version of the corpus, each line corresponds to a token, each annotation is provided in a column and empty lines indicate the end of the sentence. The content of the given columns is: domain_filename (1), sentence number within domain_filename (2), token number within sentence (3), word (4), lemma (5), part-of-speech (6), part-of-speech type (7); if the sentence has no negations, column 8 has a "***" value and there are no more columns. If the sentence has negations, the annotation for each negation is provided in three columns. The first column contains the word that belongs to the negation cue. The second and third

[14]The inter-annotator agreement values have been corrected with respect to those published in (Jiménez-Zafra et al., 2018) due to the detection of an error in the calculation thereof.

columns contain "-", because the proposed task was only negation cue detection. Figure 2.15 shows an annotated example.

27. Aquí estoy esperando que me carguen los puntos en mi tarjeta más, no sé dónde tienen la cabeza pero no la tienen donde deberían.
Here I am waiting for the points to be loaded on my card and I don't know where they have their head but they don't have it where they should.

hoteles	1	Aun	aun	np00000	proper	-	-	-	-	-	-
hoteles	2	estoy	estar	vaip1s0	auxiliary	-	-	-	-	-	-
hoteles	3	esperando	esperar	vmg0000	main	-	-	-	-	-	-
hoteles	4	que	que	cs	subordinating	-	-	-	-	-	-
hoteles	5	me	me	pp1cs000	personal	-	-	-	-	-	-
hoteles	6	carguen	cargar	vmsp3p0	main	-	-	-	-	-	-
hoteles	7	los	el	da0mp0	article	-	-	-	-	-	-
hoteles	8	puntos	punto	ncmp000	common	-	-	-	-	-	-
hoteles	9	en	en	sps00	preposition	-	-	-	-	-	-
hoteles	10	mi	mi	dp1css	possessive	-	-	-	-	-	-
hoteles	11	tarjeta	tarjeta	ncfs000	common	-	-	-	-	-	-
hoteles	12	más	más	rg	-	-	-	-	-	-	-
hoteles	13	,	,	fc	-	-	-	-	-	-	-
hoteles	14	no	no	rn	negative	no	-	-	-	-	-
hoteles	15	sé	saber	vmip1s0	main	-	-	-	-	-	-
hoteles	16	dónde	dónde	pt000000	interrogative	-	-	-	-	-	-
hoteles	17	tienen	tener	vmip3p0	main	-	-	-	-	-	-
hoteles	18	la	el	da0fs0	article	-	-	-	-	-	-
hoteles	19	cabeza	cabeza	ncfs000	common	-	-	-	-	-	-
hoteles	20	pero	pero	cc	coordinating	-	-	-	-	-	-
hoteles	21	no	no	rn	negative	-	-	-	no	-	-
hoteles	22	la	lo	pp3fsa00	personal	-	-	-	-	-	-
hoteles	23	tienen	tener	vmip3p0	main	-	-	-	-	-	-
hoteles	24	donde	donde	pr000000	relative	-	-	-	-	-	-
hoteles	25	deberían	deber	vmic3p0	main	-	-	-	-	-	-
hoteles	26	.	.	fp	-	-	-	-	-	-	-

Figure 2.15: Annotated example from the SFU ReviewSP-NEG corpus for negation cue detection in CoNLL format.

UHU-HUVR

The UHU-HUVR (Cruz Díaz et al., 2017) is the first Spanish corpus in which affixal negation is annotated. It is composed of 604 clinical reports from the Virgen del Rocío Hospital in Seville (Spain). 276 of these clinical documents correspond to radiology reports and 328 to the personal history of anamnesis reports written in free text.

In this corpus all types of negation were annotated, syntactic, morphological (affixal negation), and lexical. It was annotated with negation markers, their scopes and the negated events by two domain expert annotators following closely the Thyme corpus guidelines (Styler IV et al., 2014) with some adaptations. In the anamnesis reports, 1,079 sentences (35.20%) were found to contain negations out of 3,065 sentences. On the other hand, 1,219 sentences (22.80%) out of 5,347 sentences were annotated with negations in the radiology reports. The Dice coefficient for inter-annotator agreement was higher than 0.94 for negation markers and higher than 0.72 for negated events. Most of the disagreements were the result of human errors, i.e., the annotators missed a word or included a word that did not belong either to the event or to the marker. However, other cases of disagreement can be explained by the difficulty of the task and the lack of clear guidance. They encountered the same type of disagreements as Jiménez-Zafra et al. (2016) when annotating the SFU Review$_{SP}$-NEG corpus.

The format of the corpus is not specified and the authors say that the annotated corpus will be made publicly available, but it is not currently available probably because of legal and ethical issues.

IULA Spanish Clinical Record

The IULA Spanish Clinical Record (Marimon et al., 2017) corpus contains 300 anonymized clinical records from several services of one of the main hospitals in Barcelona (Spain) that was annotated with negation markers and their scopes. It contains 3,194 sentences, out of which 1,093 (34.22%) were annotated with negation cues.

In this corpus, syntactic and lexical negation were annotated but not morphological negation. It was annotated with negation cues and their scopes by three computational linguists annotators advised by a clinician. The inter-annotator agreement Kappa rates were 0.85 between annotators 1 and 2, and annotators 1 and 3; and 0.88 between annotators 2 and 3. The authors defined their own annotation guidelines taking into account the currently existing guidelines for corpora in English (Mutalik et al., 2001; Szarvas et al., 2008; Morante & Daelemans, 2012). Differently from previous work, they did not include the negation cue nor the subject in the scope (except when the subject is located after the verb).

The corpus is publicly available with a CC-BY-SA 3.0 license and it can be downloaded at `http://eines.iula.upf.edu/brat//#/NegationOnCR_IULA/`. The annotations can be exported in ANN format and the raw text in TXT format. Below, Figure 2.16 presents how sentence of Example (28) is annotated in this corpus.

28. AC: tonos cardíacos rítmicos sin soplos audibles.

    CA: rhythmic heart tones without audible murmurs.

```
T215 NegMarker  119 122 sin
T269 DISO   123 138 soplos audibles
R3 Scope   Arg1:T215 Arg2:T269
```

Figure 2.16: Annotated example from the IULA Spanish Clinical Record corpus.

2.1.3.3 Other corpora

Some corpora have been created for languages other than Spanish and English. To the best of our knowledge, there are also corpora annotated for Swedish, Dutch, Japanese, Chinese, German and Italian. They are presented below.

Swedish uncertainty, speculation and negation corpus

Dalianis & Velupillai (2010) annotated a subset of the Stockholm Electronic Patient Record corpus (Dalianis et al., 2009) with certain and uncertain expressions as well as speculative and negation keywords. The Stockholm Electronic Patient Record Corpus is a clinical corpus that contains patient records from the Stockholm area stretching over the years 2006 to 2008. From this corpus, 6740 sentences were randomly extracted and annotated by three annotators: one senior level student, one undergraduate computer scientist, and one undergraduate language consultant. For the annotation, guidelines similar to those of the BioScope corpus (Vincze et al., 2008) were applied. The inter-annotator agreement was measured by pairwise F-measure. In relation to the annotation of negation cues, only syntactic negation was considered and the agreement obtained was of 0.80 in terms of F-measure. The corpus was annotated with a total of 6996 expressions, out of which 1008 were negative keywords.

The corpus is in XML format, according to the example provided by the authors (Figure 2.17), but there is no information about availability.

29. Statusmässigt inga säkra artriter. Lungrtg Huddinge ua. Leverprover ua.

    Status-wise no certain arthritis. cxr Huddinge woco. Liver samples woco.

```
Bedömning:
<sentence_1>
<Uncertain_expression>Statusmässigt
<Speculative_words><Negation>inga
</Negation> säkra</Speculative_words>
artriter</Uncertain_expression>.
<Certain_expression>Lungrtg Huddinge ua
</Certain_expression>.</sentence>
Leverprover ua.
```

Figure 2.17: Annotated example from the Stockholm Electronic Patient Record corpus.

EMC Dutch clinical corpus

The EMC Dutch clinical corpus was created by Afzal et al. (2014) and it contains four types of anonymized clinical documents: entries from general practitioners, specialists' letters, radiology reports, and discharge letters. Medical terms were annotated using a list of terms extracted from the Unified Medical Language System, and the identified terms were annotated for negation, temporality and experiencer properties. In relation to negation, a term is labeled as 'Negated' if there is evidence in the text suggesting that the condition does not occur or exist, otherwise it is annotated as 'Not negated'. The corpus was annotated by two independent annotators and differences resolved by an expert who was familiar with the four types of clinical texts. An annotation guideline explaining the process and each of the contextual properties was provided, but it is not available. The kappa inter-annotator agreement for negated terms was of 0.90, 0.90, 0.93 and 0.94 for entries from general practitioners, specialists' letters, radiology reports, and discharge letters, respectively. The percentage of negated terms is similar for the different report types:

- Out of a total of 3626 medical terms from general practitioners, 12% were annotated as negated (435).
- Out of a total of 2748 medical terms from specialists' letters, 15% were annotated as negated (412).
- Out of a total of 3684 medical terms from radiology reports, 16% were annotated as negated (589).

- Out of a total of 2830 medical terms from discharge letters, 13% were annotated as negated (368).

This is the first publicly available Dutch clinical corpus, but it can not be accessed online. It is necessary to send an email to the authors.

Japanese negation corpus

citematsuyoshi2014annotating proposed an annotation scheme for the focus of negation in Japanese and annotated a corpus of reviews from "Rakuten Travel: User review data"[15] and the newspaper subcorpus of the "Balanced Corpus of Contemporary Written Japanese (BCCWJ)"[16] in order to develop a system for detecting the focus of negation in Japanese.

The Review and Newspaper Japanese corpus is composed of 5,178 sentences of facilities reviews and 5,582 sentences of Group "A" and "B"of the newspaper documents from BCCWJ. It was automatically tagged with POS tags using the MeCab analyzer[17] so that this information could be used to mark negation cue candidates. After a filtering process, 2,147 negation cues were annotated (1,246 from reviews and 901 from newspapers). Of the 10,760 sentences, 1,785 were found to contain some negation cue (16,59%).

For the annotation of the focus of negation, two annotators marked the focus for Group "A" in the newspaper subcorpus. They obtained an agreement of 66% in terms of number of segments. Disagreement problems were discussed and solved. Then, one of the annotators annotated reviews and Group "B" and the other checked the annotations. After a discussion, a total of ten labels were corrected.

The format of the corpus is not specified, although the authors show some examples of annotated sentences in their work. In Example (30) we present one of them, corresponding to a hotel review. The negation cue is written in boldface and the focus is underlined. In relation to the availability, the authors plan to freely distribute the corpus in their web site: `http://`

[15] `http://rit.rakuten.co.jp/rdr/index_en.html`
[16] `http://www.ninjal.ac.jp/english/products/bccwj/`
[17] `http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html`

cl.cs.yamanashi.ac.jp/nldata/negation/, although it is not available yet.[18]

30. heya ni reizoko ga **naku** robi ni aru kyodo reizoko wo tsukatta.

    The room where I stayed had no fridge, so I used a common one in the lobby.

Chinese Negation and Speculation corpus

Zou et al. (2016) recently presented the Chinese Negation and Speculation (CNeSp) corpus, which consists of three types of documents annotated with negative and speculative cues and their linguistic scopes. The corpus includes 19 articles of scientific literature, 821 product reviews and 311 financial articles. It is composed of 16,841 sentences, out of which 4,517 (26.82%) contain negations.

For the annotation, the guidelines of the BioScope corpus (Szarvas et al., 2008) were used with some adaptation in order to fit the Chinese language. The minimal unit expressing negation or speculation was annotated and the cues were included within the scope, as with the BioScope corpus. However, the following adaptations were realized: i) the existence of a cue depends on its actual semantic in context, ii) a scope should contain the subject which contributes to the meaning of the content being negated or speculated if possible, iii) scope should be a continuous fragment in sentence and iv) a negative or speculative word may not be a cue (there are many double negatives in Chinese, used only for emphasizing rather than expressing negative meaning). The corpus was annotated by two annotators and disagreements were resolved by a linguist expert who modified the guidelines accordingly. The inter-annotator agreement was measured in terms of Kappa. It was of 0.96, 0.96 and 0.93 for negation cue detection and 0.90, 0.91 and 0.88 for scope identification, for scientific literature, financial articles and product reviews, respectively. In this corpus, only lexical and syntactic negation were considered.

The corpus is in XML format and the authors state that it is publicly available for research purposes at http://nlp.suda.edu.cn/corpus/CNeSp/. Below, Figure 2.18 shows an annotation example of a hotel review sentence (Example 2.1.3.3).

[18] Accessed by June 27, 2019.

31. 标准间太差房间还不如3星的而且设施非常陈旧.

    The standard room is too bad, the room is not as good as the 3 stars, and the facilities are very old.

```
<sentence id="S1">
标准间太差
  -<xcope id="X1.1">
    房间还
      <cue type="negation" ref="X1.1">
    不如
      </cue>
    3星的
  </xcope>-
而且设施非常陈旧.
</sentence>
```

Figure 2.18: Annotated example from the CNeSp corpus.

German negation and speculation corpus

The German negation and speculation corpus (Cotik, Roller, et al., 2016) consists of 8 anonymized German discharge summaries and 175 clinical notes of the nephrology domain. It was first automatically annotated using an annotation tool. Medical terms were pre-annotated using data of the UMLS Methathesaurus, and later a human annotator corrected wrong annotations and included missing concepts. Furthermore, the annotator had to decide and annotate whether a given finding occurs in a positive, negative or speculative context. Finally, the annotations were corrected by a second annotator with more experience. There is no mention of annotation guidelines, and inter-annotator agreement is not reported. In relation to negation, out of 518 medical terms from discharge summaries, 106 were annotated as negated. On the other hand, out of 596 medical terms from clinical notes, 337 were annotated as negated.

The format of the corpus is not mentioned by authors and it is not publicly available.

Italian negation corpus

Altuna et al. (2017) proposed an annotation framework for negation in Italian based on the guidelines proposed by Morante et al. (2011) and Blanco & Moldovan (2011a), and they applied it to the annotation of news articles and tweets. They provided annotations for negation cues, negation scope and focus, taking into account only syntactic negation. As a general rule, they do not include the negation cue inside the scope, except when negation has a richer semantic meaning (e.g. nessun / "no" (determiner), mai / "never", nessuno / "nobody", and nulla / "nothing").

The corpus is composed of 71 documents from the Fact-Ita Bank corpus (Minard et al., 2014), which consists of news stories taken from Ita-TimeBank (Caselli et al., 2011), and 301 tweets that were used as the test set in the FactA task presented at the EVALITA 2016 evaluation campaign (Minard et al., 2016). On the one hand, the Fact-Ita Bank Negation corpus consists of 1,290 sentences, out of which 278 contain negations (21.55%). On the other hand, the tweet corpus has 301 sentences and 59 were annotated as negated (19.60%).

The annotation process was carried out by four annotators, whose background is not specified, and the inter-annotator agreement was measured using the average pairwise F-measure. The agreement on the identification of negation cues, scope and focus was of 0.98, 0.67 and 0.58, respectively.

The corpus is in XML format and it can be downloaded under a Creative Commons Attribution-NonCommercial 4.0 International License at `https://hlt-nlp.fbk.eu/technologies/fact-ita-bank` . It should be mentioned that only news annotations are available. Tweets are not available because they are from another corpus that has copyright. Below, Figure 2.19 shows how Example (32), a negation sentence of a new of the corpus, is represented.

32. Pare che, concluso questo ciclo, il docente non si dedichera solo all' insegnamento.

    It seems that, at the end of this cycle, the teacher will not only devote himself to teaching.

```
...
<token t_id="148" sentence="9" number="7">il</token>
<token t_id="149" sentence="9" number="8">docente</token>
<token t_id="150" sentence="9" number="9">non</token>
<token t_id="151" sentence="9" number="10">si</token>
<token t_id="152" sentence="9" number="11">dedichera</token>
<token t_id="153" sentence="9" number="12">solo</token>
<token t_id="154" sentence="9" number="13">all'</token>
<token t_id="155" sentence="9" number="14">insegnamento</token>
<token t_id="156" sentence="9" number="15">.</token>
...
<CUE-NEG m_id="56" focus="62" comment="" reinforcement=""
scope="63" >
<token_anchor t_id="150"/>
</CUE-NEG>
<FOC-NEG m_id="62" comment="" >
<token_anchor t_id="153"/>
</FOC-NEG>
<SCOPE-NEG m_id="63" comment="" >
<token_anchor t_id="148"/>
<token_anchor t_id="149"/>
<token_anchor t_id="151"/>
<token_anchor t_id="152"/>
<token_anchor t_id="153"/>
<token_anchor t_id="154"/>
<token_anchor t_id="155"/>
</SCOPE-NEG>
```


Figure 2.19: Annotated example from the Fact-Ita Bank Negation corpus.

2.2 Sentiment analysis

2.2.1 Definition of sentiment analysis

The web has evolved progressively since its beginning in 1990. At first, the user was almost a passive subject who received information or published it, without many possibilities to generate any interaction. The emergence of the Web 2.0 was a social revolution, because it offered users the possibility of producing and sharing contents, opinions, experiences, etc.

Some years ago it was common to ask family and friends their opinion about a particular topic, but in recent years the number of people using the Internet for this function has greatly increased. The exponential growth of the subjective information available has generated a great interest in the analysis of this information.

Sentiment analysis is an area of Natural Language Processing that focuses on the computational treatment of opinions, sentiments, evaluations, attitudes, and emotions in texts (Liu, 2015). Currently, it is a fashionable task in the field of Natural Language Processing, due mainly to the growing interest in the knowledge of the opinions and emotions of people from different sectors of the society.

We can find complete overviews of the research in sentiment analysis in the literature (Pang et al., 2008; Cambria et al., 2013; Liu, 2015). Sentiment analysis includes the study of several sub-tasks, but perhaps the best known are subjectivity detection, polarity classification and emotion recognition. In this doctoral thesis, we focus on polarity classification task, which is the task of determining the semantic orientation of a subjective text (e.g. positive, negative or neutral). We can distinguish three levels of analysis:

- Document-level polarity classification: identification of the overall opinion expresses in the document (Pang et al., 2002; Turney, 2002; Rosenthal et al., 2017).
- Sentence-level polarity classification: identification of the level of polarity of each sentence of the document (Yu & Hatzivassiloglou, 2003; Wilson et al., 2005; Appel et al., 2016).

- Aspect-level polarity classification: identification of the sentiment of the author towards each entity or aspect reviewed in the document (Thet et al., 2010; Jiménez-Zafra et al., 2016; Pontiki et al., 2016).

Different techniques have been applied to polarity classification, but they can be grouped into machine learning approach, lexicon-based approach and hybrid approach (Maynard & Funk, 2011). On the one hand, the machine learning approach can be divided into supervised and unsupervised learning methods. The supervised methods are based on using a labeled collection of data to train the classifiers and the unsupervised systems are those in which we do not have a set of previously labeled data, but only from the properties of them we try to classify data according to their similarity. On the other hand, the lexicon-based approach consists of computing the semantic orientation of the words in the text taking into account the positive or negative orientation of words (Turney, 2002). For this, lexicons are essential, which are lists of opinion bearing words that allow the identification of positive and negative words in texts. Both methodologies have their advantages and drawbacks. For example, the machine learning approach requires training data, which in many cases are difficult or impossible to obtain. On the other hand, the lexicon-based approach requires a large amount of linguistic resources which generally depend on the language and the domain. Finally, the hybrid approach combines both methods.

Although polarity classification is the most widely studied task in sentiment analysis, several challenges still remain open and are attracting the attention of researchers. One of these is the treatment of some linguistic phenomena such as irony, metaphors or negation. This doctoral thesis focus on the study of one of these phenomena, negation in Spanish texts.

2.2.2 Negation and sentiment analysis

Negation is an open challenge within Natural Language processing in general and within sentiment analysis in particular, since negation can be used to express a negative opinion from the negation of positive terms (Example 33) or, by contrast, a positive opinion can be expressed by the negation of negative terms (Example 34).

33. [El ordenador **no** *funciona* <u>bien^{+}</u>]$^{+}$.

    The computer doesn't work well.

34. [El ordenador **no** *tiene* **ningún** <u>problema^{-}</u>]$^{+}$, funciona bien^{+}.

    The computer doesn't have any problem, it works well.

Example 33 presents a negative opinion about a computer. The word "bien" *[well]* is a polar word, also known as a sentiment word, that is, it is a word with a semantic orientation (Wiegand et al., 2010). Specifically, it has a positive orientation. However, the negation cue "no" *[not]* changes the polarity of it, making the opinion negative.

Notwithstanding, the presence of negation in a sentence does not imply a negative opinion nor that all words of the sentence reverse their polarity. In Example 34, there is a sentence with a discontinuous negation cue "no-ningún" *[doesn't any]* that changes the prior polarity of the negative word "problema" *[problem]* into positive, making the opinion positive. In addition, we can see that the polarity of the word "bien" *[well]* is not reversed, because it is not in the scope of negation[19].

Moreover, negation does not always imply polarity reversal, it can also modify the degree of sentiment expressions due to the presence of intensifiers and diminishers (Example 35) or have no effect on them (Example 36), which makes the task even more difficult.

35. [La película **no** *es* **muy** <u>buena</u>]

    The film is not very good.

36. No solo es barato sino que también funciona muy bien

    Not only is it cheap but it also works very well.

[19] In the examples presented in this doctoral thesis, the negation cue appears in bold, the event in italics, the focus underlined and the scope between brackets.

There are even some good surveys about the study of negation as a linguistic phenomenon (Morante & Sporleder, 2012a) and concerning sentiment analysis(Wiegand et al., 2010). Below, it is presented the state of the art related to the application of negation for the improvement of sentiment analysis systems in the reference language, English, and in Spanish, the language of study in this doctoral thesis.

2.2.2.1 Related research applying negation for sentiment analysis in English

Most research has focused on opinions written in English. They are not accurate enough since they do not assess the processing of negation. They identify negation cues using a lexicon and they have relatively straightforward conceptualizations of the scope of negation. One of the first approaches was proposed by Das & Chen (2001) who use a simple method that adds "NOT" to the terms of the sentence that appear next to a negation cue, such as "no" or "don't". Pang et al. (2002) follow the same approach, but they assume that the negation cues ("not", "isn't", "didn't", etc.) affect all the terms from the cue to the end of the sentence. The authors carry out different experiments with and without negation using machine learning algorithms. However, the results show no significant differences considering negation or not. Polanyi & Zaenen (2006) not only considered negation but they also study intensifiers and diminishers, introducing the new concept "Contextual valence shifters" (i.e., negations, intensifiers and diminishers). They present the first model that assigns scores to opinionated words, reversing the polarity of negated words. However, they do not implement this model, so we can only speculate on its effectiveness. Kennedy & Inkpen (2006) used a similar methodology where negations reverse the semantic polarity of a particular term, while intensifiers and diminishers increase and decrease, respectively, the degree to which a term is positive or negative. They take as the scope of negation cues, intensifiers and diminishers the first sentiment-carrying word following them. They used two methods to classify opinions, the first one consists of classifying a review according to the number of positive and negative opinion words it contains, and the second is based on the use of the Support Vector Machines (SVM) algorithm, concluding that the treatment of negation is an important fact. Wilson et al. (2005) propose to use a fixed window of 4 words to determine the scope of negation. These works are the pioneers in the modeling of negation for English sentiment analysis, but the scientific community is still working on this issue since the approaches presented so far are not accurate enough.

Other works propose more robust approaches, based on the definition of linguistic rules from syntactic dependence trees, as the one of Jia et al. (2009), applying more complex calculations in order to obtain polarity in opinions (Taboada et al., 2011) or using deep learning (Socher et al., 2013). Jia et al. (2009) develop a rule-based system that uses information derived from a parse tree. This algorithm computes a candidate scope, which is then pruned by removing those 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 do not assess the processing of negation, probably due to the lack of annotated corpora for negation in the review domain, but evaluate the effectiveness of their approach on polarity determination. The results show that the identification of the scope of negation improves both the accuracy of sentiment analysis and the retrieval effectiveness of opinion retrieval. Taboada et al. (2011) define SO-CAL, a lexicon-based model that deal with negation and intensification. They handle negation by first identifying a sentiment word and tracking back to the previous words searching for a negation cue. Moreover, they introduce a new way of dealing with negation that consists of reducing the polarity value of negated words instead of reversing it. Socher et al. (2013) propose the Recursive Neural Tensor Network (RNTN) model, which represents a phrase through word vectors and a parse tree and then compute vectors for higher nodes in the tree using the same tensor-based composition function. They use a test set of positive and negative sentences and show that the RNTN model accurately captures the sentiment change and scope of negation.

Some of the works described in Subsection 2.1.2 also explore how the proposed negation detection systems can improve the results of the sentiment analysis task. For instance, Councill et al. (2010) explain that, as they expected, the performance of their sentiment-analysis system is improved dramatically by introducing negation scope detection. Lapponi et al. (2012) use their system for negation resolution as a component in a simple negation-aware testbed for sentiment classification. Results show that all negation-aware configurations are beneficial in terms of the combined F1-score. Reitan et al. (2015) develop a sentiment classifier for Twitter data, confirming that taking negation into account tends to improve sentiment classification performance significantly. Cruz et al. (2016) conclude in their study that the correct identification of negation and speculation cues and their scopes is vital for the task of sentiment analysis. Pröllochs et al. (2016) examine how detecting negation scopes can improve the accuracy of sentiment analysis for financial news which reveals negation scope detection as a crucial leverage in decision support from sentiment.

Among the research that has examined the role of negation in sentiment analysis, we will highlight a few more studies. For instance, Dadvar et al. (2011) investigate the problem of determining the polarity of sentiments in film reviews when negation cues, such as *not* and *hardly*, occur in sentences. The authors observe significant improvements in the classification of the documents after applying negation detection. Hogenboom et al. (2011) show that properly accounting for negation when analyzing sentiment in natural language texts may help improve the classification of unseen natural language text as carrying either a positive or a negative sentiment. Asmi & Ishaya (2012) propose a framework for automatic identification of opinions in textual data, including rules for negation recognition and calculation especially designed to improve sentiment text analysis. For ChandraKala & Sindhu (2012), negation detection is one of the most important pre-processing steps in identifying opinions efficiently.

More recently, Ohana et al. (2016) investigate whether the treatment of negative sentiment in negated text can improve the performance of sentiment classification tasks. They propose a novel adjustment factor based on negation occurrences as a proxy for negative sentiment polarity. This shows statistically significant performance improvements on all domains tested. Sharif et al. (2016) detect the effect of negation on consumer reviews which appear positive but are in fact completely negative in meaning. Their proposed negation approach presents a way of calculating negation identification that helps to improve review classification accuracy. Diamantini et al. (2016) experiment with different datasets, proving that their proposed negation-handling algorithm based on dependency-based parse trees achieves better sentiment analysis accuracy. Farooq et al. (2017) show that their proposed negation-handling method improves the accuracy of both negation scope identification and overall sentiment analysis. In a recent work, Hussein (2018) find that negation is the most important challenge with the greatest impact on any sentiment analysis. They come to this conclusion through a comparison between 41 papers in sentiment analysis challenges.

2.2.2.2 Related research applying negation for sentiment analysis in Spanish

The first work we are aware of in which negation is applied for improving Spanish sentiment analysis is the one of Brooke et al. (2009). In this work, authors adapt to Spanish the tool that they developed for the classification of English reviews, SO-CAL (Taboada et al., 2011). It is a lexicon-based sentiment analysis system that integrates dictionaries with positive and negative terms. Moreover, it takes into account intensifiers and negation, using a rule-based method for the identification of the scope of negation. As it has been previously described, they first identifying a sentiment word and tracking back to the previous words searching for a negation cue. If the word is in the scope of negation, they reduced its polarity value instead of reversing it.

We can also find window-based approaches for the identification of the scope of negation, as those proposed by (Anta et al., 2013) and (Gamallo et al., 2013). (Anta et al., 2013) conduct experiments for the classification of tweets according to sentiment and topic. They evaluated different features, including negation, and different algorithms. When a negation cue is detected, they reversed the sign of the 3 terms that follow it. However, they do not evaluate the effect of negation independently and do not provide any conclusions regarding it. (Gamallo et al., 2013) also classify tweets taking into account negation using different configurations of the Naive Bayes classifier. For scope detection, they take into account the PoS tag of the negation cues and search for a polarity word (noun, verb, or adjective) within a window of 2 words after it. If a polarity word is found and is syntactically linked to the negation cue, then its polarity is reversed. They define rules to specify when a word is syntactically linked to a negation cue. For example, for the adverb “no” *[not]*, the system only reverses the polarity of verbs or adjectives, and for determiner “ninguno” *[none]*, only the polarity of nouns is reversed. However, these authors do not evaluate the effect of negation in sentiment analysis, they simply take it into account as one more feature.

Vilares et al. (2013, 2015) have also worked on this task. They develop a syntactic-based method for sentiment analysis and use dependency-based methods for the treatment of negation, intensification and subordinate sentences. They study the negation cues "no" *[not]*, "sin" *[without]* and "nunca" *[never]*. Their results show that taking into account the syntactic structure of the text improves over machine learning and lexicon-based approaches on the review domain. However, they do not analyze the gain obtained using negation individually and, therefore, it is not possible to determine which is the module responsible for the improvement obtained.

We have also contributed to this task and the research carried out will be described on detail in Chapter 3: "'*Preliminary research*'. We study the most important cues[20] according to La Real Academia Española (Española, 2009) and propose a set of rules based on dependency trees for identifying the scope of these negation cues (Jiménez-Zafra et al., 2015). This module is integrated into a lexicon-based sentiment analysis system for polarity classification. We experiment on different domains, but we study Twitter in more depth, taking into account the peculiarities of the language used in this social medium (Jiménez-Zafra et al., 2019). We statistically demonstrate that the results obtained considering the negation module are significantly greater than those obtained without taking negation into account. Moreover, we 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 our approach is better.

Miranda et al. (2016) present an approach based on the use of the lexicon ANEW adapted to Spanish (Redondo et al., 2007). They conduct different experiments considering negation on a set of hotels reviews from the TripAdvisor website (Molina-González et al., 2013) using the negation cues studied by Jiménez-Zafra et al. (2015). They obtained slightly better results incorporating negation and outperformed other approaches tested on the same corpus, although they only report results on the precision of the system.

Amores et al. (2016) combine different methods for handling negation. They adapt the rules

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

proposed by Vilares et al. (2013) for determining the scope of the negation cues "no" *[not]*, "sin" *[without]* and "never" [never]. Moreover, they define a set of negation cues from already published lists, but do not specify which ones, and consider as scope the two following terms. Finally, they also take into account the use of some negation affixes. To evaluate the impact of negation on sentiment analysis they use the lexicon-based system PosNeg (Amores Fernández, 2016) and conduct experiments on a subset of the Amazon corpus[21] (Wang et al., 2011) composed only of those comments with some of the negation cues under consideration in their work. The results obtained show that negation improves the quality of PosNeg system for the polarity detection of opinions.

[21] `http://times.cs.uiuc.edu/~wang296/Data/`

Chapter 3

Preliminary research

The beginning of this research takes place thanks the book of Bing Liu on sentiment analysis Liu (2015). In this book Bing Liu mentions that, although polarity classification is the most widely studied task in sentiment analysis, there are some open challenges, such as negation, the treatment of which could improve the predictive capacity of sentiment analysis systems.

The study of the state of the art on systems integrating negation to improve the classification of opinions in Spanish make us realize that, in most of the works, negation is taken into account as one more feature, but its effect on the classification is not evaluated.

Therefore, in a first stage, we conduct a study to check whether the detection and integration of negation into a Spanish polarity classification system can improve the accuracy of the final system. To this end, we define rules based on dependency trees to identify the scope of the main negation cues according to the grammar of La Real Academia Española (Royal Spanish Academy) (Española, 2009). We develop a system for negation scope identification based on these rules and integrate it into a sentiment analysis system for polarity classification. In addition, we propose a method to correctly evaluate the role of negation in sentiment analysis.

We experiment on different domains, but we study Twitter in more depth as it is one of the main social media where people publish their opinion these days. This is the study presented in this chapter. It constitutes the preliminary research and has been decisive in the development of this doctoral thesis since it has allowed to detect the importance of a correct negation processing and the deficiencies of the sentiment analysis systems developed until now.

Below, we present the negation detection module and talk about how Twitter is considered one of the main sources of opinion that can be exploited by the sentiment analysis community. We then introduce the resources used in this study and describe the architecture of the sentiment analysis system in which we integrate the negation detection module. Later, we present the experiments carried out, the results obtained and an analysis of them. Finally, we report the conclusions of this preliminary research.

3.1 Negation detection module

Bearing in mind that negation is a linguistic phenomenon and the structure of the sentence clearly influences which words are within the scope of negation, our first approach for the identification of the scope of negation makes use of syntactic relations.

Table 3.1: Rules for identifying the scope of negation cues.

Cue	Rule for scope identification
no (not), *tampoco* (neither), *nadie* (nobody), *jamás* (never), *ninguno* (none)	Parent node and the tree formed by the brother of the right, included
ni (nor), *sin* (without)	All children and all trees formed by them until reaching leaf nodes
nada (nothing), *nunca* (never)	Parent node

We propose a set of rules based on dependency trees for identifying the scope of some negation cues. In particular, we study the most important according to the grammar of La Real Academia Española (Royal Spanish Academy) (Española, 2009): "no" *[not]*, "tampoco" *[neither]*, "nadie" *[nobody]*, "jamás" *[never]*, "ni" [nor], "sin" [without], "nada" [nothing], "nunca" *[never]* and "ninguno" *[none]* . For each negation cue, a rule for determining its scope is defined. For this, we analyze the dependency trees of diverse sentences extracted from different websites in which some of the cues considered appear. To build the dependency trees we use the dependency parser of Freeling[1] (Atserias Batalla et al., 2005), which generates the dependency tree of a sentence based on its syntactic structure. After the study of these trees, we realize that it is possible to generalize the treatment of these negation cues in 3 rules (Table 3.1), so we decide to continue with the research and apply them to Spanish sentiment analysis.

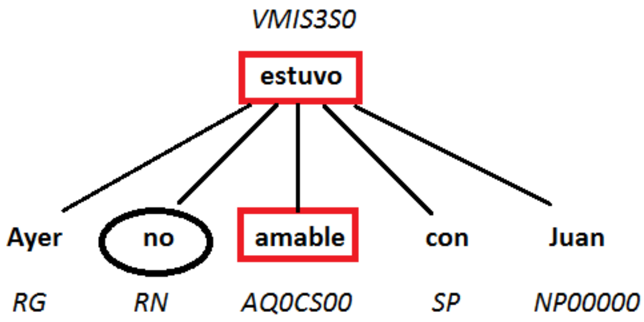

Figure 3.1: Dependency tree of the negation cue *no* (not). Sentence: "Ayer no estuvo amable con Juan" *[Yesterday he was not kind to Juan]*.

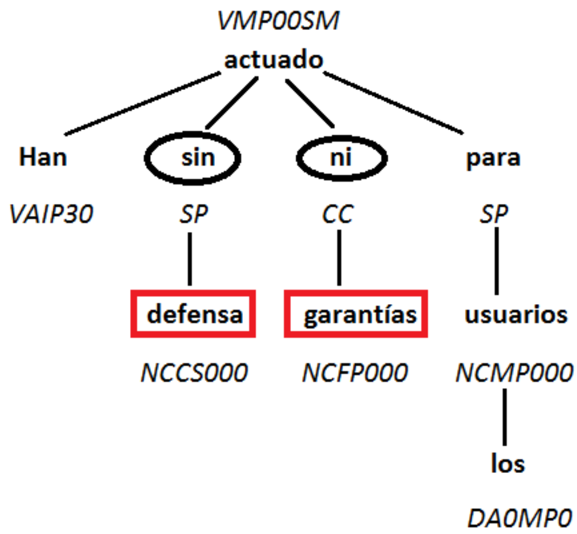

Figure 3.2: Dependency tree of the negation cues *sin* (without) and *ni* (nor). Sentence: "Han actuado sin defensa ni garantías para los usuarios" *[They have acted without defense nor guarantees for the users]*.

In order to clarify the rules that have been defined, an example of the application of each rule is shown in Figure 3.1, Figure 3.2 and Figure 3.3. Each figure represents the dependency tree

[1]Freeling (Padró & Stanilovsky, 2012) is an open-source language-analysis toolkit that is available for several languages, including Spanish. It is available at http://nlp.lsi.upc.edu/freeling/

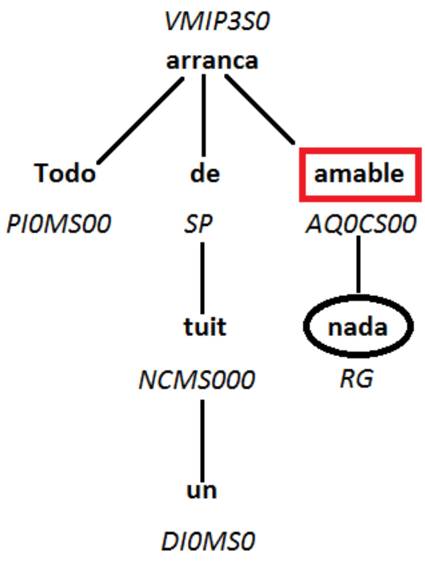

Figure 3.3: Dependency tree of the negation cue *nada* (nothing). Sentence: "Todo arranca de un tuit nada amable" *[It all starts with an unkind tweet]*.

related to a sentence in which the negation cue is marked with an ellipse and its scope with a box.

The integration of these rules into a polarity classification system allows us to tag the words that are in the scope of any of the negation cues studied, with the aim of taking this into account when the polarity of a text is determined. For example, in the sentence "Han actuado sin defensa ni garantías para los usuarios." *[They have acted without defense nor guarantees for the users.]* (Figure 3.2), the system will detect that there are two negation cues in the sentence, "sin" *[without]* and "ni" *[nor]*, and for each one it will determine its scope using the rules defined. In this case, both cues affect all children nodes and all trees formed by them until reaching leaf nodes. Therefore, the words "defensa" *[defense]* and "garantías" *[guarantees]* will be tagged as negated words in order to modify their polarity value.

3.2 Sentiment analysis on Twitter

Twitter is a micro-blogging service launched in October 2006 that allows users post "tweets", which are short messages currently limited to 280 characters. These messages usually include user's opinions and feelings, such as opinions about products and services or political or religious views. Thus, researchers, politicians and the public in general have realized that Twitter is a valuable source of information concerning people's opinions and sentiments. Twitter is becoming one of the main social media in our current society and sometimes it is considered as a thermometer of social problems and events. It has clearly changed how we interact and communicate with each other.

Since 2009 the sentiment analysis research community has started to face the problem of the computational treatment of opinions, sentiments and subjectivity in the short texts of Twitter. Researchers realized that Twitter is a valuable source of information concerning people's opinions and sentiments from which it was easy to download information to generate corpora and to extract knowledge. Perhaps, one of the first work related to sentiment analysis and Twitter is the one of Go et al. (2009), who analyze the most suitable lexical features to represent a tweet and use different machine learning approaches to build a classifier for determining the polarity of tweets. They follow the procedure described in Read (2005) in order to build the corpus using emoticons to tag positive and negative posts. After this study, a wide range of methods for SA on Twitter have been published, describing systems with different features and methodologies including machine learning systems (Kouloumpis et al., 2011; Tang et al., 2014), lexicon-based approaches (Montejo-Ráez, Díaz-Galiano, et al., 2014; Montejo-Ráez, Martínez-Cámara, et al., 2014) and hybrid methods (Ghiassi et al., 2013; Khan et al., 2015). Moreover, the growing interest of the research community has been reflected in the organization of several workshops which have created benchmark datasets and have enabled direct comparison between different approaches, both as part of the competition and beyond. The most relevant are the International Workshop on Semantic Evaluation (SemEval), whose first edition was held in 2013 (Nakov et al., 2013), and the Workshop on Semantic Analysis at SEPLN (TASS) that took place for the first time in 2012 (Villena-Román et al., 2013).

Twitter is one of the main channels in which opinions are expressed today. Therefore, we decided to integrate the negation detection module into a polarity classification system of tweets. Specifically, we develop a lexicon-based system. Below we present the resources used, the architecture of the system developed, the experiments carried out and an analysis of them.

3.3 Resources

The polarity classification system developed for this study follows a lexicon-based approach, so some sets of sentiment-bearing expressions have been employed. Specifically, we consider a list of opinion words, a set of emoticons, and a list of hashtags that express sentiment.

In addition, we have used a corpus of Spanish tweets for the assessment. Currently, two corpora of Spanish tweets are available for the research community. The first one is the corpus used in the TASS workshop (Villena-Román et al., 2013), and the second one is the Corpus Of Spanish Tweets COST[2] (Martínez-Cámara et al., 2015). We have chosen the TASS corpus for several reasons: i) it is broadly known by the Spanish research community, due to the fact that it has been used in the editions of the TASS workshop; ii) it has about 68,000 tweets, considerably more tweets than the COST corpus, which is only composed of 34,634 tweets; and iii) it was labelled following a semi-automatic process while the COST corpus was labelled following a noisy label approach, which is similar to the one employed in (Go et al., 2009).

3.3.1 iSOL lexicon

Although Spanish sentiment analysis is attracting more and more researchers, the number of opinion lexicons is scarce compared to those available for English. For English sentiment analysis we can find several resources such as the opinion lexicon compiled by Bing Liu (Hu & Liu, 2004), the MPQA lexicon (Wilson et al., 2005), General Inquirer (Stone et al., 1966), SentiWordNet (Baccianella et al., 2010) and so on.

However, for Spanish the number of resources is limited. We have used the iSOL lexicon because it has been successfully applied in other studies. iSOL is a Spanish lexicon composed of 8,135 opinion words (2,509 positive words and 5,626 negative words). This resource was created taking as a basis the list of opinion words compiled by Bing Liu, which was translated into Spanish. Subsequently, the translated version of the list was manually reviewed and it was completed with more Spanish terms in order to obtain a more representative list of Spanish

[2]http://sinai.ujaen.es/cost

opinion words. All the details of the compilation process of iSOL can be found thoroughly described in Molina-González et al. (2013). This resource has been successfully evaluated in several corpora which demonstrates its validity for sentiment analysis in Spanish (Martínez-Cámara et al., 2014; Molina-González et al., 2015; Jiménez-Zafra et al., 2016).

3.3.2 Hashtags, emoticons and laughs

The language used in Twitter has two special elements that are constantly typed by users, mentions and hashtags. A mention is a explicit reference from one user to another through writing the username preceded by the @ symbol. A hashtag is a string preceded by the hash key (#), and it is usually employed in order to identify the main topic, the sentiment or the semantic orientation of the tweet. Thus, taking into consideration hashtags in the process of polarity classification of tweets in Spanish could be a good idea. S. M. Mohammad (2012) studies the effect of hashtagging emotions such as joy, sadness, anger and surprise in order to express the general emotion or sentiment in a tweet. Later, S. Mohammad et al. (2013) compile a lexicon of opinion using hashtags in English. As far as we know, at the time of this study, there was not a lexicon of Spanish opinion hashtags available, so we compiled one. For this, we used a seed of positive hashtags (*#bueno* (#good), *#bien* (#well), *#positivo* (#positive), *#fantastico* (#great), *#excelente* (#excellent), etc.) and another of negative hashtags (*#malo* (#bad), *#mal* (#bad), *#terrible* (#terrible), *#negativo* (#negative), *#horrible* (#horrible), etc.) and retrieved all the tweets that had any of the seed words for three days. Then, we extracted all the hashtags present in those tweets and classified them as positive or negative depending on whether they appeared in the same tweet of a positive or negative seed. Finally, we manually reviewed these hashtags in order to obtain the final lists. In this way, the hashtags lexicon[3] was compiled and it is composed of 172 positive and 127 negative hashtags.

Emoticons are other indicators of polarity that should be taken into account and there are studies that demonstrate their potential. Read (2005) shows that when the author of an electronic communication uses an emoticon, he/she is effectively marking up the text with an

[3] http://sinai.ujaen.es/hashtags-sp

emotional state. Go et al. (2009) use emoticons to build one of the first corpus of tweets for sentiment analysis. Martínez-Cámara et al. (2015) also use emoticons to compile a corpus of positive and negative tweets written in Spanish. Therefore, according to the emotions itemized in Wikipedia[4], two lists of emoticons were generated[5]: one of them with 70 positive emoticons and another one with 46 negative emoticons.

Laughs are another element frequently used in Twitter. For identifying them we define a regular expression with the main forms of writing laughs in Spanish and variants thereof: jajaja, jaaajajaj, jijiji, jijiij, lol, loool, etc.

3.3.3 The TASS corpus

The TASS corpus was published for the first time in 2012 and since then it has been used in all the subsequent editions of the TASS workshop, so up until now it is the main corpus of Spanish tweets tagged for sentiment analysis. The corpus contains over 68,000 tweets gathered between November 2011 and March 2012. The tweets are written in Spanish by about 150 well-known personalities and celebrities of the world of politics, economy, communication, mass media and culture. The corpus is divided into two sets: training (10%) and test (90%), so the training set is composed of 7,219 tweets and the test one is formed by 60,017 tweets. Each tweet in both sets is tagged with its global polarity, indicating whether the text expresses a positive, negative or neutral sentiment, or no sentiment at all, and can belong to one of the following five levels: strong positive (P+), positive (P), neutral (NEU), negative (N), strong negative (N+) or no sentiment (NONE). In our experiments we discard tweets tagged with NONE class and only consider Positive, Negative and Neutral classes. Thus, original strong positive (P+) and positive (P) tweets are grouped into one unique positive class (P). Similarly, strong negative (N+) and negative (N) are considered as negative class (N). After all this processing, the final set of tweets used for the assessment consists of 22,233 positive tweets, 1,305 tweets labelled as neutral, and 15,844 negative tweets, which is a total of 39,381 tweets.

[4] http://es.wikipedia.org/wiki/Anexo:Emoticonos
[5] http://sinai.ujaen.es/emoti-sp

3.4 System architecture

As we have mentioned earlier, the aim of this study is to demonstrate that taking into account negation is useful for the polarity classification of Spanish tweets. To verify this assertion, we propose an unsupervised lexicon-based system made up of different components. The main contribution of this system is the development of a normalization module that corrects misspelled words, another that detects the presence of a negation cue in a tweet and determines its scope using the rules defined, and the compilation of a Spanish opinion hashtags lexicon. The approach used for determining the polarity of a tweet is straightforward because our goal is not focused on demonstrating that our system is a good polarity classifier but showing that treatment of negation is useful in such systems. The process to obtain the polarity of each tweet can be summarized in five steps:

1. Tokenize the tweet.
2. Correct misspelled words.
3. Determine the part of speech of each word and the lemma of each verb.
4. Detect the presence of negation cues and identify the scope of each of them using the rules defined.
5. Obtain the polarity of the tweet.

The process outlined is shown in Figure 3.4. Below, a detailed explanation of all elements is shown with the sample tweet “Todo arranca de un tweet nada amaaable. #maldad =(” *[It all starts with an unkind tweet. #wickedness = (].*

1. **Tokenization:** In order to process the tweet text, we perform sentence splitting and word tokenization. For this, the Freeling splitter and an adapted version to the Spanish language of the Christopher Potts' tokenizer[6] are used. The tokenizer developed takes

[6] http://sentiment.christopherpotts.net/tokenizing.html#sentiment

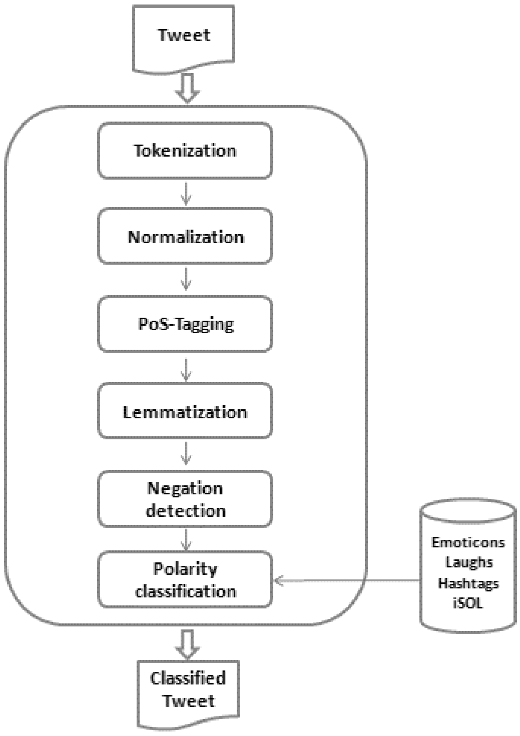

Figure 3.4: Architecture of the polarity classification system.

into account all special features of the language used in Spanish tweets: emoticons, urls, mentions, hashtags, dates, multi-words, etc. Below, the tokens that the system would identify in the sample tweet are shown in square brackets:

```
[Todo] [arranca] [de] [un] [tweet] [nada] [amaaable] [.] [#maldad] [=(]
```

2. **Normalization:** After the identification of the tokens, the next step is to perform a normalization process in order to correct all misspelled words and to mark the tokens that have repeated letters. We mark the tokens that have repeated letters to consider their intensity when we calculate the overall sentiment of the tweet. The two reasons for performing normalization to correct spelling errors are, firstly, that our system needs to build the syntactic tree of each tweet, so if there are fewer misspellings in the text the dependency parser will be more likely to be successful[7]. The second reason is that the system is based on the use of the lexical resource iSOL which is a list of words, most of

[7]Although we think that the use of a specialized parser is better for the processing of tweets, we also support the idea that while specialized parsers are not available, a standard parser can be used. At the time of this study, there is no specific parser for Spanish tweets so the Natural Language Processing tool most used for Spanish (Freeling) has been used. Moreover, due to the fact that tweets are informal texts, we apply a spelling checker in order to keep the number of errors as low as possible and to make the dependency parser work successfully.

them well-written. The spelling corrector of Peter Norvig[8] has been modified with the aim of correcting misspellings in Spanish texts. This spelling corrector only needs a large corpus in the target language. In our case, the target language is Spanish, so we have to compile a representative corpus of Spanish. This large corpus is composed of a list of Spanish lemmas, a list of Spanish verb conjugations and a list of Spanish names and surnames. All the lists were compiled by Ismael Olea[9]. Moreover, we complemented it with the list of words of the corpus CREA[10], which was compiled by La Real Academia Española (Royal Spanish Academy). Normalization of the sample tweet would correct the token `[amaaable]` and would also mark it as a token with repeated letters:

[Todo] [arranca] [de] [un] [tweet] [nada] **[amable]** [.] [#maldad] [=(]
Repeated letters

3. **PoS-Tagging and Lemmatization:** The third step is to learn the PoS-tag of each token in order to obtain the lemma of each verb, because iSOL does not have all the verbal forms of polar verbs, it only has the lemma of each one. Therefore, we use the Part-of-Speech tagger module of Freeling. This resource has two different modules for performing PoS tagging (Padró & Stanilovsky, 2012). The first one is the hmm tagger which is a classical trigram Markovian tagger (Brants, 2000) and the second one, named the relax tagger, is a hybrid system capable of integrating statistical and hand-coded knowledge (Padró, 1998). We used the hmm tagger because it is faster than the relax tagger. In the case of the sample tweet, the system would tag each token with its pertinent part of speech and would obtain the lemma of the token `[arranca]` because it is a verbal form.

[Todo] **[arrancar]** [de] [un] [tweet] [nada] [amable] [.] [#maldad] [=(]
Repeated letters

4. **Negation detection:** This module, in the first place, detects whether the tweet has any negation cue and if so, it determines the scope of each cue with the set of syntactic rules that has been defined (Table 3.1). In this way, if a tweet has a negation cue the system

[8] http://norvig.com/spell-correct.html
[9] http://olea.org/proyectos/lemarios/
[10] Royal Spanish Academy (http://www.rae.es): Data Bank (CREA) online. Current Spanish Benchmark Corpus. http://corpus.rae.es/creanet.html

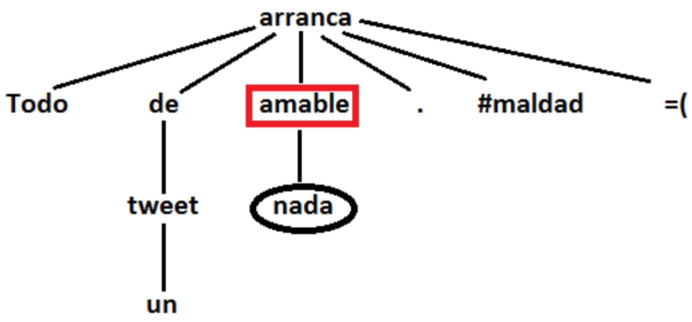

Figure 3.5: Dependency tree of the tweet: "Todo arranca de un tweet nada amable. #maldad =(" *[It all starts with an unkind tweet. #wickedness =(]*.

will generate its dependency parser and will mark each word affected by the negation as "negated" by "name_of_the_cue", in order to take this into account when the semantic orientation of the tweet is calculated. In the sample tweet there is a negation cue, the token `[nada]`. In this case, the system would generate the dependency tree of the tweet (Figure 3.5) and would mark as negated by `[nada]` the token `[amable]` that is in its scope according to the rule defined.

[Todo] [arrancar] [de] [un] [tweet] **[nada]** **[amable]** [.] [#maldad] [=(]
Repeated letters
Negated by nada

5. **Polarity classification:** The last step is to determine the polarity of the tweet. For this purpose, we develop a polarity classifier that takes into account the presence of emoticons, hashtags, expressions of laughing and negation. This component uses the resources described in Section 3.3: the bag of words of emoticons tagged as positives and negatives, the bag of hashtags and iSOL lexicon. For each tweet the classifier determines its positivity and negativity value. Thus, if a token is in the bag of positive/negative emoticons, a polarity value of 2 is added to its positivity/negativity value. If it detects that a token is an expression of laughing, the positivity value is increased by 2. In other case, if the token is in the bag of positive/negative hashtags, the counter of positivity/negativity is increased by 2. Finally, if the token is in the iSOL positive/negative list, a polarity value of 1 is added to the positivity/negativity counter and if it also has repeated letters the value is increased by 1. If the token is negated its polarity is reversed (positive →

negative, negative → positive). Using these values, the system is able to classify the tweet in one of the 3 defined classes following the Equation 3.1:

$$polarity(tweet) = \begin{cases} P & \text{if } pv > nv \\ NEU & \text{if } pv = nv \\ N & \text{if } pv < nv \end{cases} \tag{3.1}$$

where pv and nv are the positivity and negativity value of the tweet respectively.

According to this approach, the sample tweet would be classified as negative because its positivity value would be 0 and its negativity value would be 7. Below, these values are explained with details (both values are initialized to 0, $nv = 0$, $pv = 0$):

- Negativity value:
  - `Arrancar` is a verb that belongs to the list of negative words of the iSOL lexicon ($nv + 1 = 1$).
  - `Amable` is an adjective that belongs to the list of positive words of iSOL, but it is tagged as negated token because it is in the scope of the negation cue *nada*. So its polarity value is reversed (positive → negative) ($nv + 1 = 2$) and it also has repeated letters ($nv + 1 = 3$).
  - `#maldad` is a negative hashtag ($nv + 2 = 5$).
  - `=(` is a negative emoticon ($nv + 2 = 7$).
- Positivity value:
  - As we have mentioned before, `Amable` is a positive adjective, but it is in the scope of the negation cue *nada* so its polarity value is reversed. Therefore, the positivity value remains 0 ($pv = 0$).

[`Todo`] [`arrancar`] (-1) [`de`] [`un`] [`tweet`] { [`nada`] [`amable`] (+2) } (-2) [`.`] [`#maldad`] (-2) [`=(`] (-2) Note: “amable” adds two negative points because it is a positive opinion word with repeated letters and it is negated by *nada*.

3.5 Experiments

Our goal is to study whether the application of certain rules for detecting the scope of negation provides benefits in the polarity classification of Spanish tweets. Moreover, we compare the rules-based approach that we propose with the most widely used model to address negation in English Twitter sentiment analysis (Potts, 2011b). In order to explore the contribution of this linguistic phenomenon for sentiment analysis we carry out the following experiments:

- Without negation (Baseline = BS): using the system described, but without taking into account negation.
- With a baseline for negation (BSN): using the BS system with the negation approach proposed by Potts (2011b) that considers as scope of negation all terms from a negation cue to the next of the following punctuation marks: ".", ":", ";", "!", "?". This method is the most widely used in order to determine the scope of negation in English Twitter sentiment analysis.
- With Negation Rules (NR): using the BS system described in the paper, but including the module that detects the presence of a negation cue in a text and determines its scope using the syntactic rules defined.

The experiments are conducted on the tweets of the TASS corpus considering three cases:

- *Total*: all the tweets of the corpus tagged as positive (P), negative (N) or neutral (NEU) are considered. The total set is composed of 39,382 tweets.
- *NegCue*: we only take into account the tweets of the corpus that have any of the negation cues studied in this paper. The experiment is carried out over a set of 8,604 tweets.
- *RuleAffect*: we only consider the tweets that contain some of the negation cues studied and also polar tokens which are in the scope of these cues, i.e., opinion words affected by

negation. The total number of tweets is 2,326.

$$Total \supset NegCue \supset RuleAffect$$

As we can see, *RuleAffect* is a subset of *NegCue* and *NegCue* is a subset of *Total*. The reason why the dataset has been reduced to carry out experiments with two subsets is because most of the tweets in the corpus do not have negation cues (Table 3.2), and so in order to determine whether negation improves the polarity classification we need to compare the subsets with and without negation cues.

Table 3.2: Tweet sets used in the experiments.

	Tweets	Percentage
Total	39,382	100%
NegCue	8,604	22%
RuleAffect	2,326	6%

Specifically, in order to evaluate the rules defined we should consider the tweets with polar tokens affected by the negation because these are the tweets in which the rules have actually been applied. The rules are based on reversing the polarity of the words that are in the scope of negation, but if the lexicon used does not detect a polar token the rules can not be applied. The aim of this study is not to present the best polarity classification system but to check whether the rules that we have introduced can improve the classification of Spanish tweets.

In order to clarify the difference between these 3 datasets, we show an example of the type of tweets present in each one. Below we can see a sample of a tweet that belongs to the *Total* set, but is not included in any other subset (*NegCue*, *RuleAffect*) because it does not have any negation cue.

> *Gonzalo Altozano tras la presentación de su libro 101 españoles y Dios. Divertido, emocionante y brillante.* (Gonzalo Altozano after the presentation of his book 101 Spaniards and God. Fun, exciting and brilliant.)

Subsequently, a tweet of the *NegCue* subset is shown. This sample is part of the *Total* set, but it does not belong to the *RuleAffect* subset because the token *exacerbar* is not identified as a polar token by the iSOL lexicon, and consequently the rule of the negation cue *no* cannot be applied.

> *RT @usuario: Yo cuidaría como gestionar todo este panorama para no exacerbar. Nos jugamos demasiado...* (RT @user: I would look after how to manage this panorama to avoid exacerbating. We have a lot at stake...)

Finally, a tweet of the *RuleAffect* subset is presented. In this case, the term *amable* is in the scope of the negation cue *nada* and it is identified as an opinion word by the iSOL lexicon, so the rule can be applied and evaluated. As we can see, this subset only has tweets with negation cues and polar tokens affected by the negation. This type of tweet is also included in the *NegCue* subset and in the *Total* set ($Total \supset NegCue \supset RuleAffect$).

> *Todo arranca de un tuit nada amable hacia Mou. Muchas de las críticas me descalificaban por ello como periodista en RNE. De ahí la pregunta.* (It all starts with an unkind tweet to Mou. Many of the criticisms blamed me for it as a journalist in RNE. Hence the question.)

3.6 Results

We first present the evaluation measures used and then we show the polarity classification results obtained in the different experiments.

3.6.1 Evaluation measures

For the evaluation of the experiments we calculate the traditional measures in text classification: Precision (P), Recall (R), F-score (F1) and Accuracy (Acc).

$$P = \frac{TP}{TP + FP} \tag{3.2}$$

$$R = \frac{TP}{TP + FN} \tag{3.3}$$

$$F1 = \frac{2PR}{P + R} \tag{3.4}$$

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.5}$$

where TP (True Positives) are those assessments where the system and human experts agree on a label assignment, FP (False Positives) are those labels assigned by the system that do not agree with the expert assignment, FN (False Negatives) are those labels that the system failed to assign as they were given by the human expert, and TN (True Negatives) are those non-assigned labels that were also discarded by the expert. The Precision tells us how well the labels are assigned by our system (the fraction of assigned labels that are correct). The Recall measures the fraction of the expert's labels found by the system. Finally F1 combine both Precision and Recall, while Accuracy takes into account all the correct results including TN (Sebastiani, 2002). For ease of comparison, we summarize the F1 scores over the different categories (positive, negative and neutral) using the Macro-averaged F1:

$$Macro - F1 = \frac{1}{\mid c \mid} \sum_{i=1}^{|c|} \frac{2P_i R_i}{P_i + R_i} \tag{3.6}$$

In the same way, we obtain the Macro-Recall and Macro-Precision as follows:

$$Macro - R = \frac{1}{\mid c \mid} \sum_{i=1}^{|c|} \frac{TP_i}{TP_i + FN_i} \tag{3.7}$$

$$Macro - P = \frac{1}{\mid c \mid} \sum_{i=1}^{|c|} \frac{TP_i}{TP_i + FP_i} \tag{3.8}$$

3.6.2 Polarity classification results

The results of the three experiments (BS, BSN and NR) with the different datasets (*Total*, *NegCue* and *RuleAffect*) are presented below. Table 3.3 show the results for *Total* set. It can been seen that the integration of the most common approach to detect the scope of negation in English tweets (BSN - (Potts, 2011b)) does not work well in the system that we use for the polarity classification of Spanish tweets. On the other hand, when the rules-based approach that we propose is included (NR), there is an improvement, but perhaps it seems that it is not so significant. However, if we observe the confusion matrix of the experiments (Table 3.4 and Table 3.5) we can see that there is a difference of about 200 tweets which have been correctly classified with the NR experiment.

Table 3.3: Results *Total* set.

	Macro-P	Macro-R	Macro-F1	Accuracy	Improvement Accuracy
BS	0.5764	0.5235	0.5486	0.6258	-
BSN	0.5705	0.5190	0.5435	0.6205	-0.885%
NR	0.5810	0.5296	0.5541	0.6308	0.80%

Note: The improvement in the Accuracy is measured over the BS method.

Table 3.4: Confusion matrix BS experiment with *Total* set.

	Predicted P	Predicted NEU	Predicted N	Recall
Real P	16,476	4,768	989	0.7410
Real NEU	511	446	348	0.3418
Real N	2,758	5,360	7,726	0.4876
Precision	0.8344	0.0422	0.8525	

Table 3.5: Confusion matrix NR experiment with *Total* set.

	Predicted P	Predicted NEU	Predicted N	Recall
Real P	16,566	4,746	921	0.7451
Real NEU	511	457	337	0.3502
Real N	2,685	5,341	7,818	0.4934
Precision	0.8383	0.0433	0.8614	

As we have mentioned earlier, in order to evaluate the rules defined we should pay attention to the tweets with negation cues (*NegCue*) and mainly to the tweets with polar tokens in the scope of negation (*RuleAffected*). Table 3.6 and Table 3.7 present the results obtained using these subsets. As was to be expected, the values of the evaluation measures are lower than using the *Total* set (Table 3.3) because these subsets contain the most problematic tweets, i.e. the tweets with negation cues that are the most difficult to classify. In addition to these tweets, the *Total* set has other tweets without negation cues that are easier to classify, meaning that precision and recall increase. However, the improvement obtained with the rule-based approach is more evident. Furthermore, according to the results, it is reasserted the fact that the method most used to determine the scope of negation in English tweets (*BSN*) does not classify better in our system than the method that we propose for Spanish tweets (*NR*). The results achieved with the *RuleAffect* subset show the evaluation of the rules that we have presented. Of course, the rules are not perfect and can be improved to increase accuracy. However, there is apparently a significant difference between *BS* and *NR* experiments because as we can see there is an improvement of 18.57% in the accuracy and 19.19% in the F1 measure. Therefore, to avoid wrong conclusions we perform a statistical analysis to check whether the rules defined for the treatment of negation actually improve classification.

Table 3.6: Results *NegCue* set.

	Macro-P	Macro-R	Macro-F1	Accuracy	Improvement Accuracy
BS	0.4861	0.4702	0.4780	0.4866	-
BSN	0.4621	0.4514	0.4567	0.4622	-5.01%
NR	0.5060	0.4936	0.4997	0.5092	4.64%

Note: The improvement in the Accuracy is measured over the BS method.

Table 3.7: Results *RuleAffect* set.

	Macro-P	Macro-R	Macro-F1	Accuracy	Improvement Accuracy
BS	0.3971	0.3949	0.3960	0.4463	-
BSN	0.4431	0.4545	0.4487	0.5026	12.61%
NR	0.4660	0.4792	0.4725	0.5292	18.57%

Note: The improvement in the Accuracy is measured over the BS method.

3.7 Analysis of results

In order to see if there is a significant difference between the proportions of tweets correctly classified using the *BS* method and the *NR* method we carry out a hypothesis test. There are different statistical tests that can be used to test the difference in the proportions of two populations depending on whether we are going to compare measurements that have been observed in separate (independent) groups or in the same group of subjects before and after an event (matched-pairs). The most commonly used tests for comparing two independent proportions are the Z-test and the Chi-square test. In the case of matched-pairs, the most frequently used tests are the Wilcoxon-signed rank test and the sign test for quantitative data, and McNemar's test for qualitative data. We have used McNemar's test because our data are qualitative and whenever possible it is better to work with data in the original form.

Formally, McNemar's test is known as a model for matched-pairs data with a binary response (McNemar, 1947; Agresti, 2007). This test is used to compare two proportions that have been observed in the same group of subjects, but at two different times (before and after an event), that is, it is used to determine if there are differences on a dichotomous dependent variable (i.e., correctly classified = {yes, no}) between two related groups (i.e., classifier = {BS method, NR method}). It attempts to compare whether there is any significant change between the two measurements.

We want to compare the proportions of opinions correctly classified before taking into account negation (BS method) and after considering it (NR method). Thus, we formulate the following one-sided hypothesis test, at the level of significance $\alpha = 0.01$, to verify if the proportion of tweets correctly classified using the NR method is greater than the proportion using the BS method:

$$\begin{aligned} H_0 &: p_1 \leq p_2 \\ H_1 &: p_1 > p_2 \end{aligned} \tag{3.9}$$

where p_1 is the proportion of tweets correctly classified taking into account negation (*NR* model) and p_2 is the proportion of successes without considering negation (*BS* model).

We can assert at the level of significance $\alpha = 0.01$ that the proportions of tweets correctly classified taking into account negation with the rule-based method is significantly greater than the proportion of success using the classifier without considering negation (p-value $= 2.6976 * 10^{-7} \approx 0$, McNemar's test, $\alpha = 0.01$).

It has been demonstrated that the method proposed to determine the scope of negation improves the polarity classification of Spanish tweets, but we also want to analyze how each of the rules has worked. *RuleAffect* is the subset in which the rules have been evaluated, so it is interesting to know the frequency of use of each of the negation cues studied (Table 3.8).

Table 3.8: Frequency of negation cues in *RuleAffect* subset.

Negation cue	Frequency
no (not)	2,517
tampoco (neither)	36
nadie (nobody)	54
jamás (never)	10
ni (nor)	248
sin (without)	195
nada (nothing)	114
nunca (never)	58
ninguno (none)	5

Table 3.9: Tweets per negation cue in the *RuleAffect* subset.

Negation cue	Total tweets
no (not)	2,086
tampoco (neither)	35
nadie (nobody)	54
jamás (never)	10
ni (nor)	181
sin (without)	182
nada (nothing)	107
nunca (never)	52
ninguno (none)	5

The most widely used negation cue is *no* at 2,517 times, which indicates that a proper treatment of it is very important. This cue appears in 2,086 tweets, approximately 90% of the tweets evaluated (Table 3.9). This accurately reflects that most of the sentences with negation in Spanish make use of this cue [11]. Moreover, this also happens in English. The most common linguistic negation cue in English is *not*, along with contractions created with it, such as *couldn't* or *isn't* (Tottie, 1993). The classification results using only the rule for the negation cue *no* (NR_onlyno) are shown in Table 3.10. Although the improvement in performance is achieved especially when the rule for the negation cue *no* is applied (which was to be expected because it is the most representative of the whole set), when all the rules are considered the improvement is greater despite being small.

Table 3.10: Results *RuleAffect* Set - Comparative *NR_onlyno* and *NR* approaches.

	Macro-P	Macro-R	Macro-F1	Accuracy	Improvement Accuracy
NR_onlyno	0.4587	0.4693	0.4639	0.5155	-
NR	0.4660	0.4792	0.4725	0.5292	2.66%%

Note: NR_onlyno is the approach NR, but only applying the rule of the cue *no*

Table 3.11 shows the percentage of tweets per negation cue correctly classified with the approach that does not take into account negation (*BS*) and with the method that we propose to determine the scope of negation (*NR*). It can be seen that the rule-based method works better than the *BS* method for most of the cues, especially in the tweets with the cues *tampoco*, *nadie*, *ni*, *nada* and *nunca*, increasing the percentage of tweets correctly classified in a 34.61 % in the best of the cases (cue *nunca*). Notwithstanding, it is noted that the rules for the cues *jamás* and *ninguno* do not work as well as we expected because the percentage of incorrectly classified tweets overcomes the percentage of the correctly classified tweets.

In order to know the reasons for the poor performance in the tweets with the negation cues *jamás* and *ninguno*, a deeper analysis was carried out. Most of the mistakes in the classification of these tweets are produced in political tweets and in tweets belonging to the neutral class. We have to take into account that many political tweets are ironic and our system does not care about this challenge of sentiment analysis. In order to see the influence of these two negation cues in the performance of the classification system, we carry out the experimentation without

[11] According to Camarero (2008), "The most common grammatical negation procedure in Spanish is the use of the adverb *not* before the verb…".

Table 3.11: % Tweets per negation cue correctly classified in the *RuleAffect* subset.

Negation cue	BS	NR
no (not)	44.10%	52.59%
tampoco (neither)	40.00%	54.29%
nadie (nobody)	40.74%	57.41%
jamás (never)	30.00%	20.00%
ni (nor)	35.36%	53.59%
sin (without)	51.65%	55.49%
nada (nothing)	34.58%	57.01%
nunca (never)	28.85%	63.46%
ninguno (none)	20.00%	20.00%

applying the rules defined in the tweets with the cues *jamás* and *ninguno* (NR_Mod). As can be noted the non-application of these rules improves the accuracy of the final system, although not by much (Table 3.12). This is due to the fact that the number of tweets with these cues represents a very small portion of the total. Therefore, it seems that the rules defined for these two negation cues are not suitable. Maybe, they need specific treatment. We will take this into account in our future research.

Table 3.12: Results *RuleAffect* Set - Comparative *NR* and *NR_mod* approaches.

	Macro-P	Macro-R	Macro-F1	Accuracy	Improvement Accuracy
NR	0.4660	0.4792	0.4725	0.5292	-
NR_Mod	0.4661	0.4797	0.4728	0.5297	0.09%

Note: NR_Mod is the approach NR, but without the application of the rules for the cues *jamás* and *ninguno*.

In relation to the other negation cues, the cases that lead to poor performance are due mainly to the presence of irony, double negation and the absence of certain polar words in the polarity lexicon used. Another reason could be the fact that when a word is in the scope of a negation cue our module always reverses its polarity, and the negation of a word does not always mean that its polarity is reversed, it can also be increased, decreased or considered as neutral. For example, “It is not perfect” is far from meaning “It is a disaster”. Thus, although the results seem to confirm that the proposed heuristic is helpful in most cases, we conduct another experiment in which the polarity of the affected part is considered as neutral (NR_Neu). Results are presented in Table 3.13.

Table 3.13: Results *RuleAffect* Set - Comparative *NR* and *NR_Neu* approaches.

	Macro-P	Macro-R	Macro-F1	Accuracy	Improvement Accuracy
NR_Neu	0.4973	0.4404	0.4727	0.4304	-
NR	0.4660	0.4792	0.4725	0.5292	22.96%

Note: NR_Neu is the approach NR, but considering the polarity of the words that are in the scope of negation as neutral.

NR_Neu heuristic increments Precision of positive and negative classes and Recall of neutral class, but Recall of positive and negative classes decreases because this approach misclassified some tweets as neutral (Table 3.14 and Table 3.15). Consequently, the Accuracy of the approach that we propose in this study (*NR*) surpasses by 22.96% the Accuracy of the *NR_Neu* heuristic. However, it seems that it would be a good idea to study in which cases the polarity of the words that are within the scope of negation should be increased, decreased, swapped or considered as neutral rather than dealing with negation as a whole.

Table 3.14: Confusion matrix NR experiment with *RuleAffect* set.

	Predicted P	Predicted NEU	Predicted N	Recall
Real P	472	115	99	0.6880
Real NEU	87	49	52	0.2606
Real N	459	283	710	0.4890
Precision	0.4637	0.1096	0.8246	

Table 3.15: Confusion matrix NR_Neu experiment with *RuleAffect* set.

	Predicted P	Predicted NEU	Predicted N	Recall
Real P	433	212	41	0.6312
Real NEU	79	71	38	0.3777
Real N	267	688	497	0.3423
Precision	0.5558	0.0731	0.8628	

There is another important fact to discuss here and it is related to the chosen parser. The performance of traditional NLP tools is lower when tweets are analyzed, as is shown in several published papers (Kong et al., 2014; Stavrianou et al., 2014). Research on English tweets is more advanced than on Spanish, so the fact that the first dependency parser for tweets written in English was presented in 2014 (Kong et al., 2014) is evidence that the task is not easy. Some studies describe the use of standard dependency parsers on tweets written in English (Jiang et al., 2011) before the development of the specific dependency parser for tweets. As far as we know, there is no available dependency parser for Spanish tweets, but this cannot stop the advance of research. Although the use of specialized parsers is better for the processing of tweets, we also support the idea that while specialized parsers are not available, a standard parser can be used. In our study, the NLP tool most used for Spanish, Freeling, has been chosen.

3.8 Conclusion

Negation is a linguistic phenomenon that can change the meaning of a sentence, so its treatment can influence positively in the performance of Natural Language Processing tasks such as sentiment analysis. In this study, we have presented a set of syntactic rules for determining the scope of negation in Spanish. We have integrated these rules into a polarity classification system of Spanish tweets and it has been demonstrated that the results obtained with them are significantly greater than those without taking into account negation. This rule-based approach has also been compared with the method most used to determine the scope of negation in English tweets, and it has been proved that the classification with our approach is better. Moreover, we have analyzed the rules defined showing the performance of them with each negation cue.

The results obtained encourage us to follow in the study of the correct treatment of negation in the context of sentiment analysis. However, one of the main problems in this area is the lack of resources. There is no a Spanish corpus annotated with negation for sentiment analysis. In order to determine the strengths and weaknesses of sentiment analysis systems that incorporate a module for negation processing, it is necessary a corpus annotated at both levels, sentiment and negation. In this way, an error analysis could be carried out to check whether the system correctly determines the negation cues and their corresponding scopes or if some of the errors are caused by the polarity classifier used. The approaches proposed so far could not be properly evaluated due to the non-existence of a corpus annotated with such information. For this reason, another one of the contributions of this doctoral thesis is the generation of an annotated corpus with negation and its scope, which we present in Chapter 4: "*SFU Review*$_{SP}$*-NEG: a Spanish corpus annotated with negation*".

Chapter 4

SFU Review$_{SP}$-NEG: a Spanish corpus annotated with negation

The availability of corpora annotated with negation is essential to training negation processing systems in any language. Currently most of these corpora have been annotated for English, but the presence of languages other than English on the Internet, such as Chinese or Spanish, is greater every day. Available corpora that contain a representation of negation can be divided into two types (Fancellu et al., 2017): i) those that represent negation in a *logical form*, using quantifiers, predicates and relations (e.g. Groningen Meaning Bank (Basile et al., 2012), DeepBank (Flickinger et al., 2012)); and, ii) those that use a *string-level*, where the negation operator and the elements (scope, event, focus) are defined as spans of text (e.g. BioScope (Vincze et al., 2008), ConanDoyle-neg (Morante & Daelemans, 2012)). The ones of interest for the objective of this thesis are corpora dealing with *string-level* negation.

Specifically, we have developed a Spanish *string-level* corpus in the review domain, due to the fact that another of the objectives of this thesis is to show the importance of proper processing of negation for Natural Language Processing tasks, such as sentiment analysis. The corpus selected for the annotation has been the Spanish part of the SFU Review corpus (Taboada et al., 2006).[1] We have chosen it because of three reasons: i) it consists of reviews that belongs to 8

[1] https://www.sfu.ca/~mtaboada/SFU_Review_Corpus.html

different domains, which implies a greater lexical richness that is of interest for the development of the negation processing system, ii) it is widely used for sentiment analysis, task on which we are going to apply the negation processing system, and iii) there is an English version of this corpus annotated with negation, the SFU Review Corpus with negation and speculation annotations (Konstantinova et al., 2012), which would allow to study the difficulty of this phenomenon in both languages.

In this chapter, we present the SFU Review$_{\text{SP}}$-NEG corpus[2] (Jiménez-Zafra et al., 2018) and the process followed for its annotation. It represents the first corpus annotated with negation in the review domain for Spanish. In first place we define and delimit the components of negation. After that, we propose a typology of negation patterns that is the one used in the annotation. Later, we present the annotation scheme and the annotation process followed, explaining also the main sources of disagreement. Finally, we describe the corpus, provide statistics and report conclusions.

4.1 Definition and delimitation of negation components

Negation is a linguistic phenomenon that is used to change the truth-value of a linguistic unit: sentence, clause, phrase or word. In our approach to the annotation of negation in the SFU Review$_{\text{SP}}$-NEG corpus, we annotated negation cues, scopes and events, but not the focus of negation. Focus is a difficult issue to solve because it is grounded on semantic and pragmatic knowledge, and its identification requires contextual information that is not always present in the available data[3].

Specifically, we focused on **negation at the syntactic level**, i.e. negation in sentences, clauses and phrases. We did not take into account lexical negation (e.g., ignorar *[to not to know]*, falta de *[lack of]*, dudar de *[to have doubts about]*) nor morphological negation, that is, words with a negation affix (e.g., desilusionado *[disappointed]*, incapaz *[incapable/unable]*). This approach

[2]First Online: 22 May 2017 https://doi.org/10.1007/s10579-017-9391-x

[3]In oral language, the focus is often marked with specific intonation (pitch) and intensity (volume).

is in accordance with the definition of negation in the Spanish grammar of the Real Academia Española (RAE):

> *"En sus múltiples manifestaciones gramaticales, la negación se considera un operador sintáctico en un sentido similar al de los cuantificadores y determinados adverbios, es decir, un elemento que condiciona (...) la referencia de otras unidades que se hallan en su ámbito de influencia."*
> *"Negation, in its multiple grammatical expressions, is considered to be a syntactic operator that is similar to quantifiers and certain adverbs, that is, it is an element that conditions the reference of the units within its scope of influence."*
>
> (Española, 2009, p. 3631)

Words expressing syntactic negation belong to different grammatical categories: adverbs (e.g. no [*no/not*], jamás [*never/ever*], nunca [*never*], tampoco [*nor/either/neither*], nada [*nothing*]); pronouns (e.g. nada [*nothing*], nadie [*nobody/no one*], ninguno [*none/nobody*], nunca [*never*]); conjunctions (e.g. ni [*nor/neither*], sino [*but*]); prepositions (e.g. sin [*without*], en vez de [*instead of*]), and indefinite determiners (e.g. ningún [*no/neither*], ninguna [*no/neither*]). It is worth noting that some words like "nada" can belong to more than one single category. In Example (37) "nada" is an adverb whereas in Example (38) it is a pronoun.

37. No confío **nada**$_{adv}$ en él.
    I do not trust him at all.

38. No hay **nada**$_{pron}$.
    There is nothing.

Regarding the **scope**, it is the part of the sentence affected by the negation cue (Vincze et al., 2008). In the Spanish grammar of the RAE (Española, 2009), there is no mention of the inclusion or exclusion of the negation cue in the scope, and with regard to the subject, it is included inside the scope when it is in a postverbal position whereas it is excluded from the scope when it is in a preverbal position. In the SFU Review$_{\text{SP}}$-NEG corpus the scope always correspond to a syntactic component, that is a phrase, a clause or a sentence. Negation cues and events are always included within the scope of negation, and the subject is included when the word directly affected by the negation is the verb of the sentence, as in ConanDoyle-neg corpus (Morante & Daelemans, 2012).

As for the **event** of negation, it is the element directly affected by the negation, usually a noun (Example 101), a verb (Example 40) or an adjective (Example 102), always within its scope, and usually the head of the phrase in which negation appears. Focus is a semantic and pragmatic concept, whereas event is a syntactically grounded one. Consequently, the event is easier to be detected. In Example (42), the event of "No" is the head of the sentence, *pienso*, and the events of the two "ni" are *contigo* and *él*.

39. Este producto tiene [**cero** fiabilidad$_{Noun}$].

    This product offers zero reliability.

40. [La asistencia técnica **no** llegó$_{Verb}$ a tiempo].

    Technical assistance did not arrive on time.

41. Un precio [**no** precisamente barato$_{Adjective}$].

    Not really a cheap price.

42. [**No**^{1} pienso$^1{}_{Verb}$ ir al concierto [**ni**^{2} contigo$^2{}_{Pronoun}$]^2 [**ni**^{3} con él$^3{}_{Pronoun}$]^1]^3.

    I'm not going to the concert with you or him.

In the case of prepositional phrases introduced by the negation cue "sin" *[without]*, the event is the head of the nominal phrase. In the case of copulas -verbs "ser" and "estar" *[to be]* -the event is the copula and the attribute. When a verb has a predicative complement, the complement is included within the event. In the case of periphrastic verbs (e.g., no acaba de salir *[doesn't quite work out]*), verbal collocates (e.g., no da problemas *[doesn't create problems]*) and light verbs with a complement (e.g., no se dio por vencido *[he/she did not give up]*) the event includes the whole complex verbal form.

The indefinite negative pronouns constitute a special case of event and scope when they appear before the verb (Example 43). The pronoun "nadie" *[nobody]* corresponds in Spanish to "ninguna persona" *[no person]*, and therefore scope and event fuse in the same form. In Example (44), the scope is "ninguna persona" and the event is *persona*, and they fuse together in "nadie" *[nobody]* (Example 43).

43. [**<u>Nadie</u>**] vino.
    Nobody came.
44. [**Ninguna** <u>persona</u>] vino.
    Nobody came.

4.2 Typology of negation patterns

One of the most important requirements for the annotation of negation is to dispose of a reliable and comprehensive typology of language-dependent negation patterns. Practically all of the existing typologies are for English and need to be adapted in order to be applied to other languages. That is why we developed our own typology of negation patterns for Spanish (Martí et al., 2016).

We have built our typology of negation expressions taking into account the basic principles contained in the standard descriptive and normative grammars of Spanish language (Demonte & Bosque, 1999; Española, 2009), while applying the simplest and most coherent possible approach -the tagset and criteria- to the annotation process. Corpora often contain expressions and constructions that do not appear in grammars of the language, which creates difficulties when attempting to annotate them. Our typology, though based on the grammars, covers all kinds of negation patterns and the types of negation patterns correspond to clear and well defined classes. This facilitates the annotation process.

All negation patterns found in the SFU Review$_{SP}$-NEG corpus belong to one of our proposed classes. We consider that this is a valid test of our typology considering the size of the corpus and the variety of the negation patterns found.

Our typology of negation patterns was built taking into account both their syntactic structure and their semantic interpretation. Taking into account their semantic interpretation, we distinguish between those patterns that express negation (Subsection 4.2.1) and those that do not express negation (Subsection 4.2.2). In terms of syntactic structure, the negation patterns can be simple or complex.

Those patterns having the semantic value of negation are labeled with the <neg> tag (Examples 45 and 46) and those that contain a cue without a semantic value of negation are labeled with <noneg> (Examples 47 and 48), <contrast> (Example 49) or <comp> (Example 50):

45. **No**$_{<neg>}$ recomiendo el libro.

    I do not recommend the book.

46. **No**$_{<neg>}$ me pide el PIN.

    They are not asking me for the PIN.

47. **Nada más**$_{<noneg>}$ darle al contacto, se encendieron los pilotos.

    As soon as you turn the key, the pilots lights come on.

48. ¿**No**$_{<noneg>}$ prodrían hacerte otra pregunta?

    Couldn't they ask you another question?

49. BMW **no**$^{1}{}_{<contrast>}$ suele poner **más_que**$^{1}{}_{<contrast>}$ lo que considera necesario.

    BMW does not usually include more than they consider to be necessary.

50. **No**$^{1}{}_{<comp>}$ me gusta **tanto como**$^{1}{}_{<comp>}$ lo otro.

    I don't like it as much as the other one.

In Example (47), "nada más" means *[as soon as]* or *[only by]*, therefore it does not express negation. In Example (48), the cue does not have any semantic value, it is just a marker of politeness, and the meaning of the sentence is the same with or without the cue. Finally, Examples (49) and (50) show a contrasting and a comparative structure, respectively, where the cues do not negate.

4.2.1 Negation cues expressing negation

In this section, we present all those patterns that express negation in Spanish and which can be simple or complex. We also include in this section those expressions that do not contain a negative marker, but which express negation.

4.2.1.1 Simple negation

Simple negations are those negation patterns that only include one single token. This token always precedes the event and can be an adverb (e.g., no *[no/not]*, jamás *[never/ever]*, apenas *[hardly ever]*, nunca *[never]*) (Examples 51 and 52), a pronoun preceding the verb (nadie *[nobody/no one]*, nada *[nothing]*) (Example 53), or a preposition (sin *[without]*) (Example 54).

51. (...) para conductores que **apenas**$_{Adverb}$ tocan el coche.

    (...) for drivers that hardly ever drive the car.

52. Una huella que tal vez **nunca**$_{Adverb}$ se borre.

    A footprint that may never be erased.

53. **Nadie**$_{Pronoun}$ oía el dichoso ruído.

    Nobody heard the annoying noise.

54. Ha llegado a lavar un edredón nórdico **sin**$_{Preposition}$ problema.

    I've even washed a nordic quilt without trouble.

We also include in this category coordinated simple negative sentences (Examples 55 and 56).

55. **Ni**^{1} puedo desear más **ni**^{2} puedo contentarme con menos.

    I couldn't ask for more or be satisfied with less.

56. El aire acondicionado **ni**^{1} enfría **ni**^{2} calienta.

    The air conditioning doesn't heat or cool.

4.2.1.2 Complex negation

Complex negations are those cases in which negation is expressed by means of two or more tokens. These negation cues can be continuous (Example 57) or discontinuous (Examples 58 and 59). In complex negation, usually the first token is the one that carries out the main negation function whereas the second one may also express negation by reinforcing the negation content (Example 58) or may modulate the value of negation (Example 59).

57. **Casi no** llega.

    He almost didn't make it.

58. **No**^{1} vino **nunca**^{1}.

    He/she never came.

59. **No**^{1} hace **mucho**^{1} ruido.

    It isn't very noisy.

Below, we describe these two types in more detail.

a) Negation reinforcement

In Spanish, negation cues are often reinforced by a second negation cue (Examples 60a-63a). These complex structures can always be paraphrased with a simple negation structure in which the second negation cue precedes the verb (Examples 60b-63b):

60. (a) El final del libro **no**^{1} te aporta **nada**^{1}.

    (b) El final del libro **nada** te aporta.

    The end of the book doesnot add anything.

61. (a) La tecnología **no**^{1} falla **nunca**^{2}.

    (b) La tecnología **nunca** falla.

    Technology never fails.

62. (a) Allí **no**^{1} me esperaba **nadie**^{1}.

    (b) **Nadie** me esperaba allí.

    Nobody was waiting for me there.

63. (a) Como si **no**^{1} hubiera existido **jamás**^{1}.

    (b) Como si **jamás** hubieran existido.

    As if they had never even existed.

Two coordinated negative structures within the same phrase are also considered to be reinforcement because the repetition of the negation cue in the coordinated sentence (Examples 64, 65 and 66) increases the negative value of the structure.

64. No comió **ni**^{1} pan **ni**^{1} vino.

    He ate neither bread nor wine.

65. No me sentí **ni**^{1} libre **ni**^{1} poderoso.

    I felt neither free nor powerful.

66. **Sin**^{1} agua **ni**^{1} comida.

    Without water or food.

b) Negation modifiers

Negation, like many other linguistic phenomena, is not categorical (yes/no): it is a matter of degree. We grouped under the name of modifiers all the different mechanisms used for modifying the degree of negation. These modifiers can increase (Examples 67 and 68) or diminish (Examples 69 and 70) the degree of negation:

67. **No**^{1} me gustó **en absoluto**^{1}.

    I didn't like it at all.

68. **No**[1] es **nada**[1] recomendable.[4]

    It isn't recommendable at all.

69. Mi experiencia **no**[1] fue **muy**[1] buena.

    My experience was not very good.

70. **No**[1] ha quedado **demasiado**[1] claro.

    It isn't very clear.

We annotate the modifiers with the labels <increment> or <reduction> respectively. In the previous cases, the negation cue appears in the first position, but it is also possible that the negation cue appears in the second position preceded by the modifier. In Example (71) the negation cue "no" *[not]* appears in the second position, preceded by the modifier "más" *[more]*.

71. **Más**[1] equivocado **no**[1] pude estar.

    I couldn't have been more mistaken.

4.2.1.3 Lexicalized negation

Lexicalized negations are complex constructions that express negation in specific contexts. They may include a negative marker (e.g. falta de *[absence of]* in Example 72) or not (Example 73). In Example (74), the three expressions ("en la vida" *[in life]*, "en toda mi vida" *[during all my life]* and "en su vida" *[in his life]*) have literal meanings, whereas in Example (73) the same expressions have negative meanings (equivalent to "nunca" *[never]*) even though they do not contain a negative marker.

72. **A falta de** un coche en condiciones.

    In the absence of a car in good conditions.

73. (a) **En la vida** he hecho una reserva con tanta antelación.

[4] It is worth noting that *nada* has been interpreted as an adverb, with the meaning of 'not at all', but it could also be interpreted as a pronoun with the meaning of 'nothing'. Only the context can disambiguate it.

(b) **En toda mi vida** he hecho una reserva con tanta antelación.

(c) **En su vida** ha hecho una reserva con tanta antelación.

Never in my life, have I reserved so far in advance.

I have never reserved so far in advance.

74. (a) **En la vida** hay que tener paciencia.

In life, you have to be patient.

(b) **En toda mi vida** de estudiante trabajé duro.

During all my life as student I worked hard.

(c) Se mantiene activo **en su vida** diaria.

He remains active in his daily life.

4.2.2 Negation cues not expressing negation

In Spanish, there are linguistic constructions that contain negation cues but do not express negation. In these cases, the negation cues can have a rhetorical value; be part of an idiom; or appear in contrastive or comparative constructions.

4.2.2.1 Negation cues with a rhetorical value

Negation cues can appear in interrogative sentences as the one of the Example (75) with an emphatic value. They can also be part of an affirmative sentence as in Example (76), in which they have an expletive value, that is, the presence or absence of the cue "no" *[not]* does not change the meaning of the sentence. We annotated all these cases with the value "noneg".

75. El coche lo compré para viajar, **no**?

I bought this car for travelling, didn't I?

76. No pienso irme hasta que **no** vengas.

I'm not leaving until he/she comes (literally doesn't come).

4.2.2.2 Idioms containing negation cues

In Spanish, we can find idioms (Example 77) and lexicalized constructions (Example 78) containing negation cues that do not express negation. Both types of constructions are multiword expressions with a single meaning. We annotated all these cases with the value “noneg”.

77. (a) **Ni corta ni perezosa** solté algo más de 600 euros.

    Without thinking twice I coughed up more than €600.

    (b) Esta mujer es **ni más ni menos** que madame Bertholdi.

    This woman is no more, no less than Madame Bertholdi.

    (c) Personajes secundarios que pasan **sin pena ni gloria**.

    Secondary characters who pass unnoticed.

78. (a) **Hasta que no** coloca la ropa, no centrifuga.

    Until you put the clothes in the spin-dryer doesn’t start (literally “until you don’t put in”).

    (b) Un coche que sorprendió **nada más** salir al Mercado.

    A car that surprised everyone as soon as (literally “nothing more”) it appeared on the market.

    (c) **No veas** los pifostios que se montaban.

    Don’t miss the mess that resulted.

    (d) **No hay más que** llamarle a su móvil.

    All you have to do is phone him on your mobile.

    (e) Era una cosa **más que nada** para reírse.

    It was, more than anything, a laughing matter.

    (f) Eso **no hace más que** denotar su estado de pobreza absoluta.

    This only underlines his state of poverty.

4.2.2.3 Negation cues in contrastive constructions

In Spanish, contrastive constructions comparing two or more elements are used to counterpose different assessments; either to correct something previously said (Example 79a) or to introduce new information (Example 79b). In some cases, they can express obligation (Example 80).[5] All these structures are annotated with the value "contrast".

79. (a) **No**[1] vinieron 2 soldados, **sino**[1] 6.

    6 soldiers came, not 2.

    (b) Es un coche pensado **no solo**[1] para su uso, **sino**[1] para su disfrute.

    This car is designed not only to be used, but to be enjoyed.

80. (a) BMW **no**[1] suele poner **más que**[1] lo que considera necesario.

    BMW does not usually include more than they consider to be necessary.

    (b) **No**[1] veo otra salida **que**[1] pedirle otra lavadora.

    I see no other option than to order another washing machine.

    (c) **No**[1] existe otra forma de llegar al aeropuerto **que no sea**[1] en taxi.

    There is no other way to get to the airport than by taxi.

4.2.2.4 Negation cues in comparative constructions

In Spanish, comparative constructions can include a negation cue that express relative negation with respect to another element. In our annotation scheme, these structures are treated as comparative constructions containing a negation cue. They are always discontinuous constructions (Example 81).

81. (a) **No**[1] me gusta **tanto como**[1] lo otro.

    I don't like it as much as the other one.

[5] Example (80a) can be paraphrased as "BMV suele poner sólo lo que considera necesario" *[BMW only includes what they consider to be necessary]*. Example (80b) can be paraphrased as "La única salida es pedirle otra lavadora" *[The only solution is to order for another washing machine]*. Sentence (80c) can be paraphrased as "La única forma de llegar al aeropuerto es en taxi" *[The only way of getting to the airport is by car]*.

(b) El ambiente de este local es agradable pero **no**[1] (verbo elidido) **tanto como**[1] el del otro.

The atmosphere in this place is pleasant but not as much as in the other one.

(c) El motor **no**[1] es **todo lo**[1] potente **que**[1] debería.

The engine is not quite powerful as it should be.

In all of these examples, the second element in the comparison refers to a limit or parameter in relation to which a property of something/somebody is compared. We annotate these structures with the value "comp".

Figure 4.1 shows the typology of negation patterns that we proposed. Note that [+Negation] stands for cues that express negation whereas [-Negation] stands for cues not expressing negation.

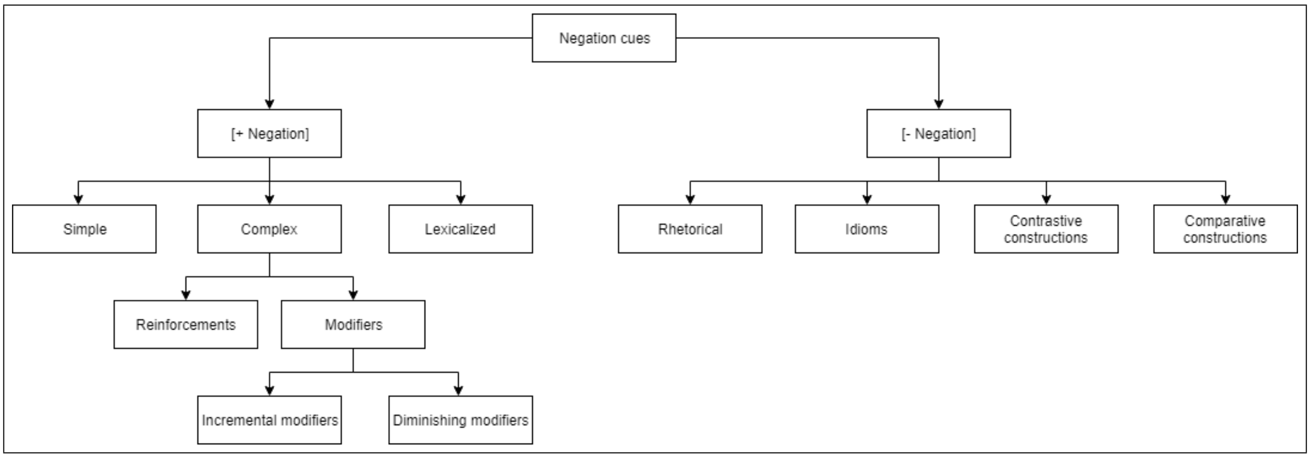

Figure 4.1: Typology of negation patterns.

4.3 Annotation scheme

In this section, it is described the annotation scheme adopted for the annotation of negation in the SFU Review$_{SP}$-NEG corpus. The general annotation scheme followed is shown in Figure 4.2 and the labels used for the annotation are described below.

<review polarity="positive/negative" >. The label *review* contains all the information about the review. It has the attribute *polarity* that describes the polarity of the review, which

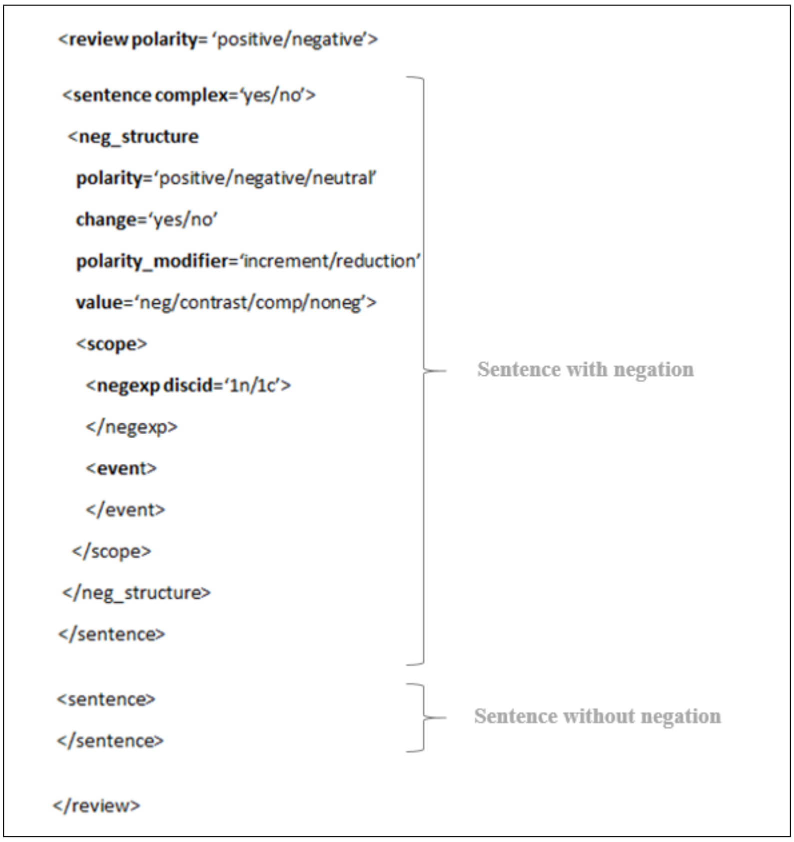

Figure 4.2: General annotation scheme.

can be positive or negative, according to the value assigned to it in the SFU Review Corpus (Taboada et al., 2006).

<sentence> or **<sentence complex="yes/no">.** The label *sentence* corresponds to a complete sentence, a phrase, a fragment or a clause with a self-contained meaning. If negation is present in the sentence, this label will have the attribute *complex* associated. It can take the value "no", if the sentence only has one negation structure (Example 82), or "yes", if there are more than one negation structure in the sentence (Examples 83 and 84). Complex structures (<**sentence complex="yes"**>) can be embedded or non-embedded. Embedded structures (Example 83) are those in which one negative structure is part of another negative structure in the same <sentence> node. Non-embedded structures are those in which two or more negative structures appear independently in the same <sentence> node (Example 84).

82. <sentence complex="no">

    El anterior coche se paró a la media hora de comprarlo <**neg_structure**> porque **no** le habían quitado el precinto de seguridad </**neg_structure**>

</sentence>

Our previous car stopped half an hour after we bought it because they had not removed the security seal.

83. <sentence complex="yes">

<**neg_structure**>$_1$ **no**^{1} quería pasarme un día entero en el aeropuerto

<**neg_structure**>$_2$ **sin**^{2} poder descansar </**neg_structure**>$_2$

<**neg_structure**>$_1$

</sentence>

I did not want to spend the whole day at the airport without resting.

84. <sentence complex="yes">

<**neg_structure**>$_1$ para que **no**^{1} les entre polvo </**neg_structure**>$_1$ o

<**neg_structure**>$_2$ para que **no**^{2} se oxiden </**neg_structure**>$_2$

</sentence>

so that dust does not get in or so that they do not rust.

<neg_structure polarity="positive/negative/neutral" change="yes/no" polarity_modifier='increment/reduction" value="neg/contrast/comp/noneg">. The label *neg_structure* represents a syntactic structure which contains a negation cue. It has four attributes:

- The attribute *polarity* shows the polarity of the negation structure, which can be one of the following values: "positive" (Example 85), "negative" (Example 86), or "neutral" (Example 87).

85. <neg_structure **polarity="positive"**> No vas a tener problemas </neg_structure>

You will have no trouble.

86. <neg_structure **polarity="negative"**> Segundas partes nunca fueron buenas </neg_structure>

    Sequels are never any good.

87. <neg_structure **polarity="neutral"**> El realismo de Flaubert no busca la precisión histórica </neg_structure>

    Flaubert's realism does not aspire to historical accuracy.

- The attribute *change* indicates whether the polarity or the meaning of the negation structure has been modified or not because of the negation. This attribute can take one of the following values: "yes", if negation modifies the polarity of the structure (Example 88), or "no", if negation does not modify the polarity of the structure (Example 89).

88. <neg_structure polarity="positive" **change="yes"**> La calidad del sonido no es mala </neg_structure>

    The sound quality is not bad.

89. <neg_structure polarity="negative" **change="no"**> Ni siquiera tengo carnet </neg_structure>

    I do not even have a card.

- The attribute *polarity_modifier* states whether there is an element in the negation structure that nuances its polarity. It has two possible values: "increment", if there is an increase in the intensity of the polarity (Example 90), or "reduction", if there is a reduction in the intensity of the polarity (Example 91).

90. <neg_structure polarity="positive" **polarity_modifier="increment"**> **No** me arrepiento **para nada** </neg_structure>

    *I do **not** regret (it) **at all**.*

91. <neg_structure polarity="negative" **polarity_modifier="reduction"**> **No** lo he utilizado **mucho** </neg_structure>

    *I have **not** used it **much**.*

- The attribute *value* shows the type of the negation structure. It can take one of these four values: "neg", if the structure expresses negation (Example 92), "contrast", if the structure expresses contrast or opposition between terms (Example 93), "comp", if the structure expresses a comparison or inequality between terms (Example 94), or "noneg", if the structure contains a negation cue but which does not negate (Example 95).

92. <neg_structure **value="neg"**> Las habitaciones no están cuidadas </neg_structure>

    The rooms are not maintained.

93. <neg_structure **value="contrast"**> Me he comprado el móvil no por las prestaciones que posee sino por el diseño </neg_structure>

    I bought the mobile not for the specs but for the design it has.

94. <neg_structure **value="comp"**> Su exterior no me gusta tanto como los de otras marcas </neg_structure>

    I don't like the outside of it as much as those of other brands.

95. <neg_structure **value="noneg"**> El coche lo compré para viajar, ¿no? </neg_structure>

    I bought this car for travelling, didn't it?

<scope>. The label *scope* delimits the part of the negation structure that is within the scope of negation (Example 96). It includes both the negation cue (<negexp>) and the event (<event>).

96. <sentence complex="no"> Y el problema es <neg_structure polarity="negative" value="neg" change="yes"> que <**scope**>no saben arreglarlo </**scope**> </neg_structure>

    And the problem is, they don't know how to fix it.

<negexp>. This label delimits the words that have the function of negation cues. It can have associated the attribute *discid* if the negation expression has more than one element and they are discontinuous. The value of this attribute is a number that represents the numerical order of the discontinuous expression in the negation structure (<neg_structure>) and a letter, 'n' or 'c', which indicates the nucleus and complement of the negation expression respectively (Example 97). In the case of coordinated negations, the <discid> label identifies the different coordinated negative elements (Example 98). The label <discid> is also used in discontinuous negation structures expressing contrast (Example 99) or comparison (Example 100).

97. El coche <negexp **discid="1n"**>**no**</neg_exp> <event>frena</event> <negexp **discid="1c"**>**en absoluto**</negexp>

    (It) does ***not*** *brake* ***at all****.*

98. (a) Permiten el paso <negexp **discid="1n"**>**sin**</neg_exp> <event> grandes contorsiones <negexp **discid="1c"**>**ni**</negexp> aspavientos </event>

    They allow one to pass ***without*** *major contortions and* ***without*** *fuss.*

    (b) <negexp **discid="1n"**>**No**</neg_exp> <event>es <negexp **discid="1c"**> **ni**</negexp> muy pesado <negexp **discid="1c"**>**ni**</negexp> muy ligero</event>

    It is ***neither*** *too heavy* ***nor*** *too light.*

99. <neg_structure value=**“contrast”** polarity=“neutral” change=“no”> La segunda parte del libro, <negexp **discid=“1c”**>**lejos de**</neg_exp> mantener mi entusiasmo <negexp **discid=“1n”**>**más bien**</negexp> lo sepultó

    </neg_structure>

    The second part of the book, ***far from*** *maintaining my enthusiasm, killed it off* ***(instead)****.*

100. <neg_structure value=**“comp”** polarity=“positive”

    polarity_modifier=“reduction”> Su exterior <negexp **discid=“1n”**>

    no</neg_exp> me gusta <negexp **discid=“1c”**>**tanto como**</negexp> el de otras marcas </neg_structure>

    I do ***not*** *like the outside of it* ***as much as*** *that of other brands.*

<event>. The label *event* denotes the word(s) directly affected by the negation. It is usually a subset of the scope, although in some cases it can match with the scope. It is generally a noun (Example 101), an adjective (Example 102) or a verb. Verbs can be a simple (Example 103) or complex verbal form, such as a passive verbal form (Example 104), a periphrastic verbal form (Example 105) or a light verb (Example 106). In the case of pronominal verbs (Example 107) or verbs with a pronoun in the passive voice (Example 108), the pronouns are also included inside the <event>, because they are part of the verbal form. The complements of copulas (Example 109) and predicative complements (Example 110) are also included inside the <event>, because the semantic content is in the complement (they are basically adjectives). Finally, the elliptical <event> is identified with the empty symbol set (∅) and manually tagged with the attribute <elliptic>. In Example (111), the antecedent of the elliptical event is “tiene un coche de segunda mano machacao” *[you have a beaten-up second-hand car]*.

101. <negexp> **Cero** </negexp> <**event**> **fiabilidad** </**event**>

    Zero reliability.

102. <negexp> **Nada** </negexp> <**event**> **bueno** </**event**>

    Not good at all.

103. <negexp> **No** </negexp> <**event**> **hablo** </**event**> de accesorios

    I am not speaking about accesories.

104. Mis peticiones <negexp> **no** </negexp> <**event**> **fueron atendidas** </**event**>

    My requests were not addressed.

105. <negexp> **No** </negexp> <**event**> **deseo regresar** </**event**> a ese hotel

    I do not want to go back to that hotel.

106. El modelo <negexp> **no** </negexp> <**event**> **da problemas** </**event**>

    The model does not create problems.

107. por lo que <negexp> **no** </negexp> <**event**> <u>te</u> **mareas** </**event**>

    so you do not feel sick.

108. aunque ya <negexp> **no** </negexp> <**event**> <u>se</u> **fabrica** </**event**>

    although it is no longer being made.

109. <negexp> **No** </negexp> <**event**> **es pesado** </**event**>

    It is not heavy.

110. <negexp> **No** </negexp> <**event**> **resulta agradable** </**event**>

   It is not pleasant.

111. Os preguntaréis: que <u>tiene un coche de segunda mano machacao</u>?

   <neg_structure> Pues <negexp> **no** </negexp> <event>

   ∅ <**elliptic="yes"**> </event> señor </neg_structure>

   You are asking yourselves: What is so special about this beaten-up second-hand car? Well, no sir, I am not.

4.4 Annotation process

In this section, we describe the procedure followed for the annotation of the SFU Review$_{\mathrm{SP}}$-NEG corpus. After the creation of initial guidelines[6] based on what had been done in previous works, the corpus was annotated by 4 annotators: two trained annotators who carried out the annotation task and two senior researchers with in-depth experience in corpus annotation who supervised the whole process.

Firstly, a training phase was carried out in which 50 files were annotated in parallel by the trained annotators in order to refine the annotation guidelines. Then, we discussed disagreements and updated the guidelines taking into account the resulting criteria. After that, a further 50 files were annotated individually by the same annotators to measure inter-annotator agreement with the aim of detecting and resolving problematic cases. A total of 528 negation structures were annotated and 49 cases of disagreement were found. An observed agreement of 90.72% corresponding to a kappa-score of 0.74 was observed in the inter-annotator agreement test. We then proceeded to annotate the whole corpus. We present this process in more detail below.

[6]Final guidelines have been presented in the previous sections: definition and delimitation of negation components, definition of a typology of negation patterns for Spanish, and annotation scheme.

Before negation was annotated, the corpus was automatically tokenized, PoS tagged and lemmatized using the Freeling library[7]. The corpus was then manually annotated with negation and polarity information applying the following procedure in all the steps of the annotation process:

- First, we selected all the structures (i.e., a sentence, a clause or a phrase) containing a negation cue and we labelled them as <neg_structure>.
- Then, we selected the fragment of text corresponding to <neg_structure> and we identified the negation cues (tagged as <negexp>), their scope (labelled as <scope>) and the word(s) corresponding to the event (labelled as <event>).
- Finally, we determined whether the polarity or sense of the <neg_structure> was affected by the negation cue (i.e., if there was a change in the polarity or an increment or reduction of its value), taking into account intensifiers and diminishers. In this step, we also annotated the negation cues that do nor negate (i.e., when the value was "noneg", "constrast" or "comp").

The annotation process was carried out in three steps: a first step for the training of the annotators and the creation of a first version of the annotation guide; a second step devoted to testing the validity of the annotation guide and to conducting an inter-annotator agreement test; and a final step in which the corpus was annotated definitively.

The training phase was carried out in two months. We elaborated a working annotation guide reflecting the knowledge about negation obtained from the literature. In this phase, two annotators tagged 50 files obtained randomly from the corpus. The files were annotated in parallel, all problems were discussed weekly, and the resulting agreements were introduced into the annotation guide. After two months, we stopped this training phase, when problems and disagreements had practically disappeared.

After the training phase, the 50 files corresponding to cell phone reviews (25 positive reviews and 25 negative reviews) were individually annotated by the two trained annotators. Then, we

[7] http://nlp.lsi.upc.edu/freeling/node/30

conducted an inter-annotator agreement test using these files, in which a total of 528 negation structures were annotated. These files were compared manually and we found 49 cases of disagreement, which were analyzed in order to detect misunderstandings of the annotation guidelines and problems not detected in the training phase. We discussed them, proposed solutions and, as a result, the guidelines were updated.

Table 4.1: Inter-annotator agreement.

Attributes	% Agreement	Kappa	#Total agreement	#Total disagreement
<scope>	96.97%	0.94	528	16
<event>	97.16%	0.95	528	15
<neg_structure> attributes	98.11%	0.97	528	10
<negexp>	98.48%	0.97	528	8
All	**97.68%**	**0.96**	**528**	**49**

In the inter-annotator agreement test in terms of Kappa coefficient, we obtained 0.97 for negation cues, 0.95 for negated events, 0.94 for scopes and 0.97 for the attributes of the negation structures. This information is shown in detail in Table 4.1. It is worth mentioning that we treated partial matches in scope as disagreement.

During the third phase, in order to ensure the consistency and quality of the annotation, the team met once a week to discuss the problems arising during the annotation process and to resolve doubts and specific cases.

Annotation was performed using the AnCoraPipe annotation tool[8] to facilitate the annotation task and to minimize errors in the annotation process. The texts annotated were XML documents with UTF-8 encoding. Next, we present an example of annotation.

4.4.1 An example of annotation

Figure 4.3 is an example extracted from a review of hotels domain. It corresponds to the sentence “Las habitaciones son pequeñas, casi no tienen camas de matrimonio, ni tienen terraza.”

[8] http://annotation.exmaralda.org/index.php/AnCoraPipe - User’s guide: http://clic.ub.edu/ca/ancorapipe

[The rooms are small, they have almost no double beds, nor do they have a terrace.].[9]

```xml
<sentence complex="yes">
    <d gen="f" lem="el" name="d" num="p" pos="da0fp0" postype="article" wd="Las"/>
    <n gen="f" lem="habitación" name="n" num="p" pos="ncfp000" postype="common" wd="habitaciones"/>
    <v lem="ser" mood="indicative" name="v" num="p" person="3" pos="vsip3p0" postype="semiauxiliary" tense="present" wd="son"/>
    <a gen="f" lem="pequeño" name="a" num="p" pos="aq0fp0" postype="qualificative" wd="pequeñas"/>
    <f lem="," name="f" pos="fc" punct="comma" wd=","/>
    <neg_structure polarity="negative" polarity_modifier="reduction" value="neg">
      <scope>
        <negexp>
          <r lem="casi" name="r" pos="rg" wd="casi"/>
          <r lem="no" name="r" pos="rn" postype="negative" wd="no"/>
        </negexp>
        <event>
          <v lem="tener" mood="indicative" name="v" num="p" person="3" pos="vmip3p0" postype="main" tense="present" wd="tienen"/>
        </event>
        <n gen="f" lem="cama" name="n" num="p" pos="ncfp000" postype="common" wd="camas"/>
        <s complex="no" lem="de" name="s" pos="sps00" postype="preposition" wd="de"/>
        <n gen="m" lem="matrimonio" name="n" num="s" pos="ncms000" postype="common" wd="matrimonio"/>
      </scope>
    </neg_structure>
    <f lem="," name="f" pos="fc" punct="comma" wd=","/>
    <neg_structure polarity="negative" change="yes" value="neg">
      <scope>
        <negexp>
          <c lem="ni" name="c" pos="cc" postype="coordinating" wd="ni"/>
        </negexp>
        <event>
          <v lem="tener" mood="indicative" name="v" num="p" person="3" pos="vmip3p0" postype="main" tense="present" wd="tienen"/>
        </event>
        <n gen="f" lem="terraza" name="n" num="s" pos="ncfs000" postype="common" wd="terraza"/>
      </scope>
    </neg_structure>
    <f lem="." name="f" pos="fp" punct="period" wd="."/>
</sentence>
```

Figure 4.3: Example of annotation. Sentence: "Las habitaciones son pequeñas, casi no tienen camas de matrimonio, ni tienen terraza." *[The rooms are small, they have almost no double beds, nor do they have a terrace.]*

In the SFU Review$_{SP}$-NEG corpus only the structures that contain at least one negation cue are manually annotated with negation and polarity information. For this reason, the first part of the sentence exemplified in Figure 4.3 ("Las habitaciones son pequeñas" *[The rooms are small]*) is only morphologically annotated: each token (wd) is annotated with information about its PoS (pos) and lemma (lem). Each type of word has a different tag related to its PoS[10] and tags related to different morphological features according to its PoS, such as subtype of PoS (postype)[11], number (num), gender (gen), tense (tense) and mood (mood), among others[12].

The <sentence complex="yes"> attribute indicates that the sentence contains more than one negation structure <neg_structure> ("casi no tienen camas de matrimonio" *[they have almost no double beds]* and "ni tienen terraza" *[nor do they have a terrace]*) and each of them contains a negation cue, that is, a <negexp> ("casi no" *[almost no]* and "ni" *[nor]*, respectively).

[9]Sentence extracted from the review: no_2_20.txt – Domain: hotels - SFU Review$_{SP}$-NEG corpus.
[10]d=determinant, n=noun, v=verb, a=adjective, r=adverb, c=conjunction, s=preposition, f=punctuation mark.
[11]For instance, <postype="article"> indicates that the determiner is an article and <complex="no"> indicates that the preposition is not complex.
[12]For a complete description of the morphological tags, see http://clic.ub.edu/corpus/ancora-documentacio.

The <polarity="negative"> attribute associated to <neg_structure> shows that these structures represent a negative opinion and <value="neg"> indicates that both structures contain negation cues that negate. If we remember, there are negation cues that do not negate or express contrast or comparison. The value "reduction" of the attribute <polarity_modifier> states that negation diminishes the polarity value, while <change="yes"> indicates that the polarity of the structure is changed by the negation cue.

Finally, the <scope> attribute delimits the words of the negation structure within the scope of negation, including the negation cue <negexp> and the event <event>. In this example, the scope of the first negation structure is "casi no tienen camas de matrimonio" *[they have almost no double beds]*, whereas the negation cue is "casi no" *[almost no]* and its event is "tienen" *[have]*. In the second negation structure, the scope is "ni tienen terraza" *[nor do they have a terrace]*, the negation cue is "ni" *[nor]* and the event is "tienen" *[have]*.

4.5 Problematic cases

Two types of annotations problems should be distinguished concerning negation (Jiménez-Zafra et al., 2016): a) those that are related to the lack of agreement between the annotators, since what it is being annotated is complex: especially the scope, but also the event, and the discontinuities; and b) the problems arising from how the negation pattern is interpreted. These cases occur in constructions that are at the limit of what can be considered negation. They are semantic problems, i.e., problems involved in interpreting these constructions. In our typology, these cases mainly correspond to negation patterns in comparative and contrastive constructions.

4.5.1 Disagreement cases

As it has been mentioned before, after the training phase, 50 files were individually annotated by the two trained annotators in order to measure inter-annotator agreement with the aim of detecting and resolving problematic cases. A total of 528 negative structures were annotated and 49 cases of disagreement were found. The main sources of disagreement are presented in Table 4.2.

Table 4.2: Disagreement cases.

Type of disagreement	#Total	% disagreement in 528 <neg_structure>	% disagreement of 49 disagreement elements
<scope> boundary	16	3.03%	32.65%
<event> boundary	15	2.84%	30.61%
<neg_structure> attributes	10	1.89%	20.41%
Discontinuous elements	8	1.52%	16.33%
Disagreements (total)	**49**	**9.28%**	**100.00%**

Most of the problematic cases (63.26%) were related to the scope of the negation and the event, though disagreements related to the value of the attributes of the <neg_structure> label and to discontinuities were also observed. Below, we describe these cases with a representative example[13]:

- Disagreements related to the scope of negation: 16 disagreements were due to the non-inclusion of the relative pronoun within the scope (Example 112a). We decided to include the relative pronoun (the subject of the relative clause) in the scope, therefore in the SFU Review$_{SP}$-NEG corpus the subject is always included within the scope when the word directly affected by the negation is the verb of the sentence (Example 112b):

112. (a) Una cámara de fotos **que** <**scope**> no es una maravilla </**scope**>

  (b) Una cámara de fotos <**scope**> **que** no es una maravilla </**scope**>

  A photo camera that is not so fantastic.

- Disagreements related to the event were mainly due to the treatment of verbal forms: pronominal verbs and light verbs. We observed a total of 15 cases. The problem with the pronominal verbs was the non-inclusion of the pronoun inside the event (Example

[13]For all cases, the annotation used in the second example (labeled with letter *b*) was selected. Disagreements were discussed by all the annotators and solutions were proposed by the senior researchers.

113a). In this case, we opted to include the pronoun inside the event (Example 113b), since it is part of the verb. On the other hand, the problem with the light verbs arose from the incorrect identification of the lexicalized arguments. In Example (114a) the argument “una rallada” *[a scratch]* was incorrectly treated as a lexicalized form, whereas in Example (115a) the opposite is the case: “tan mal” *[so badly]* is part of the verbal form (the complete verbal form should be: “dejar (tan) mal” *[leave someone (so) badly]*).

113. (a) <negexp> No </negexp> <**event**> he podido resistir </**event**> **me**

(b) <negexp> No </negexp> <**event**> he podido resistir **me** </**event**>

I could not resist myself.

114. (a) <negexp discid="1n"> No </negexp> <**event**> tenía <negexp discid="1c"> ni </negexp> **una rallada** </**event**>

(b) <negexp discid="1n"> No </negexp> <**event**> tenía </**event**> <negexp discid="1c"> ni </negexp> una rallada

It did not have a single scratch.

115. (a) <negexp> No </negexp> lo <**event**> dejaré </**event**> **tan mal**

(b) <negexp discid="1n"> No </negexp> lo <**event**> dejaré <negexp discid="1c"> **tan** </negexp> **mal** </**event**>

I will not leave him so badly.

- 10 disagreements were found in the value of the attributes of the <neg_structure> label. Most of them were related to the attributes *polarity* and *value*. For instance, in Example (116a) the negation structure was annotated as if it expressed negation (value="neg"), whereas the correct value should be "contrast" (Example 116b). In Example (117a), the annotator forgot to assign the value of the attribute *value* to the negation structure.

116. (a) Los motorolas a mí <**neg_structure value= "neg" polarity="negative"**> no hacen más que darme problemas <**neg_structure**>

     (b) Los motorolas a mí <**neg_structure value= "contrast" polarity="negative"**> no hacen más que darme problemas <**neg_structure**>
     Motorolas (devices) have not given me anything but trouble.

117. (a) <**neg_structure value=**> no me puedo mover <**neg_structure**>

     (b) <**neg_structure value="neg"**> no me puedo mover <**neg_structure**>
     I can not move (about).

- 8 disagreements related to discontinuities were due to the non-identification of intensifiers (Example 118) and diminishers (Example 119). In both of the following examples, the annotator failed to identify the discontinuous negation cue, the intensifier "para nada" *[at all]* and the diminisher "del todo" *[completely]* were not annotated.

118. (a) <negexp> no </negexp> me <event> extraña< /event> **para nada** los problemas que tiene

     (b) <**negexp discid="1n"**> no </**negexp**> me <event> extraña< /event> <**negexp discid="1c"**> **para nada** </**negexp**> los problemas que tiene
     'I am not surprised at all by the problems he is having.'

119. (a) <negexp> no </negexp> <event> estaba **del todo** acertado < /event>

     (b) <**negexp discid="1n"**> no </**negexp**> <event> estaba <**negexp discid="1c"**> **del todo** </**negexp**> acertado < /event>
     'It was not completely right.'

4.5.2 Semantic interpretation of negation patterns

The cases that generated the greatest controversy during the annotation process were those related to the interpretation of the negation patterns. They are borderline cases in which it is difficult to determine whether negation patterns express negation or not. These cases are related to comparative constructions and contrastive constructions:

4.5.2.1 Comparative constructions

In the case of comparative constructions, the negation simply places an entity below or above another entity on a scale. What is negated is the predicate expressing somebody's beliefs. In Example (120), what is negated is the predicate "imaginaba" *[imagined]*. In this type of constructions we decided that there is no negation, strictly speaking, and we annotated them with the value "comp" for "comparative". The example can be paraphrased as "Me lo imaginaba más grande" *[I imagined it bigger]* or "Es más pequeño de lo que me imaginaba" *[It is smaller than I imagined]*. In both cases no negation is present.

120. **No** es **tan** grande **como** me lo imaginaba.
    *It is **not as** big **as** I imagined.*

Many of these cases are examples of what is called "downward entailment operators", which are controversial and closely related to negation, but are not featured in this version of the corpus.

4.5.2.2 Contrastive constructions

Contrastive constructions are used to counterpoise different assessments, either to make a correction (Example 121) or to add new information (Example 122). In other cases, they can express obligation (Example 123). We agreed to annotate these structures with the value "contrast".

121. **No** vinieron 2 soldados, **sino** 6.

    *Six soldiers came, **not** two.*

122. **No solo** lleva rueda de recambio **sino también** caja de herramientas.

    *It **not only** has a spare tire **but also** a toolbox.*

123. **No** hay más solución **que** comprar una lavadora.

    *There is **no** other solution **than** to buy a washing machine.*

Example (121) declares/states that six soldiers came and the negation refers to a supposed information about the number of soldiers who came. The function of the negation is to contrast the belief with what really happened.

Example (122) is a very common coordination construction: "no solo... sino también" *[not only... but also]*. The sentence can be paraphrased as "Lleva rueda de recambio y caja de herramientas" *[It has spare tire and toolbox]*.

Finally, Example (123) is another case of a pattern that is used to reinforce what is said. The sentence can be paraphrased as an affirmative one "La única solución es comprar una lavadora" *[The only solution is to buy a washing machine]*.

4.6 Corpus description and statistics

The SFU Review$_{SP}$-NEG corpus[14] (Jiménez-Zafra et al., 2018)[15] is an extension of the Spanish part of the SFU Review corpus[16] (Taboada et al., 2006) and it could be considered the counterpart of the SFU Review Corpus with negation and speculation annotations (Konstantinova et al., 2012). It is composed of 400 product reviews, 25 positive reviews and 25 negative reviews

[14]It is publicly available and can be downloaded at `http://sinai.ujaen.es/sfu-review-sp-neg-2/` under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.
[15]First Online: 22 May 2017 `https://doi.org/10.1007/s10579-017-9391-x`
[16]`https://www.sfu.ca/~mtaboada/SFU_Review_Corpus.html`

from eight different domains: cars, hotels, washing machines, books, cell phones, music, computers and movies. We automatically annotated each review at the token level with PoS-tags and lemmas using Freeling (Padró & Stanilovsky, 2012), and manually annotated it 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. It is the first Spanish corpus that includes the event in the annotation of negation and that takes into account discontinuous negation cues. Moreover, it is the first corpus in which it is annotated how negation affects the words that are within its scope.

General characteristics of the corpus are presented in Table 4.3: total of positive and negative reviews (columns 2 and 3), number of sentences (column 4), average of sentences per document (column 5), number of tokens (column 6) and average of tokens per sentence (column 7). As the fourth and sixth columns show, the corpus is of considerable size, 221,884 tokens and 9,446 sentences. It is larger than the existing Spanish corpora: UAM Spanish Treebank (Sandoval & Salazar, 2013), UHU-HUVR (Cruz Díaz et al., 2017) and IULA Spanish Clinical Record (Marimon et al., 2017) corpora consist of 1,500, 8,412 and 3,194 sentences, respectively. Regarding the length of the reviews, there are differences depending on the domain in terms of sentences (column 5), but not in terms of tokens (column 7). In particular, notable differences are observed in the average number of sentences of movies and books reviews with respect to the other domains because most of them contain plot fragments.

Table 4.3: Statistics of the SFU Review$_{\mathrm{SP}}$-NEG corpus.

Domain	#Positive documents	#Negative documents	#Sentences	Avg. sentences per document	#Tokens	Avg. tokens per sentence
Books	25	25	1,840	36.80	42,171	22.92
Cars	25	25	756	15.12	18,697	24.73
Cells phones	25	25	1,021	20.42	23,286	22.81
Computers	25	25	651	13.02	16,554	25.43
Hotels	25	25	853	17.06	19,235	22.55
Movies	25	25	2,472	49.44	59,680	24.14
Music	25	25	953	19.06	23,928	25.11
Washing machines	25	25	900	18.00	18,333	20.37
Total	200	200	9,446	23.62	221,884	23.49

In this corpus we distinguish three type of sentences: i) sentences without a negation structure

(<sentence>), ii) sentences with one negation structure (<sentence complex="no">), and iii) sentences with two or more negation structures (<sentence complex="yes">). Table 4.4 shows the distribution of sentences per domain according to the number of negation structures they contain. Out of 9,446 sentences, 2,191 sentences have been annotated with one negation structure and 887 with two or more. Each negation structure has a value assigned to it according to the semantic interpretation of the negation cue: negation (label "neg"), opposition or contrast between two or more elements (label "contrast"), comparison between two or more elements (label "comp") and, no negation (label "noneg"). Table 4.5 presents the distribution of negation structures and their semantic value per domain. Notice that the semantic value negation "neg" is the most frequent for the negation structures (91.08%), and that the movies domain has the highest number of negation structures, followed by the books and cell phones domains.

Table 4.4: Distribution of sentences per domain according to the number of negation structures (one <complex="no">, two or more <complex="yes">).

Domain	#Sentences complex="no"	#Sentences complex="yes"	Total
Books	451	191	642
Cars	184	76	260
Cells phones	219	120	339
Computers	165	65	230
Hotels	208	76	284
Movies	560	208	768
Music	195	85	280
Washing machines	209	66	275
Total	2,191	887	3,078

Table 4.5: Distribution of negation structures taking into accounts their semantic value: "neg", "contrast", "comp" and "noneg".

Domain	neg	contrast	comp	noneg	Total
Books	805 (88.07%)	65 (7.11%)	4 (0.44%)	40 (4.38%)	914 (100%)
Cars	324 (91.79%)	12 (3.40%)	7 (1.98%)	10 (2.83%)	353 (100%)
Cells phones	501 (94.89%)	6 (1.13%)	1 (0.19%)	20 (3.79%)	528 (100%)
Computers	305 (93.27%)	13 (3.97%)	1 (0.31%)	8 (2.45%)	327 (100%)
Hotels	360 (90.91%)	8 (2.02%)	2 (0.50%)	26 (6.57%)	396 (100%)
Movies	936 (89.83%)	48 (4.61%)	10 (0.96%)	48 (4.60%)	1,042 (100%)
Music	373 (92.33%)	15 (3.71%)	3 (0.74%)	13 (3.22%)	404 (100%)
Washing machines	337 (92.84%)	8 (2.20%)	2 (0.55%)	16 (4.41%)	363 (100%)
Total	3,941 (91.08%)	175 (4.04%)	30 (0.70%)	181 (4.18%)	4,327 (100%)

Negation structures with a semantic value different from "noneg" have a label that indicates how negation affects the words that are within their scope, that is, whether negation changes the polarity (label change="yes") or not (change="no"), or whether it modifies the polarity reducing its intensity (polarity_modifier="reduction") or increasing it (polarity_modifier="increment"). As shown in Table 4.6, most of these negation structures are used to change the polarity of the elements included in their scope. Only 5.67% of negation structures do not change the polarity, they correspond to those cases that affect words expressing facts or non-internal states. Regarding the structures with modifiers, they are more frequently use to increment than to decrease the positive or negative value of the polarity.

Table 4.6: Distribution of negation structures according to the influence of negation in the words of its scope.

Domain	change "yes"	change "no"	polarity_modifier "increment"	polarity_modifier "reduction"	Total
Books	606 (69.34%)	76 (8.69%)	143 (16.36%)	49 (5.61%)	874 (100%)
Cars	234 (68.22%)	15 (4.37%)	55 (16.04%)	39 (11.37%)	343 (100%)
Cells phones	392 (77.17%)	13 (2.56%)	56 (11.02%)	47 (9.25%)	508 (100%)
Computers	231 (72.41%)	15 (4.70%)	54 (16.93%)	19 (5.96%)	319 (100%)
Hotels	248 (67.03%)	17 (4.60%)	55 (14.86%)	50 (13.51%)	370 (100%)
Movies	731 (73.54%)	59 (5.93%)	112 (11.27%)	92 (9.26%)	994 (100%)
Music	285 (72.90%)	18 (4.60%)	51 (13.04%)	37 (9.46%)	391 (100%)
Washing machines	241 (69.45%)	22 (6.34%)	49 (14.12%)	35 (10.09%)	347 (100%)
Total	2,968 (71.59%)	235 (5.67%)	575 (13.87%)	368 (8.87%)	4,146 (100%)

The structures that a Spanish negation processing system should identify are those of the type "neg". In the following we analyze the information of the corpus related to them. The other structures evidence that negation cues do not always negate, which makes it difficult to deal with this phenomenon.

The corpus consists of 9,446 sentences, out of which 2,825 (29.91%) contains at least one structure of type "neg". We say at least because we find sentences with one negation cue (2,028), two negation cues (578) and even three or more negation cues (219), as it is reported in Table 4.7. As we can see in this table, 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 4.7: Statistics about the sentences that contain negation cues that negate.

	#Sentences	% Sentences	Avg. tokens per sentence
0 negations	6,621	70.09	19.47
1 negation	2,028	21.47	28.78
2 negations	578	6.12	39.10
≥3 negations	219	2.32	54.72
Total	9,446	100.00	23.49

Negation cues in this corpus can be simple, if they are composed of a single token (e.g., “no” *[not]*, “nunca” *[never]*), continuous, if they have two or more contiguous tokens (e.g., “casi no” [almost not], “en mi vida” [never in my life]) or discontinuous, if they consist of two or more non-contiguous tokens (e.g., “no-en absoluto” *[not-at all]*, “no-nada” *[not-nothing]*). Table 4.8 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 *continuous* cues (186) and a considerable amount of *discontinuous* cues (608). Table 4.9 provides the most frequent cues in the corpus. There are 246 different negation cues, being “no” *[not]* the most common with a total of 2,317 occurrences.

Table 4.8: Total and percentage of negation cues that negate by type.

	#Negation cues	% Negation cues
Simple	3,147	79.85
Continuous	186	4.72
Discontinuous	608	15.43
Total	3,941	100.00

Table 4.9: Most frequent negation cues that negate.

Cue	#	%
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

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 negation cue, the event 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 4.10 we present the total and percentage of scopes distributed by type, by the number of tokens they have and by the percentage of the sentence they cover. 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 that spans only before the cue (230). Most scopes span between 3 and 7 tokens (61.12%), but almost 25% span more than 7 tokens. They span up to 43 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.

Table 4.10: Scope statistics.

		#	%
Type	before cue	230	5.84
	after cue	2,702	68.56
	before and after cue	1,009	25.60
#Tokens	<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

The corpus constitute an invaluable resource for the study of negation in Spanish. Given the opinionated nature of the texts involved (reviews), it is also very useful to test whether the integration of negation processing systems into sentiment analysis systems boost their accuracy.

Table 4.11 shows the distribution of reviews per domain according to the presence of negation taking into account the polarity of the documents. As can be seen, negation is present in almost all negative reviews: from a total of 200 reviews, only 3 negative reviews do not have structures that negate. Regarding positive reviews, the number of documents without a structure that negates is 12.

Table 4.11: Distribution of reviews per domain and polarity according to the presence of negation.

	Positive reviews		Negative reviews	
Domain	With negation	Without negation	With negation	Without negation
Books	23	2	25	0
Cars	23	2	25	0
Cells phones	24	1	25	0
Computers	22	3	24	1
Hotels	23	2	25	0
Movies	24	1	25	0
Music	24	1	24	1
Washing machines	25	0	24	1
Total	188	12	197	3

4.7 Conclusion

This chapter introduces the existing corpora annotated with negation information in several languages, which are essential for the development of negation processing systems. An exhaustive search have been conducted, finding corpora for the following languages: English, Spanish, Swedish, Dutch, Japanese, Chinese, German and Italian.

We have presented the SFU Review$_{SP}$-NEG corpus, the largest Spanish corpus to date, and the first Spanish corpus that includes the event in the annotation of negation and that takes into account discontinuous negation cues. Moreover, it is the first corpus in which it is annotated how negation affects the words that are within its scope. For the annotation of the corpus, the components of negation have been defined and delimited, and a typology of negation patterns in Spanish has been created, which has the advantage of being easily expressed in terms of a tagset and of presenting clearly delimited types, thus avoiding ambiguity in the annotation process. The annotation process followed and the problematic cases found during it have been reported in order to facilitate the annotation task of this phenomenon for other researchers.

This corpus could be very useful for the research community for the study of negation in Spanish. It is a well-studied phenomenon from a theoretical perspective (L. R. Horn, 1989;

L. Horn, 2010), but its computational treatment has not been extensively studied for languages other than English. Given the opinionated nature of the texts involved (reviews), it could also be useful for sentiment analysis. In these systems it is essential not only to identify negation but also determine its scope and decide whether the negation changes the polarity of the sentence, or increments or reduces its intensity. Moreover, it could also be relevant in a wide range of applications, such as information retrieval (Liddy et al., 2000), information extraction (Savova et al., 2010) or machine translation (Baker et al., 2012), where it is crucial to detect when a fragment of text expresses a different meaning due to the presence of negation.

When resources are developed, the most ideal is to give them visibility in order to contribute to the advancement of the phenomenon studied. Therefore, apart from using it to develop a Spanish negation processing system (Chapter 5) and apply it to improve the task of sentiment analysis (Chapter 6), we give it visibility by organizing the first Spanish negation workshop: NEGES (Chapter 7).

Chapter 5

A system to process negation in Spanish

The computational treatment of negation has not been resolved yet due to its complexity, the multiple linguistic forms in which it can appear and the different ways it can act on the words within its scope. If we want to develop systems that approach human understanding, it is necessary to incorporate the treatment of one of the main linguistic phenomena used by people in their daily communication.

Natural Language Processing is a subfield of Artificial Intelligence that focuses on the processing and generation of human language in order for computers to learn, understand and produce human language (Hirschberg & Manning, 2015). Some linguistic phenomena such as negation, speculation, irony or sarcasm, pose challenges for computational natural language learning. One might think that, given the fact that negations are so crucial in language, most Natural Language Processing pipelines incorporate negation modules and that the computational linguistics community has already addressed this phenomenon. However, this is not the case. Work on processing negation has started relatively late as compared to work on processing other linguistic phenomena as it has been shown in Chapter 2.

Four tasks are usually performed in relation to processing negation: i) negation cue detection, in order to find the words that express negation; ii) scope identification, in order to know 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 in order to find the part of the scope that is most prominently negated. Most of the works have modeled these tasks as token-level classification tasks, where a token is classified as being at the beginning, inside or outside a negation cue, scope, event or focus. Scope, event and focus identification tasks are more complex because they depend on negation cues detection. In this chapter, it is presented the system we developed for task i) and task ii).

Most applications treat negation in an ad hoc manner by processing main negation constructions, but processing negation is not as easy as using a list of negation cues and applying look-up methods because negation cues do not always act as negators. For example, in the sentence *"You bought the car to use it, didn't you?"* the cue "not" is not used as a negation but it is used to reinforce the first part of the sentence. We believe that there are three main reasons for which most applications treat negation in an ad hoc manner: one is that negation is a complex phenomenon, which has not been completely modeled yet. In this way it is similar to phenomena like factuality for which it is necessary to read large amounts of theoretical literature in order to put together a model, as shown by Saurí's work on modelling factuality for its computational treatment (Saurí & Pustejovsky, 2009). A second reason is that, although negation is a phenomenon of habitual use in language, it is difficult to measure its quantitative impact in some tasks such as anaphora resolution or text simplification. The number of sentences with negation in the English texts of the corpora analyzed is between 9.37% and 32.16%, while in Spanish texts it is between 10.67% and 34.22%, depending on the domain. In order to evaluate the improvement that processing negation produces, it would be necessary to focus only on those parts of the text in which negation is present and perform an evaluation before and after its treatment. However, from a qualitative perspective, its impact is very high, for example, when processing clinical records, because the health of patients is at stake. A third reason is that there are no large corpora exhaustively annotated with negation phenomena, which hinders the development of machine learning systems.

In this chapter we present a machine learning system that processes negation in Spanish (Jiménez-Zafra et al., 2019). 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. In first place, we analyze existing corpora to check if it possible to merge them in order to create a larger training corpus for our system. Then, we present the system developed and the results obtained in the experiments performed with the selected data. After that, 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. Finally, we report conclusions.

5.1 Data selection

One of the first steps when attempting at developing a machine learning negation processing system is to check whether there are training data and to decide whether their quality is good enough. Differently than for other well established tasks like semantic role labelling or parsing, for negation there is no corpus of reference, but several small corpora, and, ideally a training corpus needs to be large for a system to be able to learn. This motivates our main research questions: Is it possible to merge the existing negation corpora in order to create a larger training corpus? What are the problems that arise? In order to answer the questions we first review all existing corpora and characterise them in terms of several factors: type of information about negation that they contain, type of information about negation that is lacking, and type of application they would be suitable for. Available corpora that contain a representation of negation can be divided into two types (Fancellu et al., 2017): i) those that represent negation in a logical form, using quantifiers, predicates and relations (e.g. Groningen Meaning Bank (Basile et al., 2012), DeepBank (Flickinger et al., 2012)); and, ii) those that use a string-level, where the negation operator and the elements (scope, event, focus) are defined as spans of text (e.g. BioScope (Vincze et al., 2008), ConanDoyle-neg (Morante & Daelemans, 2012)). It should be noted that we focus on corpora which deal with *string-level* negation.

In Subsection 2.1.3: "*Corpora annotated with negation*" of Chapter 2: "*Background*" the existing corpora so far have been described. To the best of our knowledge, there are corpora annotated for English, Spanish, Swedish, Japanese, Chinese, Dutch, German and Italian. For each corpus we collected information about the source of the texts, the size and the percentage of sentences that contain negation. In addition, we indicate what type of information has been annotated, whether the annotation has been thought of for a specific task and whether negation is the main focus of the annotation. In relation to negation, we specify what types of negation have been annotated (syntactic, lexical, morphological), what elements have been annotated (cue, scope, event, focus) and what guidelines have been followed for the annotation. Moreover, we include information on the number of annotators, their background and how the inter-annotator agreement was measured. Finally, we also provide information on the availability of the corpora and their format. Here, we analyze them in order to select the data that will be used in our system.

5.1.1 Analysis criteria

The criteria applied to review the corpora are listed below:

- **Language**: the language(s) of the texts included in the corpora. This characteristic should always be specified in the description of any corpus, as it conditions its use.
- **Domain**: field to which the texts belong. Although cross-domain methodologies are being used for many tasks (Szarvas et al., 2012; F. Li et al., 2012; Bollegala et al., 2016), the domain of a corpus partly determines its area of application since different areas have different vocabularies.
- **Availability**: accessibility of the corpora. We indicate whether the corpus is publicly available and we provide the links for obtaining the data when possible. Corpora annotation is time consuming and expensive, so it is not only necessary that corpora exist, but also that they be publicly available for the research community to use.
- **Guidelines**: we study the guidelines used for the annotation showing similarities and differences between corpora. The definition of guidelines for the annotation of any phenomenon is fundamental because the generation of quality data will depend on it. The

goal of annotation guidelines can be formulated as follows: given a theoretically described phenomenon or concept, describe it as generically as possible but as precisely as necessary so that human annotators can annotate the concept or phenomenon in any text without running into problems or ambiguity issues (Ide, 2017).

- **Sentences**: corpus size is measured in sentences. The number of sentences is the information that is usually provided in the statistics of a corpus to give an idea of its extension, although the important thing is the information contained in them.

- **Annotated elements**: this aspect refers to the elements on which the annotation has been performed, such as sentences, events, relationships, etc.

- **Elements with negation**: total number of elements that have been annotated with negation. The annotation should cover all the relevant cases that algorithms need to process in order to allow for a rich processing of negation.

- **Negation types**: refers to the types of negation that have been annotated[1]:

  - **Syntactic negation**.
  - **Lexical negation**.
  - **Morphological or affixal negation**.

- **Negation components**: components of negation that have been annotated[2]:

  - **Negation cue**.
  - **Scope**.
  - **Negated event**.
  - **Focus**.

[1]Negation types are described in detail in Section 2.1: "*Negation*" of Chapter 2: "*Background*".
[2]Negation components are described in detail in Section 2.1: "*Negation*" of Chapter 2: "*Background*".

5.1.2 Corpora analysis

Here we present an analysis of the corpora based on the criteria previously defined. Although this doctoral thesis focuses on the processing of negation in Spanish, in Subsection 2.1.3: "*Corpora annotated with negation*" of Chapter 2: "*Background*" it has been presented a compilation of the corpora existing so far and, here, we also analyze all of them, as it may be useful for the scientific community to advance in the study of this phenomenon in other languages. In Appendix A, the information analyzed is summarized in Table A.6, Table A.7 and Table A.8.

5.1.2.1 Language and year of publication

The years of publication of the corpora (Table A.1, Appendix A) show that interest in the annotation of negation started in 2007 with English texts. Thenceforth, a total of 11 English corpora have been presented. The following language for which annotations were made was Swedish, although we only have evidence of one corpus presented in 2010. For other languages, the interest is more recent. The first corpus annotated with negation in Spanish appeared in 2013 and since then 5 corpora have been compiled, 3 of them in the last two years. There are also corpora for Dutch, Japanese, Chinese, German and Italian, although it seems that it is an emergent task because we only have evidence of one corpus annotated with negation in each language. These corpora appeared in 2014, 2016, 2016 and 2017, respectively. From the analysis of the years of publication, it can be observed that it is a task of recent interest for Spanish, Dutch, Japanese, Chinese, German and Italian, and that for English it is something more established or at least more extensively studied. For Swedish, although annotation with negation started three years after the English annotation, no continuity is observed as there is only one corpus annotated with negation.

5.1.2.2 Domain

If we look at Table A.6, Table A.7 and Table A.8 (see Appendix A), it can be seen that in the corpora annotated so far there is a special interest in the medical domain, followed by reviews. In English, out of 11 corpora, 5 focus on the biomedical domain, 3 on reviews or opinion articles, 1 on journal stories, 1 on tutorial dialogues and 1 on the literary domain. In Spanish, 3 of the corpora are about clinical reports, 1 about movies, books and product reviews, and 1 about newspaper articles. In other languages, we have only found one corpora annotated with negation per language. For Swedish, Dutch and German, the domain is clinical reports, for Japanese news articles and reviews, for Italian it is news articles and the Chinese corpus is about scientific literature, product reviews and financial articles. This information shows that in all languages there is a common interest in processing negation in clinical/biomedical texts. This is understandable because detecting negated concepts is crucial in this domain. If we want to develop information extraction systems, it is very important to process negation because clinical texts often refer to concepts that are explicitly not present in the patient, for example, to document the process of ruling out a diagnosis:

> *"In clinical reports the presence of a term does not necessarily indicate the presence of the clinical condition represented by that term. In fact, many of the most frequently described findings and diseases in discharge summaries, radiology reports, history and physical exams, and other transcribed reports are denied in the patient."*
>
> (Chapman et al., 2001, p. 301)

Not recognizing these negated concepts can cause problems. For example, if the concept "pulmonary nodules" is recognized in the text *"There is no evidence of pulmonary nodules"* and negation is not detected, the diagnosis of a patient will be totally different.

Considering the corpora analyzed, another domain that has attracted the attention of researchers is opinion articles or reviews. The large amount of content that is published on the Internet has generated great interest in the opinions that are shared in this environment through social networks, blogs, sales portals and other review sites. This user-generated content is useful for marketing strategies because it can be used to measure and monitor customer satisfaction. It is a quick way to find out what customers liked and what they did not like. Moreover, micro-bloggings such as Twitter are being used to measure voting intention, people's moods and even to predict the success of a film. The study of negation in this domain is very important because if negation is present in a sentence and it is not taken into account, a system can extract a completely different opinion than the one published by the user. In Example (124) we can find a positive opinion that changes to negative if negation is present as in Example (125), or by contrast, in Example (126) there is a positive opinion in which negation is present whose meaning change if it does not have negation as in Example (127).

124. The camera works well.

125. The camera does not work well.

126. I have not found a camera that works better.

127. I have found a camera that works better.

Other domains for which interest has also been shown, although to a lesser extent, are journal stories, tutorial dialogues, the literary domain, newspaper articles, scientific literature and financial articles.

5.1.2.3 Availability

The extraction and annotation of corpora is time consuming and expensive. Therefore, it is not enough that corpora exists, but it must also be made available for the scientific community to allow progress in the study of the different phenomena. In this overview we focus on negation, and of the 22 corpora collected, 15 are publicly available. Of the 7 non-available corpora, 5 contain clinical reports and legal and ethical issues may be the reasons for this. The links for obtaining the data of the different corpora (when possible) are shown in Table A.2 (Appendix A).

5.1.2.4 Size

The size of a corpus is usually expressed in number of sentences and/or tokens. It is important to know the extension of the corpus, but what is really important is the number of elements of the phenomenon or concept that has been annotated. As we focus on negation, the relevant information is the total of elements (sentences, events, relationships, etc.) that have been annotated and the total of elements that have been annotated with negation. Both are very important because for a rich processing of negation, algorithms need examples of elements with and without negation in order to cover all possible cases.

In Table A.3 (Appendix A) we present information on the size of the corpora. The existing corpora are not very large and they do not contain many examples of negations. However, differences in languages are observed. According to the existing corpora, negation is used less frequently in English, Swedish, Dutch and Japanese, while in Spanish, Italian, Chinese and German, it appears more frequently. The percentage of negated elements in English ranges from 6.12% to 32.16%. It should be noted that the first percentage corresponds to relations in the biomedical domain and the second to sentences in product reviews. In Swedish we are aware of only one corpus, the Stockholm Electronic Patient Record, which consists of clinical reports and contains 10.67% of negated expressions. The EMC Dutch corpus is also composed of clinical reports and the percentage of medical terms negated is 14.04%. The Review and Japanese corpus consists of reviews and newspaper articles and 16.59% of the sentences contain negations. For Spanish the frequency of negated sentences goes from 10.67% in newspaper articles to 34.22% in clinical reports. In Italian, the existing corpus is composed of news articles and the percentage of negated sentences is 21.55%. The German negation and speculation corpus consists of clinical reports and 39.77% of the medical terms annotated are negated. Finally, the Chinese corpus of scientific literature, product reviews and financial articles contains 26.82% of negated sentences. The percentages of elements with negation do not always correspond to sentences, but in some cases are related to events, expressions, relationships, medical terms or answers, depending on the level at which the annotation has been made. Therefore, for a better comparison of the frequency of occurrence of negation in sentences we have also calculated the average per language, taking into account only those corpora that provide information at the sentence level. Thus, the average number of sentences with negation in English texts is 17.94% and in Japanese 16.59%, while for Spanish it is 29.13%, for Italian 21.55% and for Chinese 26.82%[3]. On the other hand, if we take a look at the domain of the corpora we can say that, in general, clinical reports are the type of texts that have a greater presence of negation, followed by reviews/opinion articles, and biomedical texts.

Although negation is an important phenomenon for NLP tasks, it is relatively infrequent compared to other phenomena. Therefore, in order to train a negation processing system properly, it would be necessary to merge some corpora. However, in order to do this, the annotations of the corpora must be consistent, a fact that we will analyze in the following entry *Annotation guidelines*.

5.1.2.5 Annotation guidelines

The definition of guidelines for data annotation is fundamental because the consistency and quality of the annotations will depend on it. We analyze several aspects of the annotation guidelines of the corpora reviewed:

- Existence and availability. Have annotation guidelines been defined? Are they available?
- Negation. What types of negation have been taken into account (syntactic and/or lexical and/or morphological)?
- Negation elements. What elements of negation have been annotated? Cue? Scope? Negated event? Focus?
- Tokenization. What tokenizer has been used?
- Annotation scheme and guidelines. What annotation scheme and guidelines have been used?

[3] The Italian and Chinese percentages correspond to the only existing corpus in each language. The percentages of sentences annotated with negation in Swedish and Dutch could not be calculated because the information provided by the authors corresponds to expressions and medical terms, respectively

Existence and availability

Ide (2017) indicates that the purpose of the annotation guidelines is to define a phenomenon or concept in a generic but precise way so that the annotators do not have problems or find ambiguity during the annotation process. Therefore, it is very important to define annotation guidelines that annotators can consult whenever necessary. In addition, these guidelines should be available not only for the annotators of the ongoing project but also for other researchers to use them. The definition of annotation guidelines involves a long process of study and the time spent on it should serve to facilitate the annotation process to other researchers. In Table A.4 (Appendix A), we show the link or reference to the annotation guidelines of the different corpora.

As Table A.4 (Appendix A) shows, there is information about the annotation guidelines of most corpora, although some guidelines are not complete. For one third of the corpora the guidelines are not available. In some cases, it is indicated that existing annotation guidelines were adopted with some modifications, but these modifications are not reflected.

Negation elements

Another important aspect to be analyzed from the corpora is what elements of negation have been annotated. As mentioned in Subsection 5.1.1, negation is often represented using one or more of the following four elements: cue, scope, focus and event.

The first task that a negation processing system should carry out is the identification of **negation cues**, because it is the one that will allow us to identify the presence of this phenomenon in a sentence and because the rest of the elements are linked to it. Most of the existing corpora contain annotations about negation cues. However, some of the corpora of the biomedical and clinical domain take negation into account only to annotate whether an event or relationship is negated, but not to annotate the cue. They use a clinical perspective more than a linguistic one. This is the case with the BioInfer, Genia Event, IxaMed-GS, EMC Dutch and German negation and speculation corpora.

Depending on the negation cue used, we can distinguish three main types of negation: syntactic, lexical and morphological. Most annotation efforts focus on syntactic negation. It has been difficult to summarize the types of negation considered, because in some cases they are not specified in the description of a corpus nor in the guidelines, and we have had to manually review the annotations of the corpora and/or contact the annotators. In Table A.5 (Appendix A), we determine for each corpus whether it contains annotations about negation cues (✓) or not (-), and what types of negation have been considered. In the second column, we use *CS*, *CM* and *CL* to indicate that all syntactic, morphological and lexical negation cues have been taken into account, *NA* if the information is not available or *PS*, *PM* and *PL* if syntactic, morphological and lexical negations have been considered partially, for example because only negation which acts on certain events or relationships have been considered or because a list of predefined markers have been used for the annotation.

Once the negation cue has been identified, we can proceed to the identification of the rest of the elements. The **scope** is the part of the sentence affected by the negation cue, that is, it is the set of words on which negation acts and on which to proceed depending on the objective of the final system. In most of the corpora reviewed the scope has been annotated, except in the Genia Event, Stockholm Electronic Patient Record, PropBank Focus (PB-FOC), EMC Dutch, Review and Newspaper Japanese, IxaMed-GS, and German negation and speculation corpora. The two remaining elements, **event** and **focus** have been annotated to a lesser extent. The **negated event** is the event or property that is directly negated by the negation cue, usually a verb, a noun or an adjective. It has been annotated on two English corpora (Genia Event and ConanDoyle-neg), three Spanish corpora (IxaMed-GS, SFU Review$_{SP}$-NEG and UHU-HUVR) and the EMC Dutch, the Fact-Ita Bank Negation and the German negation and speculation corpora. On the other hand, the **focus**, the part of the scope most prominently or explicitly negated, has only been annotated on three English corpora (PropBank Focus (PB-FOC), Deep Tutor Negation and SOCC) and in the Review and Newspaper Japanese corpus, which shows that it is the least studied element. In the fourth, fifth and sixth columns of Table A.5 (Appendix A) this information is represented using ✓ if the corpus contains annotations about the scope, event and focus, respectively, or - otherwise.

Tokenization

The way in which each corpus was tokenized is also important and it is only mentioned in the description of the SFU Review$_{SP}$-NEG corpus. Why is it important? The identification of negation cues and the different elements (scope, event, focus) is usually carried out at token level, that is, the system is trained to tell us whether a token is a cue or not and whether it is part of a scope or not. Tokenization is also important when we want to merge annotations. If the tokenization is different in several versions of a corpus or in different corpora, merging annotations will pose technical problems.

Annotation scheme and guidelines

In Subsection 2.1.3: "Corpora annotated with negation" of Chapter 2: "Background" an example of each corpus has been provided whenever possible. If we take look at them we can see that the annotation schemes are different. There is no uniformity between languages, nor between domains. Moreover, the annotation guidelines are different. There are divergences in the negation aspects being annotated (negation cue, scope, event, focus) and the criteria used to annotate these elements. The main differences are related to the following aspects[4]:

- Inclusion or not of the subject within the scope. For example, in the UAM Spanish Treebank corpus all the arguments of the negated events, including the subject, are included within the scope of negation (Example (128)). On the contrary, in the IULA Spanish Clinical Record corpus the subject is included within the scope (Example (129)) only when it is located after the verb (Example 130) or when there is an unaccusative verb (Example (131)).

128. Gobierno, patronal y cámaras tratan de demostrar [que *Chile*$_{SUBJ}$ **no** castiga a las empresas españolas].

     Government, employers and chambers try to demonstrate that Chile does not punish Spanish companies.

[4] In the examples provided to clarify differences, we mark in bold negation cues and enclose negation scopes between square brackets.

129. MVC_{SUBJ} **sin** [ruidos sobreañadidos].

   NBS no additional sounds.

130. Se **descarta** [$enolismo_{SUBJ}$].

   Oenolism discarded.

131. [*El dolor*]$_{SUBJ}$ **no** [ha mejorado con nolotil].

   Pain has not improved with nolotil.

- Inclusion or not of the cue within the scope. For example, in the annotation of the SOCC corpus, the negation cue was not included within the scope (Example (132)), whereas in the BioScope corpus it was included (Example (133)).

132. I **cannot** [believe that one of the suicide bombers was deported back to Belgium.]

133. Mildly hyperinflated lungs [**without** focal opacity].

- Strategy to annotate as scope the largest or shortest syntactic unit. For example, in the Product Review corpus annotators decided to annotate the minimal span of a negation covering only the portion of the text being negated semantically (Example (134)), whereas in ConanDoyle-neg corpus the longest relevant scope of the negation cue was marked (Example (135)).

134. Long live ambitious filmmakers with **no** [talent]

135. [It was] suggested, but **never** [proved, that the deceased gentleman may have had valuables in the house, and that their abstraction was the motive of the crime].

- Use a set of predefined negation cues or all the negation cues present in a text. For example, for scope annotation in the Product Review corpus, a lexicon of 35 explicit negation cues was defined and, for instance, the cue "not even" was not considered, while in the SFU Review$_{SP}$-NEG corpus all syntactic negation cues were take into account.

These differences provoke that the annotations are not compatible, not even within corpora of the same language and domain.

5.1.3 Discussion

The perspective that we have taken when analyzing the corpora annotated with negation is computational, since our final goal is not to evaluate the quality of the annotations from a theoretical perspective, but to determine whether corpora can be used to develop a negation processing system. In order to achieve this we need a significant amount of training data, even more taking into consideration that negation is a relatively infrequent phenomenon as compared to tasks like semantic role labeling. Additionally, we need qualitative data that cover all possible cases of negation. Since the existing corpora are small, we have analyzed them in order to evaluate whether it is possible to merge the corpora into a larger one. Two features that are relevant when considering merging corpora are the language and the domain. Next, we discuss the possibility of merging corpora according to each of these aspects.

Table 5.1: Overall negation processing tasks for which the corpora could be used, by language.

	Negation cues detection	**Scope identification**	**Event extraction**	**Focus detection**
English	BioScope PropBank Focus (PB-FOC) ConanDoyle-neg SFU Review$_{EN}$ NEG-DrugDDI NegDDI-DrugBank Deep Tutor Negation SOCC	BioInfer BioScope ConanDoyle-neg SFU Review$_{EN}$ NEG-DrugDDI NegDDI-DrugBank Deep Tutor Negation SOCC	Genia Event ConanDoyle-neg	PropBank Focus (PB-FOC) Deep Tutor Negation SOCC
Spanish	UAM Spanish Treebank SFU Review$_{SP}$-NEG UHU-HUVR IULA Spanish Clinical Record	UAM Spanish Treebank SFU Review$_{SP}$-NEG UHU-HUVR IULA Spanish Clinical Record	IxaMed-GS SFU Review$_{SP}$-NEG UHU-HUVR	
Swedish	Stockholm Electronic Patient Record			
Dutch		EMC Dutch		
Japanese	Review and Newspaper Japanese			Review and Newspaper Japanese
Chinese	CNeSP	CNeSP		
German			German negation and speculation	
Italian	Fact-Ita Bank Negation	Fact-Ita Bank Negation		

On the one hand, it could be necessary to merge corpora for processing negation in a specific language. As we have mentioned before, there are four general tasks related to negation processing: negation cue detection, scope identification, negated event extraction and focus detection. In Table 5.1 we show for which of these tasks each corpus can be used. Negation cue detection and scope identification are the tasks for which there are more corpora. However, it is noteworthy that in some of the corpora (BioInfer, Genia Event, Product Review, EMC Dutch, IxaMed-GS and, German negation and speculation corpus) negation cues have not been annotated, despite the fact that the cue is the element that denotes the presence of negation in a sentence and the one to which the rest of elements (scope, event and focus) are connected. The task with the fewest annotated corpora is focus detection, probably because annotating focus is a difficult task that depends on stress and intonation. For the event extraction task there are also few corpora, most of them belonging to the biomedical and clinical domains.

Table 5.2: Specific tasks for which the corpora could be used to evaluate the impact of processing negation.

	Information extraction in the biomedical and clinical domain	**Drug-drug interactions**	**Clinical events detection**	**Bio-molecular events extraction**	**Sentiment Analysis**	**Constructiveness and toxicity detection**
English	BioInfer Genia Event BioScope	NEG-DrugDDI NegDDI-DrugBank		Genia Event	Product Review SFU Review$_{EN}$	SOCC
Spanish	IxaMed-GS UHU-HUVR IULA Spanish Clinical Record		IxaMed-GS UHU-HUVR		SFU Review$_{SP}$-NEG	
Swedish	Stockholm Electronic Patient Record					
Dutch	EMC Dutch					
Japanese					Review and Newspaper Japanese	
Chinese					CNeSp	
German	German negation and speculation					
Italian					Fact-Ita Bank Negation	

On the other hand, it could be necessary to merge corpora in order to evaluate the impact of processing negation in specific tasks such as information extraction in the biomedical and clinical domain, drug-drug interactions, clinical events detection, bio-molecular events extraction, sentiment analysis and, constructiveness and toxicity detection. Moreover, corpora can be used to improve information retrieval and question answering systems. In Table 5.2, we show for each language the specific tasks for which the corpora could be used. The applicability tasks of most of the corpora analyzed are i) information extraction in the biomedical and clinical domain; and ii) sentiment analysis. For the first task, the role of negation could be evaluated in English, Spanish, Swedish, Dutch and German (5 of the 8 languages analyzed) and, for the second task, it could be analyzed in English, Spanish, Japanese, Chinese and Italian (5 of the 8 languages analyzed). For drug-drug interactions, bio-molecular events extraction and constructiveness and toxicity detection, it could only be analyzed in English; and for clinical events detection, it could only be evaluated in Spanish.

However, our analysis shows that merging the corpora is not an option in their current state. There are corpora for which it is not possible to make the union simply because they are not publicly available. Of the 22 corpora collected, 7 are non-available, and 5 of them consist of clinical reports. These corpora are not available due to legal and ethical issues, which makes it difficult to study negation in this domain, a domain in which processing negation is crucial because the health of patients is at stake. In general, we find the following problems that are related to the aspects analyzed regarding annotation guidelines:

1. There are corpora for which the annotation guidelines are not available or are not complete. This is a problem because in order to merge corpora we need to know the criteria followed for the annotation and we need to know whether the corpora are consistent. For example, if negation cues are included within the scope of negation, this rule must be satisfied in all the corpora used to train a negation processing system.

2. Corpora have been annotated with different purposes. Some corpora have been annotated taking into account the final application, whereas others are annotated from a linguistic point of view. There are cases in which not all types of negation have been considered or

they have only partially been taken into account. Therefore, when merging the corpora it is very important to take into consideration the types of negations (syntactic, morphological, lexical) and merge only those corpora completely annotated with the same types to avoid the system being trained with false negatives.

3. The way in which each corpus was tokenized is not specified in most of the cases, whereas annotations are carried out at token level. If we would like to expand the corpora, we would need to have more technical information available to make sure that the annotations are compatible. If we want to run the negation processing system on new test data, we need to make sure that in both training and test data, the tokenization should be the same.

4. The annotation formats are different. This problem could be resolved by reconverting the corpora annotations, but the process is time consuming. The different corpora must be pre-processed in a different way in order to obtain the information related to negation and to represent it according to the input format for the machine learning system.

5. Finally, the annotation guidelines are different. This is a great problem because it means that the criteria used during the annotation process are different. For example, some authors include the subject within the scope of negation and others leave it out. If the training examples are contradictory, the system will not be reliable.

5.1.4 Conclusions and selected corpus

As our analysis and discussion show, the main problem for merging the corpora is related to the non-existence of a common scheme and annotation guidelines. In view to future work, the annotation of negation should be standardized in the same way as has been done for other annotation tasks such as semantic role labeling. Moreover, there are languages for which the existence of corpora annotated with negation is limited, for example Spanish, Swedish, Dutch, Japanese, Chinese, German and Italian, and there are even languages for which no corpora have been annotated with this information, such as Arabic, French or Russian. This is a sign that we must continue working to try to advance in the study of this phenomenon that is so important to the development of systems that approach human understanding.

We have analyzed whether it is possible to make these corpora compatible. First, we focus on overall negation processing tasks (Table 5.1).

For **negation cue detection**, we could merge the corpora that have been completely annotated for the same type of negation (Table A.5). Taking this into account, we could merge BioScope, ConanDoyle-neg, SFU Review$_{EN}$, NEG-DrugDDI, NegDDI-DrugBank, Deep Tutor Negation and SOCC corpora for the identification of syntactic cues in English; NEG-DrugDDI and NegDDI-DrugBank for morphological cues detection; and BioScope, NEG-DrugDDI, NegDDI-DrugBank and Deep Tutor Negation for lexical cues identification. For Spanish, UAM Spanish Treebank, SFU Review$_{SP}$-NEG, UHU-HUVR and IULA Spanish Clinical Record corpora could be merged for syntactic cues detection. UHU-HUVR and IULA Spanish Clinical Record corpora could also be merge for the identification of lexical cues. However, we can not merge corpora in their actual form because, as we have analyzed before, the annotation formats and guidelines are different. It would be necessary to pre-process the corpora in order to get negation cues information and convert them into a common format. However, one more problem should be surmounted because each corpus has been tokenized in a different way. The most difficult task would be to establish a correspondence between each new token and its initial annotation. Suppose a corpus with Example (136) that corresponds to the following list of tokens: “I”, “don’t”, “like”, “meat”, “.”, in which the third token (“don’t”) is a negation cue. Suppose that the new tokenizer returns as list of tokens the following: “I”, “do”, “n’t”, “like”, “meat”, “.”. How do we know which token is the negation cue in the new tokenization list? This can be further complicated in sentences with multiple markers in which not all act as negation cues (Example (137)), with non-contiguous cues (Example (137)) or with multi-words expressions (Example (138)). An additional problem is that most existing annotation schemes do not account for the complexity of the linguistic structures used to express negation, so most of them do not differentiate between simple, continuous and discontinuous negation cues. The annotation of these structures needs to be unified.

136. I **don't** like meat.

137. El final del libro **no** te aporta **nada**, **no** añade **nada** nuevo, no crees?

    The end of the book doesn't give you anything, it doesn't add anything new, didn't you?

138. He is a well-known author but he is **not** the best for me.

For **scope identification**, we would have the same problems as for cue detection, but we would also have to solve additional aspects, such as unifying the inclusion or not of the subject and the cue within the scope, and unifying the length of the scope to the largest or shortest syntactic unit. We would have to use the same syntactic analyzer to process the texts and convert the manual annotations into annotations that follow the new standards in relation to inclusion of subject and length of scope. For **event extraction** the main problem is that in most of the corpora events have only been annotated if they are clinically or biologically relevant, so not all negated events are annotated. Finally, for **focus detection**, we would be able to merge PropBank Focus (PB-FOC), Deep Tutor Negation and SOCC English corpora.

Once the problems related to negation processing had been solved, it would be possible to merge corpora for specific tasks (Table 5.2). This would require a study of the annotation schemes, the labels used and their values. For example, for sentiment analysis, we would have to make sure that the corpora use the same polarity labels. If not, we would have to analyze the meaning of the labels, define a new tag set and convert the real labels of these corpora to those of the new tag set.

In view that we can not merge corpora in their actual form, for the development of the negation processing system we have selected the SFU Review$_{SP}$-NEG corpus (Jiménez-Zafra et al., 2018) due to the following reasons. Most existing annotation schemes for Spanish do not account for the complexity of the linguistic structures used to express negation and the SFU Review$_{SP}$-NEG corpus is an exception to this. For its annotation it has been defined a reliable and comprehensive typology of language-dependent negation patterns. Moreover, it consists of reviews that belongs to 8 different domains, which implies a greater lexical richness that is of interest for the development of the negation processing system. In addition, it is widely used for sentiment analysis, task on which we are going to evaluate the role of negation. Finally, there is an English version of this corpus annotated with negation, the SFU Review Corpus with negation and speculation annotations (Konstantinova et al., 2012), which would allow to study the difficulty of this phenomenon in both languages.

5.2 System architecture

We propose a supervised machine learning system that model negation processing as two consecutive classification tasks (Figure 5.1). The first one for detecting negation cues and the second one for determining the scope of the identified negation cues. As in previous works, we approach each of them as a sequence labelling task (Morante et al., 2008). Sequence labelling tasks involve the assignment of a label to each member of the sequence. In our case, each sequence is a sentence consisting of a set of tokens. Therefore, in the first task, each token is classified as being at the beginning of a negation cue (B-cue), inside a negation cue (I-cue), or outside it (O-cue), using the BIO encoding (Ramshaw & Marcus, 1995). This also makes it possible to find continuous and discontinuous negation cues. In the second task, for each negation cue detected in the first task, the system determines for each token of the sentence whether it is the beginning of the scope of the cue (B-scope), the continuation of the scope (I-scope) or if it does not belong to the scope (O-scope). Regarding the classifier, we choose the 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., 2012; Reitan et al., 2015; Loharja et al., 2018). 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] The corpus used for training and testing the system is the SFU Review$_{SP}$-NEG corpus (Jiménez-Zafra et al., 2018), whose choice has

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

been justified in Section 5.1: "*Data selection*", and selected features are described in Section 5.3: "*Experiments*".

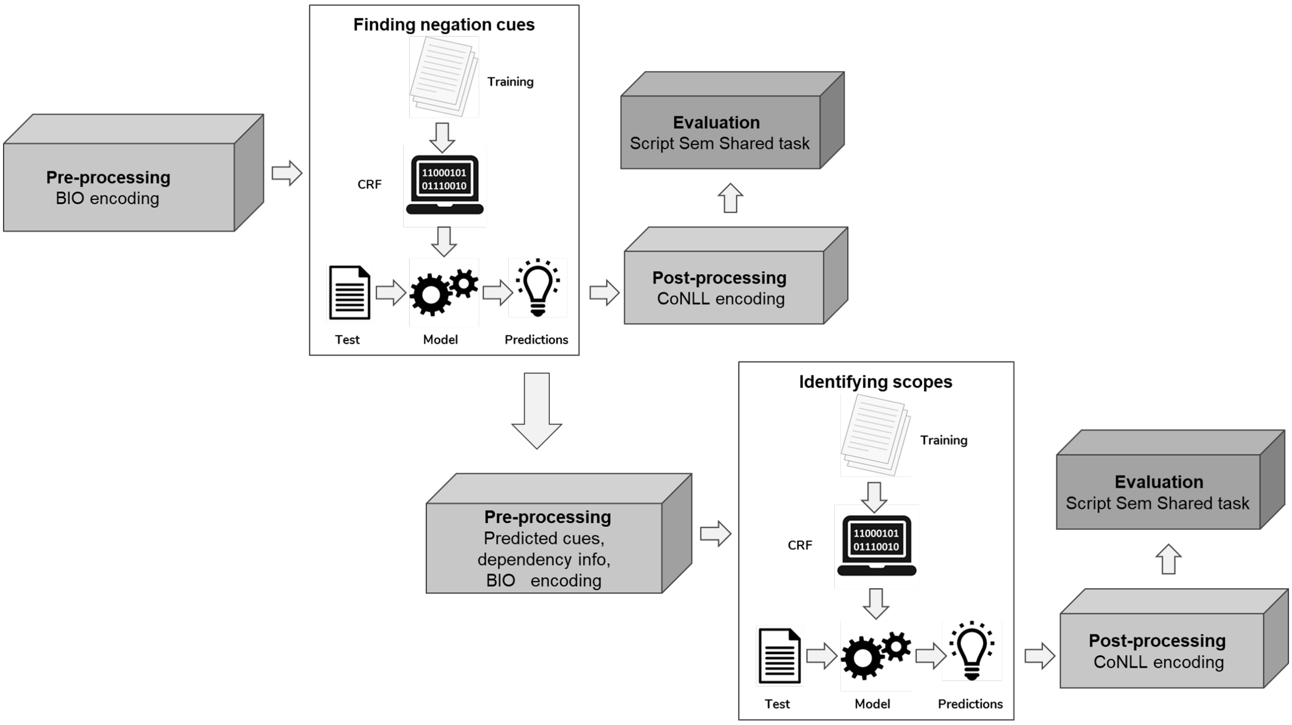

Figure 5.1: Architecture of the negation processing system.

In the following, some examples are provided to help in understanding how the system works. Let's suppose that the system receives as input the sentences of Examples (139) and (140).

139. [**No** voy a volver **en absoluto**].

 I am not going back at all.

140. [**No**[1] soy alta][1], aunque [**tampoco**[2] soy un pitufo][2].

 I'm not tall, but I'm not a smurf either.

In the first phase, the system predicts the negation cues. It classifies each of the tokens of each of the sentences as is represented in Figure 5.2 and Figure 5.3. On the one hand, the sentence of Example (139) has a discontinuous negation cue, "No-en absoluto" *[not-at all]* and the system assigns the label B-cue to the first token, *No*, the label I-cue to the fifth and sixth tokens, *en* and *absoluto*, and the label O-cue to the rest of tokens. On the other hand, the sentence of

Example (140) contains two simple negation cues, that is, two cues composed of a single token. Therefore, the system classifies the first token, *No*, and the sixth token, *tampoco*, with the label B-cue, and the rest of tokens with the label O-cue.

```
ID  TOKEN        OUTPUT
1   No           B-cue
2   voy          O-cue
3   a            O-cue
4   volver       O-cue
5   en           I-cue
6   absoluto     I-cue
7   .            O-cue
```

Figure 5.2: System output for negation cues detection task - Predicting a discontinuous cue in the sentence of Example (139).

```
ID  TOKEN        OUTPUT
1   No           B-cue
2   soy          O-cue
3   alta         O-cue
4   ,            O-cue
5   aunque       O-cue
6   tampoco      B-cue
7   soy          O-cue
8   un           O-cue
9   pitufo       O-cue
10  .            O-cue
```

Figure 5.3: System output for negation cues detection task - Predicting two simple cues in the sentence of Example (140).

In the second phase, another classifier determines the scope of the predicted negation cues. Therefore, for each sentence in which the first classifier has detected negation cues, the second classifier predicts the scope of each negation cue. The sentence of Example (139) has one negation cue, the discontinuous cue "No-en absoluto" *[not-at all]*. Consequently, the system predicts the scope for it and assigns the label B-scope to the first token of the scope, the token *No*, the label I-scope, to the tokens that are inside the scope (*voy*, *a*, *volver*, *en*, *absoluto*), and the label O-scope to the last token, because it is outside the scope of negation (Figure 5.4). The sentence of Example (140) has two negation cues. Therefore, the system has to predict two scopes (Figure 5.5). For the first predicted cue, the token *No* with ID=1, the system predicts as scope the tokens *No*, *soy* and *alta*. For the second predicted cue, the token *tampoco* with ID=6, it determines as scope the tokens *tampoco*, *soy*, *un* and *pitufo*.

ID	TOKEN	ID_NEG_CUE	OUTPUT
1	No	1#5#6	B-scope
2	voy	1#5#6	I-scope
3	a	1#5#6	I-scope
4	volver	1#5#6	I-scope
5	en	1#5#6	I-scope
6	absoluto	1#5#6	I-scope
7	.	1#5#6	O-scope

Figure 5.4: System output determining the scope of the predicted discontinuous cue in the sentence of Example (139).

ID	TOKEN	ID_NEG_CUE	OUTPUT	ID	TOKEN	ID_NEG_CUE	OUTPUT
1	No	1	B-scope	1	No	6	O-scope
2	soy	1	I-scope	2	soy	6	O-scope
3	alta	1	I-scope	3	alta	6	O-scope
4	,	1	O-scope	4	,	6	O-scope
5	aunque	1	O-scope	5	aunque	6	O-scope
6	tampoco	1	O-scope	6	tampoco	6	B-scope
7	soy	1	O-scope	7	soy	6	I-scope
8	un	1	O-scope	8	un	6	I-scope
9	pitufo	1	O-scope	9	pitufo	6	I-scope
10	.	1	O-scope	10	.	6	O-scope

Figure 5.5: System output determining the scope of the two predicted simple cues in the sentence of Example (140).

5.3 Experiments

As it has been previously mentioned, the corpus used for the experimentation is the SFU Review$_{SP}$-NEG corpus (Jiménez-Zafra et al., 2018). This corpus is in XML format (Figure 4.3), but in order to use the evaluation script released by the *SEM-2012 Shared Task[6], which is widely used for the evaluation of negation cues detection, scope identification and event recognition, we convert the data to CoNLL format (Buchholz & Marsi, 2006). Very briefly, in CoNLL format, each line corresponds to a token, each annotation (lemma, PoS-tag, etc.) is provided in a column, and empty lines indicate end of sentence. Figure 5.6, Figure 5.7 and Figure 5.8 show an example of a corpus sentence without negation, a sentence with a negation and a sentence with two negations, respectively, after conversion to CoNLL format. Each line corresponds to a token and each annotation is provided in a column. The content of the columns given is:

- Column 1: domain_filename

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

- Column 2: sentence number within domain_filename
- Column 3: token number within sentence
- Column 4: word
- Column 5: lemma
- Column 6: part-of-speech
- Column 7: part-of-speech type
- Columns 8 to last:
  - If the sentence has no negations, column 8 has a "***" value and there are no more columns.
  - If the sentence has negations, the annotation for each negation is provided in three columns. The first column contains the word that belongs to the negation cue, the second contains the word that belongs to the scope of the negation cue, and the third column contains the word that is the negated event or property.

On the one hand, the sentence of Figure 5.6 has no negations and, consequently, the 8th column is "***" for all tokens. On the other hand, the sentence of Figure 5.7 contains one negation cue and the columns for negation components start at the 8th column. Finally, the sentence of Figure 5.8 has two negations. The information for the first negation is provided in columns 8-10, and for the second in columns 11-13. If the token is part of a negation cue, scope or event, the corresponding column will have as value the word form of the token, if it is not it will contain a "-".

coches_yes_5_2	2	1	Quedé	quedar	vmis1s0	main	***
coches_yes_5_2	2	2	prendado	prendar	vmp00sm	main	***
coches_yes_5_2	2	3	de	de	sps00	preposition	***
coches_yes_5_2	2	4	ese	ese	dd0ms0	demonstrative	***
coches_yes_5_2	2	5	coche	coche	ncms000	common	***
coches_yes_5_2	2	6	.	.	fp	-	***

Figure 5.6: Sentence without negation in CoNLL format.

```
coches_yes_5_2  9   1   Para     para     sps00       preposition -       Para     -
coches_yes_5_2  9   2   mí       mí       pp1cso00    personal    -       mí       -
coches_yes_5_2  9   3   no       no       rn          negative    no      no       -
coches_yes_5_2  9   4   tiene    tener    vmip3s0     main        -       tiene    tiene
coches_yes_5_2  9   5   ningún   ninguno  di0ms0      indefinite  ningún  ningún   -
coches_yes_5_2  9   6   defecto  defecto  ncms000     common      -       defecto  -
coches_yes_5_2  9   7   .        .        fp          -           -       -        -
```

Figure 5.7: Sentence with one negation in CoNLL format.

```
hoteles_no_2_20 2   1   Las           el          da0fp0  article        -     -           -       -   -       -
hoteles_no_2_20 2   2   habitaciones  habitación  ncfp000 common         -     -           -       -   -       -
hoteles_no_2_20 2   3   son           ser         vsip3p0 semiauxiliary  -     -           -       -   -       -
hoteles_no_2_20 2   4   pequeñas      pequeño     aq0fp0  qualificative  -     -           -       -   -       -
hoteles_no_2_20 2   5   ,             ,           fc      -              -     -           -       -   -       -
hoteles_no_2_20 2   6   casi          casi        rg      -              casi  casi        -       -   -       -
hoteles_no_2_20 2   7   no            no          rn      negative       no    no          -       -   -       -
hoteles_no_2_20 2   8   tienen        tener       vmip3p0 main           -     tienen      tienen  -   -       -
hoteles_no_2_20 2   9   camas         cama        ncfp000 common         -     camas       -       -   -       -
hoteles_no_2_20 2   10  de            de          sps00   preposition    -     de          -       -   -       -
hoteles_no_2_20 2   11  matrimonio    matrimonio  ncms000 common         -     matrimonio  -       -   -       -
hoteles_no_2_20 2   12  ,             ,           fc      -              -     -           -       -   -       -
hoteles_no_2_20 2   13  ni            ni          cc      coordinating   -     -           -       ni  ni      -
hoteles_no_2_20 2   14  tienen        tener       vmip3p0 main           -     -           -       -   tienen  tienen
hoteles_no_2_20 2   15  terraza       terraza     ncfs000 common         -     -           -       -   terraza -
hoteles_no_2_20 2   16  .             .           fp      -              -     -           -       -   -       -
```

Figure 5.8: Sentence with two negations in CoNLL format.

Once the corpus was converted to CoNLL format, it was randomly split into three set: training, development and test. 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). These splits in CoNLL format, but with information only about the negation cues, were 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) (see Chapter 7: "*NEGES: Workshop on Negation in Spanish*").

For finding negation cues, we pre-process the data to convert them to BIO encoding (Figure 5.9), which is the format expected by the system. For identifying scopes, we pre-process the reviews to add dependency relations using Freeling (Padró & Stanilovsky, 2012), which was the tool used to annotate the original version of the corpus with PoS-tags, PoS-types and lemmas, and convert them to BIO encoding taking into account the negation cues (Figure 5.10).

```
hoteles_no_2_20 2   1   Las           el          da0fp0  article        O-Cue
hoteles_no_2_20 2   2   habitaciones  habitación  ncfp000 common         O-Cue
hoteles_no_2_20 2   3   son           ser         vsip3p0 semiauxiliary  O-Cue
hoteles_no_2_20 2   4   pequeñas      pequeño     aq0fp0  qualificative  O-Cue
hoteles_no_2_20 2   5   ,             ,           fc      -              O-Cue
hoteles_no_2_20 2   6   casi          casi        rg      -              B-Cue
hoteles_no_2_20 2   7   no            no          rn      negative       I-Cue
hoteles_no_2_20 2   8   tienen        tener       vmip3p0 main           O-Cue
hoteles_no_2_20 2   9   camas         cama        ncfp000 common         O-Cue
hoteles_no_2_20 2   10  de            de          sps00   preposition    O-Cue
hoteles_no_2_20 2   11  matrimonio    matrimonio  ncms000 common         O-Cue
hoteles_no_2_20 2   12  ,             ,           fc      -              O-Cue
hoteles_no_2_20 2   13  ni            ni          cc      coordinating   B-Cue
hoteles_no_2_20 2   14  tienen        tener       vmip3p0 main           O-Cue
hoteles_no_2_20 2   15  terraza       terraza     ncfs000 common         O-Cue
hoteles_no_2_20 2   16  .             .           fp      -              O-Cue
```

Figure 5.9: Sentence of Figure 5.8 pre-processed for negation cues detection phase.

hoteles_no_2_20	2	6#7	casi#no	casi#no	rg#rn	-#negative	mod#mod	8#8	1	Las	el	da0fp0	article	spec	2	O-scope
hoteles_no_2_20	2	6#7	casi#no	casi#no	rg#rn	-#negative	mod#mod	8#8	2	habitaciones	habitación	ncfp000	common	suj	3	O-scope
hoteles_no_2_20	2	6#7	casi#no	casi#no	rg#rn	-#negative	mod#mod	8#8	3	son	ser	vsip3p0	semiauxiliary	sentence	0	O-scope
hoteles_no_2_20	2	6#7	casi#no	casi#no	rg#rn	-#negative	mod#mod	8#8	4	pequeñas	pequeño	aq0fp0	qualificative	atr	3	O-scope
hoteles_no_2_20	2	6#7	casi#no	casi#no	rg#rn	-#negative	mod#mod	8#8	5	,	,	fc	-	f	3	O-scope
hoteles_no_2_20	2	6#7	casi#no	casi#no	rg#rn	-#negative	mod#mod	8#8	6	casi	casi	rg	-	mod	8	B-scope
hoteles_no_2_20	2	6#7	casi#no	casi#no	rg#rn	-#negative	mod#mod	8#8	7	no	no	rn	negative	mod	8	I-scope
hoteles_no_2_20	2	6#7	casi#no	casi#no	rg#rn	-#negative	mod#mod	8#8	8	tienen	tener	vmip3p0	main	S	3	I-scope
hoteles_no_2_20	2	6#7	casi#no	casi#no	rg#rn	-#negative	mod#mod	8#8	9	camas	cama	ncfp000	common	cd	8	I-scope
hoteles_no_2_20	2	6#7	casi#no	casi#no	rg#rn	-#negative	mod#mod	8#8	10	de	de	sps00	preposition	sp	9	I-scope
hoteles_no_2_20	2	6#7	casi#no	casi#no	rg#rn	-#negative	mod#mod	8#8	11	matrimonio	matrimonio	ncms000	common	sn	10	I-scope
hoteles_no_2_20	2	6#7	casi#no	casi#no	rg#rn	-#negative	mod#mod	8#8	12	,	,	fc	-	f	3	O-scope
hoteles_no_2_20	2	6#7	casi#no	casi#no	rg#rn	-#negative	mod#mod	8#8	13	ni	ni	cc	coordinating	coord	3	O-scope
hoteles_no_2_20	2	6#7	casi#no	casi#no	rg#rn	-#negative	mod#mod	8#8	14	tienen	tener	vmip3p0	main	S	3	O-scope
hoteles_no_2_20	2	6#7	casi#no	casi#no	rg#rn	-#negative	mod#mod	8#8	15	terraza	terraza	ncfs000	common	cd	14	O-scope
hoteles_no_2_20	2	6#7	casi#no	casi#no	rg#rn	-#negative	mod#mod	8#8	16	.	.	fp	-	f	3	O-scope
hoteles_no_2_20	2	13	ni	ni	cc	coordinating	coord	3	1	Las	el	da0fp0	article	spec	2	O-scope
hoteles_no_2_20	2	13	ni	ni	cc	coordinating	coord	3	2	habitaciones	habitación	ncfp000	common	suj	3	O-scope
hoteles_no_2_20	2	13	ni	ni	cc	coordinating	coord	3	3	son	ser	vsip3p0	semiauxiliary	sentence	0	O-scope
hoteles_no_2_20	2	13	ni	ni	cc	coordinating	coord	3	4	pequeñas	pequeño	aq0fp0	qualificative	atr	3	O-scope
hoteles_no_2_20	2	13	ni	ni	cc	coordinating	coord	3	5	,	,	fc	-	f	3	O-scope
hoteles_no_2_20	2	13	ni	ni	cc	coordinating	coord	3	6	casi	casi	rg	-	mod	8	O-scope
hoteles_no_2_20	2	13	ni	ni	cc	coordinating	coord	3	7	no	no	rn	negative	mod	8	O-scope
hoteles_no_2_20	2	13	ni	ni	cc	coordinating	coord	3	8	tienen	tener	vmip3p0	main	S	3	O-scope
hoteles_no_2_20	2	13	ni	ni	cc	coordinating	coord	3	9	camas	cama	ncfp000	common	cd	8	O-scope
hoteles_no_2_20	2	13	ni	ni	cc	coordinating	coord	3	10	de	de	sps00	preposition	sp	9	O-scope
hoteles_no_2_20	2	13	ni	ni	cc	coordinating	coord	3	11	matrimonio	matrimonio	ncms000	common	sn	10	O-scope
hoteles_no_2_20	2	13	ni	ni	cc	coordinating	coord	3	12	,	,	fc	-	f	3	O-scope
hoteles_no_2_20	2	13	ni	ni	cc	coordinating	coord	3	13	ni	ni	cc	coordinating	coord	3	B-scope
hoteles_no_2_20	2	13	ni	ni	cc	coordinating	coord	3	14	tienen	tener	vmip3p0	main	S	3	I-scope
hoteles_no_2_20	2	13	ni	ni	cc	coordinating	coord	3	15	terraza	terraza	ncfs000	common	cd	14	I-scope
hoteles_no_2_20	2	13	ni	ni	cc	coordinating	coord	3	16	.	.	fp	-	f	3	O-scope

Figure 5.10: Sentence of Figure 5.8 pre-processed for scope identification phase.

The experimentation was organized in the following phases:

- Phase A: Negation cues detection

    1. We train with the train split and select features based on results with the development set.
    2. We report results about the prediction of negation cues using the test set.
    3. We compare results with those of state-of-the-art.

- Phase B: Scope identification

    1. We train with the train split and select features based on results with the development set.
    2. We report results about the identification of scopes on the test set using the predicted cues in Phase A-2.
    3. We report results about the identification of scopes on the test set using the gold cues.
    4. We develop some baseline systems and compare results with those of the proposed system.

Table 5.3: Features used to train the CRF classifiers to detect negation cues and scopes (two separate classifiers). We use t to refer to the token to be predicted, and m to the negation cue.

	Name	Description	Marker?	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 m	✗	✓
34	location	Location of current t with respect to m (before, inside or after)	✗	✓
35	distance	Number of tokens between t and m	✗	✓
36	chain_pos_f	Sequence of fine part-of-speech tags between t and m	✗	✓
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 m 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 m	✗	✓
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 m and vice versa	✗	✓
50, 51	path_types	Dependency types in the syntactic path from t to m 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	✗	✓

The feature set is inspired by the work of Cruz et al. (2016), who train their system on the SFU Review corpus. We decided to use similar features because the SFU Review$_{\text{SP}}$-NEG corpus (Spanish) is the comparable version of the SFU Review corpus (English) (Konstantinova et al., 2012). The final feature set used for each classifier is presented in Table 5.3 and the feature selection process is described below.

5.3.1 Features for negation cues detection

In first place, we experiment with the lemma and PoS-tag of the token in focus, boolean tag 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). We find that the most useful features are lemmas and part-of-speech tags, according to the chi-squared feature selection method. Therefore, we discard the rest of features and conduct experiments to find out which is the optimal window for which lemma and PoS-tags features should be added. We decide to use as features the lemma and PoS-tags of the current token as well as 7 tokens before and after (31 features). These features are positional, we do not use a bag-of-words representation.

Additionally, we use a binary flag (known_cue) to indicate whether the token was seen as part of a negation cue in the training instances. 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), 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).

5.3.2 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 (1-4), location of the token respect the cue (5) (before, inside or after), distance in number of tokens between the cue and the current token (6), chain of PoS-tags and chain of types between the the cue and the token (7-8), lemma and PoS-tags of the token to the left and right of the token in focus (9-12), relative position of the cue and the token in the sentence (13-14), dependency relation and direction (head or dependent) between the token and the cue (15-16), PoS-tags of the first and second order syntactic heads of the token (17-18), whether the token is ancestor of the token and vice versa (19-20), dependency shortest path from the token in focus to the cue and vice versa (21-22), dependency shortest path from the token in focus to the cue but including direction (up or down) (23), and length of the short path between the token and the cue (24).

During the feature tuning process, we discover that the least informative features are *dep_rel* (15-16), *is_ancestor* (19-20), *heads* second order (18) and *path_length* (24). Therefore, we conduct experiments removing all these features and the two least informative[7] (16, 20), but the results do not improve the initial experiment. Consequently, we decide to select the initial set (24 features) as features for reporting results with the test set.

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

5.4 Results

The results reported for identifying cues and detecting scopes have been obtained over the test set using the features previously described. Baseline models have also been considered in both phases to compare with the results obtained by the proposed system. The following subsections present the evaluation measures used and detail the results for the negation cue detection and scope identification tasks compared to baseline systems.

5.4.1 Evaluation measures

The evaluation script used to evaluate both tasks is the same as the one used to evaluate the *SEM 2012 Shared Task: "Resolving the Scope and Focus of Negation"[8] (Morante & Blanco, 2012a), which reports results in terms of 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.

$$P = \frac{TP}{TP + FP} \tag{5.1}$$

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

$$R = \frac{TP}{TP + FN} \tag{5.2}$$

$$F1 = \frac{2PR}{P + R} \tag{5.3}$$

5.4.2 Negation cues detection results

Table 5.4 provides experimental results for negation cue detection. Our system is in general accurate: precision is between 83% and 99% in the different domains, 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 5.5 we provide an error analysis.

Table 5.4: System results on the test set for negation cues detection.

	P	R	F1
Books	**83.47**	80.16	**81.78**
Cars	93.44	83.82	88.37
Cell phones	90.57	84.21	87.27
Computers	89.29	**92.59**	90.91
Hotels	**98.11**	88.14	**92.86**
Movies	90.79	84.66	87.62
Music	95.83	**79.31**	86.79
Washing machines	94.44	**73.91**	82.92
All	91.99	83.35	87.32

At the time we developed our system, the state of the art comprised the systems presented at NEGES 2018 (Jiménez-Zafra et al., 2019a). However, with the new edition of the task in NEGES 2019 (Jiménez-Zafra et al., 2019b), new systems have appeared with which to compare our results. 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

[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)

Focus of Negation"[10] (Morante & Blanco, 2012a). In Table 5.5 we present the overall results of our system and those of the state of the art in chronological order.

Taking as baseline the results of existing systems we can say that our results (87.32 F1) 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. Regarding the approaches followed to detect negation cues, machine learning and deep learning algorithms are the selected. UNED and Aspie96 conduct experiments using deep learning, while UPC, our system, CLiC and IBI use machine learning algorithms[11], confirming that the use of CRF algorithm provides the best results in this task.

Table 5.5: System results on the test set for negation cues detection compared to existing results.

	P	R	F1
UNED (Fabregat et al., 2018)	79.45	59.58	67.97
UPC (Loharja et al., 2018)	91.48	82.18	86.45
Our results	**91.99**	**83.35**	**87.32**
Aspie96 (Giudice, 2019b)	18.80	28.34	22.58
CLiC (Beltrán & González, 2019)	89.67	79.40	84.09
IBI (Domínguez-Mas et al., 2019)	91.22	72.16	80.50
UNED (Fabregat et al., 2019)	91.82	75.98	82.99

5.4.3 Scope identification results

For scope identification, comparison with other scope detection systems is not possible because ours are the first results. Therefore, we calculate two baselines:

1. From cue to end of sentence: scope is identified as all tokens from the cue to the token previous to the end of the sentence.

2. From cue to first punctuation mark: scope labels are assigned to all the tokens from the cue to the token previous to the first punctuation mark.

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

[11] A complete description of the approaches followed by each team is provided in Chapter 7: "*NEGES: Workshop on Negation in Spanish*".

For example, for the sentence "No soy alta, aunque tampoco soy un pitufo" *[I'm not tall, but I'm not a smurf either.]*, the first method will detect as scope of the negation cue *No* all the tokens of the sentence except the full stop (Example 141), while the second method will tag as scope the tokens *No*, *soy* and *alta* (Example 142). However, for the negation cue *tampoco*, both systems will detect as scope the same set of tokens: *tampoco*, *soy*, *un* and *pitufo*.

141. [**No**^{1} soy alta, aunque [**tampoco**^{2} soy un pitufo]^2]^1.
    I'm not tall, but I'm not a smurf either.

142. [**No**^{1} soy alta]^1, aunque [**tampoco**^{2} soy un pitufo]^2.
    I'm not tall, but I'm not a smurf either.

Table 5.6 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 get the upper bound of the system, and with *predicted cues* (Table 5.7). 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 study in Section 5.5, what types of scopes have been the most difficult to predict.

The results obtained by our system, with an F1 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 the most optimal for Spanish. We perform an error analysis in order to detect where does the system fail. 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.

Table 5.6: Baseline results on the test set for scope detection using predicted cues.

	From cue to end of sentence			From cue to first punctuation mark		
	P	R	F1	P	R	F1
Books	55.56	19.84	29.24	70.37	37.70	49.10
Cars	71.43	14.71	24.40	87.10	39.71	54.55
Cell phones	64.29	15.79	25.35	83.87	45.61	59.09
Computers	68.96	24.69	36.36	79.07	41.98	54.84
Hotels	88.89	13.56	23.53	94.44	28.81	44.15
Movies	76.27	27.61	40.54	84.62	47.24	60.63
Music	84.21	18.39	30.19	91.89	39.08	54.84
Washing machines	83.33	21.74	34.48	90.00	39.16	54.55
All	74.12	**19.54**	30.51	85.17	**39.91**	53.97

Table 5.7: System results on the test set for scope detection.

	Scope (gold cues)			Scope (predicted cues)		
	P	R	F1	P	R	F1
Books	100	67.06	80.28	**79.38**	61.11	**69.06**
Cars	100	61.76	76.36	90.48	**52.88**	69.09
Cell phones	100	68.42	81.25	87.50	61.4	72.16
Computers	100	**61.73**	**76.34**	84.75	61.73	71.43
Hotels	100	71.19	83.17	**97.50**	66.1	**78.79**
Movies	100	72.39	**83.98**	88.98	**69.33**	77.94
Music	100	66.67	80.00	94.34	57.47	71.43
Washing machines	100	**72.46**	84.03	93.75	65.22	76.92
All	100	67.71	80.68	**89.59**	**61.91**	**73.35**

5.5 Error analysis

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

5.5.1 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 the sentence of Example (143),[12] the system has identified the negation cue *no* instead of *ya no*.

143. **Ya** **<u>no</u>** 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[13] or with a frequency of occurrence between 1 and 2[14], 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 the sentence of Example (144), the system predicts *no* as negation cue, rather than the discontinuous cue *no muy*.

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

[13] 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*.

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

144. 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 (Example 143). Moreover, in some cases it is wrongly identified as negation cue when it is part of a contrasting structure (Example 145). 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 correctly identified by the system in most cases, not even in simple sentences such as the one of Example (146).

145. No exige sino aquello que se le da.

    He demands only that which is given to him.

146. **Jamás** compréis un ordenador de marca.

    Never buy a branded computer.

147. 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 (Example 143 and Example 144), ii) cues wrongly identified as negation cues (Example 147) and iii) cues identified as negation cues when they are part of a contrasting structure (Example 145). 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.

5.5.2 Negation scopes

The error analysis of scopes[15] is based on predictions of the scope processing module using gold cues. We can make several general observations.

1. 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.

2. The system never includes in the scope punctuation that signals the end of the sentence.

3. 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.

4. The system includes generally all tokens of a syntactic phrase in the scope, so it does not split phrases. Example (148) 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. In Example (150) and Example (149) the system excludes from the scope the second element of the coordination.

5. Except for a few cases with the cue *ningún*, as in Example (151), 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 (Example 152 and Example 153).

148. {[**No** me lo pensaría dos} veces].

    I would not think twice.

149. [{**No** paso a calificar las prestaciones} y características del móvil]...

    I do not go on to qualify the performance and characteristics of the mobile....

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

150. la batería ... [{**no** dura más de un día} y medio]... *the battery... does not have a life of more than a day and a half...*

151. ... a mi gusto no cuenta [con **ningún** temazo]...
    ... I don't think there's any hit...

152. ... [{**ni** lo que aspira aún a ser]...
    ... nor what he still aspires to be...

153. [{**Ni** quienes la vieron lo saben cómo fue]...
    Even those who saw her do not know how it was...

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 of negation 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 nominal or pronominal (Example 154),[16] or the adverbial complements of the verb (Example 155), when gold does include them, or including them when gold does not (Example 156 and Example 157). 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 Example (154) the gold annotations include the subject in the scope, whereas in Example (156) the subject is not included.

154. ... el motor ... [que además {**no** es el que menos gasta}]...
    ... the engine ... that also is not the one that spends less...

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

155. Vamos, [por 11900 euros {yo **no** me lo compraba}].

    Well, for 11900 euros, I wouldn't buy it.

156. Los plásticos resultan demasiado evidentes y {la tapicería [**no** es nada del otro mundo]}.

    Plastics are too obvious and upholstery is nothing new.

157. {En mi opinión, [no lo compréis]}...

    In my opinion, do not buy it...

The errors due to wrong system predictions at the end of the scope are mostly due to the system adding complements when gold does not. In Example (158) the system adds to the scope a clause that acts as causal complement of the verb, in Example (159) it adds a relative clause that is a complement of the direct object of the negated verb, and in Example (160), 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 token previous to the final punctuation mark. However, there are also some errors due to the system shortening the scope, as in Example (161), where the second element of the coordinated adjectival phrase is not included, or Example (162), where the complement of the noun *cable* is not included.

158. Por cierto, [{no lo probé] porque en ningún sitio lo tenían}.

    By the way, I did not prove it because nowhere did they have it.

159. ... [{que **no** se adapta a la caja de cambio] que lleva}.

    ... which does not adapt to the gearbox that it carries.

160. La pila de ropa [{**sin** lavar] sigue subiendo}.

    The pile of unwashed clothes continues to rise.

161. {[**No** me sentí **ni** libre} **ni** poderoso] en aquella suntuosa mañana.

    I felt neither free nor powerful in that sumptuous morning.

162. ... [{**sin** cable} para el pc]...

    ... without cable for pc...

Errors at the beginning and at the end of the scope are less frequent. In Example (163) 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 Example (164) the system excludes the subject of the verb affected by the cue and adds the quotation marks at the end. In Example (165) the system excludes the subject, but includes a causal complement at the end.

163. ... [(que encima, según Opel, {**no** es un fallo}])...

   ... (which, according to Opel, is not a fault)...

164. "No trates de arreglar [lo que {**no** está descompuesto]"}.

   "Do not try to fix what is not broken".

165. [Los antiguos PC, {**no** metían casi ruido] debido a la carencia de ventiladores}...

   The old PCs did not make much noise due to the lack of fans...

In sum, it seems that some errors might be due to inconsistencies in the annotations of 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.

5.6 Conclusion

In this chapter it has been presented a machine learning negation processing system for Spanish which addresses two tasks: negation cues detection and scope identification. The system has been trained and tested on a corpus of product reviews, the SFU Review$_{SP}$-NEG corpus (Jiménez-Zafra et al., 2018). The choice of this corpus has been justified after an exhaustive analysis of the existing corpora. Although the system focuses on the processing of negation in

Spanish, we have analyzed all the corpora presented in Subsection 2.1.3: "*Corpora annotated with negation*" of Chapter 2: "*Background*", as it may be useful for the scientific community to advance in the study of this phenomenon in other languages. After the analysis, we conclude that the lack of a standard annotation scheme and guidelines as well as the lack of large annotated corpora make it difficult to progress in the treatment of negation. As future work, the community should work on the standardization of negation as has been done for other well established tasks like semantic role labelling and parsing. A robust and precise annotation scheme should be defined for the different elements that represent the phenomenon of negation (cue, scope, negated event and focus) and researchers should work together to define common annotation guidelines.

For cue identification the system outperforms state-of-the-art results, while for scope detection we provide the first experimental results. A qualitative error analysis has shown that 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 is also 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*. Regarding 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 capture 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.

In Chapter 6: "*Applying negation to improve Spanish sentiment analysis*", we integrate the negation processing system presented in this chapter into a sentiment analysis system in order to show the importance of the development of accurate negation processing systems for Natural Language Processing tasks.

Chapter 6

Applying negation to improve Spanish sentiment analysis

Negation is a complex phenomenon in natural language that can change the polarity of a sentence, creating an opposition between the positive and negative counterparts (L. R. Horn, 1989). It is a primarily 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 systems (Blanco & Moldovan, 2014).

When applying negation to a specific task, the first step should be to process negation with an accurate system and, second, to use the output to improve the task in which negation has a key role, for example, information retrieval (Liddy et al., 2000), information extraction (Savova et al., 2010), machine translation (Baker et al., 2012) or sentiment analysis (Liu, 2015).

In the context of sentiment analysis, accurate negation identification is one of the most important tasks. In order to correctly interpret the sentiment value of a particular expression, it is imperative to identify whether it is in the scope of negation. While much of the work on negation detection has been focused on English (see Chapter 2: "*Backgroung*" - Subsection 2.1.2.1: "*Negation processing in English*"), the accurate identification of negation in other languages is a necessity. This doctoral thesis focuses on the study of this phenomenon in Spanish, and this chapter presents the application to the sentiment analysis task of the system described in Chapter 5: "*A system to process negation in Spanish*". First, we introduce the methodology followed to study the effect of the negation processing system on sentiment analysis. Then, we present the experiments conducted to evaluate the effect of accurate negation detection and the results obtained. After that, error analysis is provided and, finally, we report conclusions.

6.1 Methodology

To study the effect of our Spanish negation processing system on sentiment analysis we must consider three elements: i) the corpus we are going to work with, ii) the specific task on which the experimentation is going to be carried out, and iii) the sentiment analysis system in which we are going to integrate the negation processing system.

6.1.1 Data

The only Spanish corpus annotated with negation that could be used to asses the impact of negation on sentiment analysis is the SFU Review$_{SP}$-NEG corpus (see Subsection 5.1.3: "*Discussion*"). This corpus is one of the results of this doctoral thesis and the details of it have been presented in Chapter 4: "*SFU Review$_{SP}$-NEG: a Spanish corpus annotated with negation*". We use it to evaluate the role of negation in sentiment analysis.

6.1.2 Task

Sentiment analysis includes the study of several sub-tasks, but perhaps the best known are subjectivity detection, polarity classification and emotion recognition (see Subsection 2.2.1: "*Definition of sentiment analysis*"). The SFU Review$_{SP}$-NEG corpus is annotated with overall sentiment of reviews. Therefore, the selected task for carrying out the experimentation is polarity classification at the document-level.

Polarity classification at the document-level is the task which aims to determine the overall sentiment orientation (e.g. positive, negative or neutral) of the opinion given in a text (e.g. a review, a news article, a headline, a tweet). Approaches to this problem can be broadly classified into two types: lexicon-based or machine learning (Taboada et al., 2011; Taboada, 2016). In the following subsection we explain both approaches and present the system chosen to integrate the negation processing system.

6.1.3 Sentiment analysis system

Polarity classification at the document level can be addressed using lexicon-based methods, machine learning algorithms or hybrid approaches.

In lexicon-based methods, dictionaries of positive and negative words are compiled, perhaps adding the strength of the valence (e.g., *accolade* is strongly positive, whereas *accept* is midly 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. Or the system may take into account which words are intensified or negated, changing the value of, respectively, *good, very good* and *not good.*

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 & Inkpen, 2006). In these methods, negation may be picked up by unigrams as a single feature (more instances of *not* in some texts), or by bigrams and trigrams if those are used (*not good; not very good*), but otherwise the method is not able to detect whether an individual phrase is being negated.

In short, in order to study the impact of accurate negation detection in sentiment analysis it is necessary to determine how to efficiently represent negation, in the case of machine learning systems, or how to modify the polarity of the words within the scope of negation in the case of lexicon-based systems. This has been studied in a broader sense in the case of lexicon-based methods. Therefore, we focus on them.

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 (L. R. 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 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, we have used the lexicon-based system known as SO-CAL (Taboada et al., 2011). We have selected it because it is broadly used for the classification of reviews and has been shown to work well on other texts such as blog posts or headlines (Taboada et al., 2011).

6.1.3.1 SO-CAL

SO-CAL, the Semantic Orientation CALculator,[1] is a lexicon-based sentiment analysis system that was specifically designed for customer reviews, but has been shown to work well on other texts such as blog posts or headlines (Taboada et al., 2011). It contains dictionaries[2] 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 polarity of the word it accompanies.

Negation in the standard SO-CAL system for both English and Spanish takes the shift method,

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

[2] https://github.com/sfu-discourse-lab/Sentiment_Analysis_Dictionaries

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 it is 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 cue. If a negation cue 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 cue in the same sentence. The system may continue to track back and keep looking left for negation cues if a "skipped" word is present, such as adjectives, copulas, determiners and certain verbs. Skipped words allow the system to look for cues 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 is to investigate whether a more accurate method for negation detection in Spanish can improve those results.

6.2 Experiments

Experiments are conducted on the corpus developed in this doctoral thesis (see Chapter 4: "*SFU Review$_{SP}$-NEG: a Spanish corpus annotated with negation*"), the SFU Review$_{\mathrm{SP}}$-NEG (Jiménez-Zafra et al., 2018), and they are organized in three phases. In first place, we use the negation processing system presented in Chapter 5: "*A system to process negation in Spanish*" to detect the negation cues present in the texts (Phase A). Later, we use the negation processing system to identify the scopes of the predicted cues (Phase B). We apply in both phases 10-fold cross validation in order to classify the polarity of all the reviews. Finally, we classify the overall sentiment orientation of the reviews in three different ways (Phase C): i) using the SO-CAL system without negation, ii) using the SO-CAL system with the rule-based method that incorporates for negation handling, and iii) using the SO-CAL system by applying our negation processing system, that is, with the output of Phases A and B.

- Phase A: Negation cues detection. Prediction of the negation cues on the texts of the SFU Review$_{SP}$-NEG corpus using the negation processing system (Jiménez-Zafra et al., 2019) and 10-fold cross validation.

- Phase B: Scope identification. Identification of the scopes corresponding to the predicted cues in Phase A.

- 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.

6.3 Results

The objective of this research is to evaluate the effect of the negation processing system developed for Spanish on sentiment classification. The following subsections present the evaluation measures used and report the results for identifying cues, detecting scopes and classifying the sentiment on the reviews of the SFU Review$_{SP}$-NEG corpus.

6.3.1 Evaluation measures

Negation cues detection and scope identification experiments (Phase A and Phase B) have been evaluated with the script used in the *SEM 2012 Shared Task "Resolving the Scope and Focus of Negation" (Morante & Blanco, 2012a), which reports results in terms of Precision (P), Recall (R) and F-score (F1). It has been previously described in Subsection 5.4.1: "*Evaluation measures*" of Chapter 5: "*A system to process negation in Spanish*".

For the evaluation of the sentiment analysis experiments (Phase C), the traditional measures used in text classification have been applied Sebastiani (2002): P, R, F1 and Accuracy (Acc). P, R and F1 have been measured per class (positive and negative) and averaged using macro-average method.

$$P = \frac{TP}{TP + FP} \tag{6.1}$$

$$R = \frac{TP}{TP + FN} \tag{6.2}$$

$$F1 = \frac{2PR}{P + R} \tag{6.3}$$

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{6.4}$$

6.3.2 Negation cues detection and scope identification results

Table 6.1 details the results for negation cues detection and scope resolution on the SFU Review$_{SP}$-NEG corpus using 10-fold cross-validation.

In general, the results for negation cue detection and scope identification are encouraging. We can say that the cue detection module is very precise (92.70%) and provides a good recall

(82.09%), although there are some types of negation cues that are more difficult to identify, such as the negation cue *no*, the discontinuous cues and infrequent cues, as we discussed in Section 5.5: "*Error analysis*" of Chapter 5: "*A system to process negation in Spanish*". On the other hand, the scope identification module is also very precise (90.77%), but its recall is not very high (63.64%). 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. In addition, 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. Moreover, we have also to consider the errors that the classifier introduces in the cue detection phase and which are accumulated in the scope recognition phase.

Table 6.1: Results for negation cues detection (Phase A) and scope identification (Phase B) using 10-fold cross validation.

	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

6.3.3 Sentiment analysis results

Table 6.2 and Table 6.3 present the results for the sentiment classification of the reviews integrating into SO-CAL the negations detected by our system (SO-CAL with negation processing system), compared to those obtained by using the search heuristics implemented in the SO-CAL system (SO-CAL with negation) and a simple baseline model which involves not applying any negation identification (SO-CAL without negation).

As expected, 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, being the configuration of SO-CAL with the negation processing system the one that achieves the best performance (Table 6.3). In general, SO-CAL without negation is biased towards positive polarity, with the F1-score for positive reviews (74.00%) higher than for negatives ones (70%) (Table 6.2). 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 sentiment is more difficult than detecting positive opinions (Ribeiro et al., 2016), for a host of reasons, including a possible universal positivity bias (Boucher & Osgood, 1969).

Table 6.2: Sentiment analysis results per class (positive and negative) 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).

	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 6.3: Overall sentiment analysis results 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).

	SO-CAL without negation				SO-CAL with built-in negation				SO-CAL with negation processing system			
	P	R	F1	Acc.	P	R	F1	Acc.	P	R	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**

6.4 Error analysis

In this section we conduct an analysis of the SO-CAL system using as negation detector the one proposed in this doctoral thesis, compared to SO-CAL's built-in detection system, which simply traces back until it finds a negation cue, without explicitly detecting scope.

The configuration of SO-CAL with our negation processing system achieves the best performance, improving on the baseline by 3.3% and the search heuristic by almost 3% in terms of overall accuracy (Table 6.3). These results can be explained by two factors. Firstly, the negation detector that we propose benefits from a wider list of cues (the built-in search heuristic in SO-CAL include 13 different negation cues while the SFU Review$_{\text{SP}}$-NEG corpus contains 245 different negation cues). Secondly, the scope detection approach goes beyond the window-based heuristic that the SO-CAL system incorporates. In search heuristics based approaches, a cue is identified only if it appears in the predefined list of cues, without taking into consideration whether it is actually acting as such, and the scope is limited to certain parts of the sentence, but in many cases the scope spans beyond the distance of words that the search heuristic methods consider. Below, we illustrate with examples[3] these two situations.

- Case 1: Negation cue predicted by the negation processing system, but not present in the SO-CAL list. For example, in the sentence of Example (166), the negation cue *ningun* has been predicted by the negation detector but it is not present in SO-CAL list. Therefore, *ningun temazo* is correctly classified as negative (-1.5 points) by SO-CAL when we integrate our Spanish negation detector, but with the heuristic that it incorporates by default it is incorrectly classified as positive (3.0 points).

166. 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...*

    a. SO-CAL with built-in negation:

       temazo 3.0 = 3.0

[3]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.

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 our 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 Subsection 6.1.3.1). The sentence in Example (167) is correctly classified as negative when we integrate our 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. However, *una* is not in the skipped list either and therefore the sentence is incorrectly classified as positive.

167. 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

However, we also find cases in which the sentiment classification of the reviews using our Spanish negation detector system is not correct. Analyzing the reviews that were incorrectly classified, we find the following types of errors:

- Case 1: Words correctly identified as scope by the Spanish negation detector that are present in SO-CAL dictionary, but are not sentiment words in the domain under study.

In Example (168), the word *official* is in the scope of negation and belongs to the positive dictionary of SO-CAL. 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 a incorrect way.

168. 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

- Case 2: 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). On the other hand, if the word is negative, it is annulled, i.e., to the scoring of the word it is added its opposite value. In Example (169), 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).

169. 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 (REPEATED) = 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

- Case 3: Sentiment words not present in the SO-CAL dictionary. In Example (170) 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.

170. 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

- Case 4: Negation used in an ironic way. In Example (171), 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 our 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.

171. 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

- Case 5: Scope erroneously predicted by our Spanish negation detector. In Example (172), the negation processing system has predicted as scope of the last negation cue, *no*, the following: *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.

172. 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) = -6.075000000000001

       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) = -6.075000000000001

       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 (CAPITALIZED) = 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

- Case 6: Negation cue detected by the SO-CAL system, but not predicted by our Spanish negation detector. In Example (173), 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.

173. 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

In most of the cases where the Spanish SO-CAL did not see any improvements through negation detection, we can attribute that to problems with the system's search heuristics or with its dictionaries. 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 English version). It is clear, then, that better performance can be achieved by developing the system in conjunction with adopting a state-of-the-art negation processing system.

In summary, we have shown that accurate negation detection is possible and that improvements in sentiment analysis can be gained from detecting negation and its scope with sophisticated negation processing systems.

6.5 Conclusion

In this chapter, we have integrated the Spanish negation processing system presented in Chapter 5: "*A system to process negation in Spanish*" into a well-known sentiment polarity classifier, the SO-CAL system, for the polarity classification of the reviews of the SFU Review$_{SP}$-NEG corpus. We have compared the results obtained with those produced by using the search heuristics implemented in the SO-CAL system and a baseline model which involves not applying any negation identification (SO-CAL without negation), in order to show the importance of developing accurate negation processing systems for Natural Language Processing tasks.

The results obtained show that accurate recognition of negation cues and scopes is of paramount importance to the sentiment classification task and reveal that simplistic approaches to negation are insufficient for sentiment detection. Error analysis shows that our 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 English version). We plan to check the SO-CAL Spanish dictionaries. There are some words that are clearly sentiment word, such as *encanta* ('love'), that are not included in these dictionaries. In addition, the negation weighting factor of the sentiment heuristic of SO-CAL should be reviewed. Authors introduced it because positive 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 sometimes results in a very high negative score.

Besides the sentiment analysis task, accurate negation detection is useful for other tasks, such as information retrieval, information extraction, or machine translation. The system presented in this doctoral thesis could also be tested on them.

Chapter 7

NEGES: Workshop on Negation in Spanish

Negation is a complex linguistic phenomenon that has been widely studied from a theoretical perspective (L. R. Horn, 1989; L. Horn, 2010), and less from an applied point of view. However, interest in the computational treatment of this phenomenon is of growing interest, because it is relevant for a wide range of Natural Language Processing applications such as sentiment analysis or information retrieval, where it is crucial to know when the meaning of a part of the text changes due to the presence of negation. In fact, in recent years, several challenges and shared tasks have focused on negation processing: the BioNLP'09 Shared Task 3 (Kim et al., 2009), the NeSp-NLP 2010 Workshop: Negation and Speculation in Natural Language Processing (Morante & Sporleder, 2010), the CoNLL-2010 shared task (Farkas et al., 2010), the i2b2 NLP Challenge (Uzuner et al., 2011), the *SEM 2012 Shared Task (Morante & Blanco, 2012a), the ShARe/CLEF eHealth Evaluation Lab 2014 Task 2 (Mowery et al., 2014), the ExProM Workshop: Extra-Propositional Aspects of Meaning in Computational Linguistics (Morante & Sporleder, 2012b; Blanco et al., 2015, 2016) and the SemBEaR Workshop: Computational Semantics Beyond Events and Roles (Blanco et al., 2017; Blanco & Morante, 2018).

However, most of the research on negation has been done for English. Therefore, in order to advance the study of this phenomenon in Spanish, the second most widely spoken language in the world and the third most widely used on the Internet, we create NEGES. NEGES is the acronym for "NEGación en ESpañol" (Negation in Spanish).

In this chapter, we first present the origin of this workshop, its objective and the editions that have been held. Next, we describe the tasks that have been organized within it along with the datasets provided. Then, we summarize the participants of each edition and task and the results obtained. Finally, we report conclusions.

7.1 Origin, objective and editions held

In the 32th International Conference of the Spanish Society for Natural Language Processing (SEPLN 2016[1]) held in Salamanca, we observe special interest in the issue of negation and we decide to create a group on this topic, the NEGES group. At the end of 2016 we create a distribution list with researchers interested in studying this phenomenon. This list is currently made up of 26 researchers (June, 2019) from the Computational Linguistics and Natural Language Processing fields who aim to contribute to the ongoing research on negation in Spanish in the Language Technology community. To subscribe to this list, l-neg-sp, simply go to https://listas.ujaen.es/mailman/listinfo/l-neg-sp and fill in the details of the form.

NEGES group promotes research in negation detection in Spanish, provides members with a means of exchanging news of recent research developments and other matters of interest as well as it makes available resources relevant to negation detection in Spanish, including corpora, annotation guidelines, evaluation scripts, etc.

Activities of the NEGES group include the holding of an annual meeting each September at the International Conference of the Spanish Society for Natural Language Processing (SEPLN). Up to now, three editions of NEGES have been held: the first two as a workshop (NEGES: Workshop on Negation in Spanish) and the third as a task in a evaluation forum (NEGES task: Negation in Spanish). We present them below.

[1] http://cedi2016.scie.es/es/sepln

The first edition takes place in 2017 in the context of the 33th International Conference of the Spanish Society for Natural Language Processing (SEPLN 2017[2]) held in Murcia (Spain). The objectives of this edition are to bring together the scientific community interested in negation, to determine how the study of this phenomenon is being addressed, to identify the main problems encountered and to share resources and tools. In order to do so, the teams have to make an oral presentation related to some of these topics. A total of 7 teams participate and in view of the interest shown, in the next edition we decide to propose tasks to solve the main problems encountered throughout our research on this phenomenon: non-existence of a standard guide for the annotation of negation in Spanish, non-existence of a Spanish negation processing system and relevance of evaluating the role of negation in Natural Language Processing tasks.

The 2018 edition is held in Seville (Spain) as part of the 34th International Conference of the Spanish Society for Natural Language Processing (SEPLN 2018[3]). It consists of three tasks related to different aspects of negation (Jiménez-Zafra et al., 2019a): Task 1 on reaching an agreement on the guidelines to follow for the annotation of negation in Spanish, Task 2 on identifying negation cues, and Task 3 on evaluating the role of negation in sentiment analysis. A total of 4 teams participate in the workshop, 2 for developing annotation guidelines (Jiménez-Zafra et al., 2018a) and 2 for negation cues detection (Jiménez-Zafra et al., 2018b). Task 3 has no participants.

In the 2019 edition, NEGES is organized as a task of IberLEF[4] (Iberian Languages Evaluation Forum), with the aim of joining forces with other researchers to create a reference forum in Spanish with tasks of relevance to processing some of the languages spoken in the Iberian Peninsula. It is held in Bilbao (Spain) as part of the 35th International Conference of the Spanish Society for Natural Language Processing (SEPLN 2019[5]). In this edition (Jiménez-Zafra et al., 2019b), two sub-tasks are proposed as a continuation of the tasks carried out in NEGES 2018: Sub-task A: “Negation cues detection” (equivalent to Task 2 in NEGES 2018) and Sub-task B: “Role of negation in sentiment analysis” (equivalent to Task 3 in NEGES

[2] http://sepln2017.um.es/
[3] http://www.sepln2018.com/
[4] https://sites.google.com/view/iberlef-2019/
[5] http://hitz.eus/sepln2019/?language=es

2018). About 13 teams show interest in the task and 5 teams finally submit results, 4 for negation cues detection and 1 for evaluating the role of negation in sentiment analysis.

7.2 Tasks description

In the 2017 edition of NEGES, Workshop on Negation in Spanish, no task is organized. The aim is to bring together the scientific community interested in negation to determine future directions.

In the 2018 edition of NEGES, Workshop on Negation in Spanish, three tasks are proposed:

- **Task 1**: "Annotation guidelines"
- **Task 2**: "Negation cues detection"
- **Task 3**: "Role of negation in sentiment analysis"

In the 2019 edition of NEGES task, Negation in Spanish, two sub-tasks are proposed as a continuation of the tasks carried out in NEGES 2018:

- **Sub-task A**: "Negation cues detection"
- **Sub-task B**: "Role of negation in sentiment analysis"

As can be seen, Sub-task A of NEGES 2019 corresponds to Task 2 of NEGES 2018 and Sub-task B to Task 3. Therefore, in the editions of NEGES have been organized three different tasks, which are described below.

7.2.1 Annotation guidelines

This task is only proposed in NEGES 2018 and corresponds to Task 1. It has as goal to reach an agreement on the guidelines to follow for the annotation of negation in Spanish texts.

Although there have already been several annotation efforts, the community lacks a standard for the annotation of negation, contrary to what happens with other phenomena, such as semantic roles.

The corpora annotated so far in Spanish belong to 3 domains (news, clinical reports and product reviews) and are based on different guidelines. In this task, the guidelines used for the annotation of the corpora are made available to the participants so that they can analyze them (Table A.4). A period of analysis is provided and once it is over, participants send a document indicating which aspects of the guidelines they agree with and which they do not, all duly justified. The documents describing the perspective of each team are sent to the rest of participants prior to the workshop in order to enhance a discussion about the main aspects of interest and try to reach a consensus.

Table 7.1: Annotation guidelines provided for NEGES 2018 - Task 1: "Annotation guidelines".

Corpus	Domain	Annotation guidelines
UAM Spanish TreeBank	News	pp. 51-55 (Sandoval & Salazar, 2013)
IxaMed-GS	Clinical reports	pp. 322 (Oronoz et al., 2015)
SFU ReviewSP-NEG	Product reviews	pp. 538-559 (Jiménez-Zafra et al., 2018)
UHU-HUVR	Clinical reports	pp. 54-57 (Cruz Díaz et al., 2017)
IULA Spanish Clinical Record	Clinical reports	pp. 45-49 (Marimon et al., 2017)

7.2.2 Negation cues detection

This task is proposed in NEGES 2018 as Task 2 and corresponds to Sub-task A of NEGES 2019. Its aim is to promote the development and evaluation of systems for identifying negation cues in Spanish. Negation cues can be *simple*, if they are expressed by a single token (e.g., "no" *[no/not]*, "sin" *[without]*), *continuous*, if they are composed of a sequence of two or more contiguous tokens (e.g., "ni siquiera" *[not even]*, "sin ningún" *[without any]*), or *discontinuous*, if they consist of a sequence of two or more non-contiguous tokens (e.g., "no...apenas" *[not...hardly]*, "no...nada" *[not...nothing]*). For example, in sentence (174) the systems have to identify four negation cues: i) the discontinuous cue "No...nada" *[Not...nothing]*, ii) the simple cue "no" *[no/not]*, iii) the simple cue "no" *[no/not]* again, and iv) the continuous cue "ni siquiera" *[not even]*.

174. **No**[1] tengo **nada**[1] en contra del servicio del hotel, pero **no**[2] pienso volver, **no**[3] me ha gustado, **ni siquiera**[4] las vistas son buenas.

 I have nothing against the service of the hotel, but I do not plan to return, I did not like it, not even the views are good.

Participants receive a set of training and development data consisting of reviews of movies, books and products from the SFU Review $_{SP}$-NEG corpus(Jiménez-Zafra et al., 2018) to build their systems during the development phase. At a later stage, a set of tests are made available for evaluation. Finally, the participant's submissions are evaluated against the gold standard annotations.

7.2.3 Role of negation in sentiment analysis

This task is proposed in NEGES 2018 as Task 3 and corresponds to Sub-task B of NEGES 2019. It aims to evaluate the impact of accurate negation detection in sentiment analysis. In this task, participants have to develop a system that uses the negation information contained in the SFU Review$_{SP}$-NEG corpus (Jiménez-Zafra et al., 2018) to improve the task of polarity classification. They have to classify each review as *positive* or *negative* using an heuristic that incorporates negation processing. For example, systems should classify a review such as the one of Example (175) as *negative* using the negation information provided by the organization, a sample of which is shown in Figure 7.1.

175. El 307 es muy bonito, pero no os lo recomiendo. Por un fallo eléctrico te puedes matar en la carretera.

 The 307 is very nice, but I don't recommend it. An electrical failure can kill you on the road.

```xml
<sentence complex="no">
    <d wd="El" postype="article" pos="da0ms0" name="d" lem="el" num="s" gen="m"/>
    <z wd="307" pos="z" name="z" lem="307"/>
    <v wd="es" postype="semiauxiliary" pos="vsip3s0" name="v" lem="ser" person="3" num="s" tense="present" mood="indicative"/>
    <r wd="muy" pos="rg" name="r" lem="muy"/>
    <a wd="bonito" postype="qualificative" pos="aq0ms0" name="a" lem="bonito" num="s" gen="m"/>
    <f wd="," pos="fc" name="f" lem="," punct="comma"/>
    <c wd="pero" postype="coordinating" pos="cc" name="c" lem="pero"/>
    <neg_structure polarity="negative" value="neg" change="yes">
        <scope>
            <negexp>
                <r wd="no" postype="negative" pos="rn" name="r" lem="no"/>
            </negexp>
            <p wd="os" postype="personal" pos="pp2cp000" name="p" lem="os" person="2" num="p" gen="c"/>
            <p wd="lo" postype="personal" pos="pp3cna00" name="p" lem="lo" person="3" num="n" gen="c" case="accusative"/>
            <event>
                <v wd="recomiendo" postype="main" pos="vmip1s0" name="v" lem="recomendar" person="1" num="s" tense="present"
                mood="indicative"/>
            </event>
        </scope>
    </neg_structure>
    <f wd="." pos="fp" name="f" lem="." punct="period"/>
</sentence>
<sentence>
    <s wd="Por" postype="preposition" pos="sps00" name="s" lem="por" complex="no"/>
    <d wd="un" postype="indefinite" pos="di0ms0" name="d" lem="uno" num="s" gen="m"/>
    <n wd="fallo" postype="common" pos="ncms000" name="n" lem="fallo" num="s" gen="m"/>
    <a wd="eléctrico" postype="qualificative" pos="aq0ms0" name="a" lem="eléctrico" num="s" gen="m"/>
    <p wd="te" postype="personal" pos="pp2cs000" name="p" lem="te" person="2" num="s" gen="c"/>
    <v wd="puedes" postype="main" pos="vmip2s0" name="v" lem="poder" person="2" num="s" tense="present" mood="indicative"/>
    <v wd="matar" postype="main" pos="vmn0000" name="v" lem="matar" mood="infinitive"/>
    <s wd="en" postype="preposition" pos="sps00" name="s" lem="en" complex="no"/>
    <d wd="la" postype="article" pos="da0fs0" name="d" lem="el" num="s" gen="f"/>
    <n wd="carretera" postype="common" pos="ncfs000" name="n" lem="carretera" num="s" gen="f"/>
    <f wd="." pos="fp" name="f" lem="." punct="period"/>
</sentence>
```

Figure 7.1: Review annotated with negation information.

7.3 Data

The SFU Review$_{SP}$-NEG corpus[6] (Jiménez-Zafra et al., 2018) is the collection of documents provided for "Negation cues detection" and "Role of negation in sentiment analysis" tasks[7]. It is the corpus developed in this doctoral thesis (see Chapter 4: "*SFU Review$_{SP}$-NEG: a Spanish corpus annotated with negation*").

7.3.1 Negation cues detection

For this task the SFU Review$_{SP}$-NEG corpus is randomly splitted into development, training and test sets with 33 reviews per domain in training, 7 reviews per domain in development and 10 reviews per domain in test. The data is converted to CoNLL format (Buchholz & Marsi, 2006) where each line corresponds to a token, each annotation is provided in a column and empty lines indicate the end of the sentence. The content of the given columns is:

[6]http://sinai.ujaen.es/sfu-review-sp-neg-2/

[7]To download the data in the format provided for both tasks go to http://www.sepln.org/workshops/neges2019/ or send an email to the organizers.

- Column 1: domain_filename
- Column 2: sentence number within domain_filename
- Column 3: token number within sentence
- Column 4: word
- Column 5: lemma
- Column 6: part-of-speech
- Column 7: part-of-speech type
- Columns 8 to last: if the sentence has no negations, column 8 has a "***" value and there are no more columns. Else, if the sentence has negations, the annotation for each negation is provided in three columns. The first column contains the word that belongs to the negation cue. The second and third columns contain "-".

Figure 7.2 and Figure 7.3, show examples of the format of the files with different types of sentences. In the first example (Figure 7.2) there is no negation so the 8th column is "***" for all tokens, whereas the second example (Figure 7.3) is a sentence with two negation cues in which information for the first negation is provided in columns 8-10, and for the second in columns 11-13.

hoteles_yes_4_1	9	1	Muy	muy	rg	-	***
hoteles_yes_4_1	9	2	cómodo	cómodo	aq0ms0	qualificative	***
hoteles_yes_4_1	9	3	llegar	llegar	vmn0000	main	***
hoteles_yes_4_1	9	4	en	en	sps00	preposition	***
hoteles_yes_4_1	9	5	metro	metro	ncms000	common	***
hoteles_yes_4_1	9	6	(	(	fpa	-	***
hoteles_yes_4_1	9	7	parada	parada	ncfs000	common	***
hoteles_yes_4_1	9	8	Sevilla	sevilla	np00000	proper	***
hoteles_yes_4_1	9	9	)	)	fpt	-	***
hoteles_yes_4_1	9	10	o	o	cc	coordinating	***
hoteles_yes_4_1	9	11	andando	andar	vmg0000	main	***
hoteles_yes_4_1	9	12	desde	desde	sps00	preposition	***
hoteles_yes_4_1	9	13	el	el	da0ms0	article	***
hoteles_yes_4_1	9	14	centro	centro	ncms000	common	***
hoteles_yes_4_1	9	15	.	.	fp	-	***

Figure 7.2: Sentence without negation in CoNLL format.

The distribution of reviews and negation cues in the datasets is provided in Table 7.2: 264 reviews with 2,511 negation cues for training the systems, 56 reviews with 594 negation cues for the tuning process, and 80 reviews with 836 negation cues for the final evaluation.

```
hoteles_yes_5_2 12  1   El          el          da0ms0      article        -  -  -  -  -  -
hoteles_yes_5_2 12  2   hotel       hotel       ncms000     common         -  -  -  -  -  -
hoteles_yes_5_2 12  3   es          ser         vsip3s0     semiauxiliary  -  -  -  -  -  -
hoteles_yes_5_2 12  4   muy         muy         rg          -              -  -  -  -  -  -
hoteles_yes_5_2 12  5   bueno       bueno       aq0ms0      qualificative  -  -  -  -  -  -
hoteles_yes_5_2 12  6   ,           ,           fc          -              -  -  -  -  -  -
hoteles_yes_5_2 12  7   tiene       tener       vmip3s0     main           -  -  -  -  -  -
hoteles_yes_5_2 12  8   todas       todo        di0fp0      indefinite     -  -  -  -  -  -
hoteles_yes_5_2 12  9   las         el          da0fp0      article        -  -  -  -  -  -
hoteles_yes_5_2 12  10  comodidades comodidad   ncfp000     common         -  -  -  -  -  -
hoteles_yes_5_2 12  11  típicas     típico      aq0fp0      qualificative  -  -  -  -  -  -
hoteles_yes_5_2 12  12  de          de          sps00       preposition    -  -  -  -  -  -
hoteles_yes_5_2 12  13  un          uno         di0ms0      indefinite     -  -  -  -  -  -
hoteles_yes_5_2 12  14  cuatro      4           z           -              -  -  -  -  -  -
hoteles_yes_5_2 12  15  estrellas   estrella    ncfp000     common         -  -  -  -  -  -
hoteles_yes_5_2 12  16  ,           ,           fc          -              -  -  -  -  -  -
hoteles_yes_5_2 12  17  pero        pero        cc          coordinating   -  -  -  -  -  -
hoteles_yes_5_2 12  18  no          no          rn          negative       no -  -  -  -  -
hoteles_yes_5_2 12  19  es          ser         vsip3s0     semiauxiliary  -  -  -  -  -  -
hoteles_yes_5_2 12  20  nuevo       nuevo       aq0ms0      qualificative  -  -  -  -  -  -
hoteles_yes_5_2 12  21  ,           ,           fc          -              -  -  -  -  -  -
hoteles_yes_5_2 12  22  asi_que     así_que     cs          subordinating  -  -  -  -  -  -
hoteles_yes_5_2 12  23  no          no          rn          negative       -  -  -  no -  -
hoteles_yes_5_2 12  24  está        estar       vmip3s0     main           -  -  -  -  -  -
hoteles_yes_5_2 12  25  todo        todo        pi3ms000    indefinite     -  -  -  -  -  -
hoteles_yes_5_2 12  26  reluciente  reluciente  aq0cs0      qualificative  -  -  -  -  -  -
hoteles_yes_5_2 12  27  .           .           fp          -              -  -  -  -  -  -
```

Figure 7.3: Sentence with two negations in CoNLL format.

Table 7.2: Distribution of reviews and negation cues in the datasets of "Negation cues detection" task.

	Reviews	Negation cues
Training	264	2,511
Development	56	594
Test	80	836
Total	400	3,941

7.3.2 Role of negation in sentiment analysis

For this task we provide the SFU Review$_{SP}$-NEG corpus with the original format (XML). The meaning of the labels found in the reviews are the following:

- <review polarity="positive/negative">. It describes the polarity of the review, which can be *"positive"* or *"negative"*.
- <sentence complex="yes/no">. This label corresponds to a complete phrase or fragment thereof in which a negation structure can appear. It has associated the *complex* attribute that can take one of the following values:
  - "yes", if the sentence contains more than one negation structure.
  - "no", if the sentence only has a negation structure.
- <neg_structure>. This label corresponds to a syntactic structure in which a negation cue

appears. It has 4 possible attributes, two of which (*change* and *polarity_modifier*) are mutually exclusive.

- polarity: it presents the semantic orientation of the negation structure (*"positive"*, *"negative"* or *"neutral"*).
- change: it indicates whether the polarity or meaning of the negation structure has been completely changed because of the negation (*change="yes"*) or not (*change="no"*).
- polarity modifier: it states whether the negation structure contains an element that nuances its polarity. It can take the value *"increment"* if there is an increment in the intensity of the polarity or, on the contrary, it can take the value *"reduction"* if there is a reduction of it.
- value: it reflects the type of the negation structure, that is, *"neg"* if it expresses negation, *"contrast"* if it indicates contrast or opposition between terms, *"comp"* if it expresses a comparison or inequality between terms or *"noneg"* if it does not negate despite containing a negation cue.

• <scope>. This label delimits the part of the negation structure that is within the scope of negation. It includes both, the negation cue (<*negexp*>) and the event (<*event*>).

• <negexp>. It contains the word(s) that constitute(s) the negation cue. It can have associated the attribute *discid* if negation is represented by discontinuous words.

• <event>. It contains the words that are directly affected by the negation (usually verbs, nouns or adjectives).

The distribution of reviews in the training, development and test sets is provided in Table 7.3, as well as the distribution of the different negation structures per dataset. The total of positive and negative reviews can be seen in the rows named as *+ Reviews* and *- Reviews*, respectively.

Table 7.3: Distribution of reviews and negation cues in the datasets of "Role of negation in sentiment analysis" task.

	Training	Dev.	Test	Total
Reviews	264	56	80	400
+ Reviews	134	22	44	200
- Reviews	130	34	36	200
neg	2.511	594	836	3,941
noneg	104	22	55	181
contrast	100	23	52	175
comp	18	6	6	30

7.4 Evaluation measures

The evaluation script used to evaluate the systems presented in "Negation cues detection" task is the same as the one used to evaluate the *SEM 2012 Shared Task: "Resolving the Scope and Focus of Negation" (Morante & Blanco, 2012a). It is based on the following criteria:

- Punctuation tokens are ignored.
- A True Positive (TP) requires all tokens of the negation element have to be correctly identified.
- To evaluate cues, partial matches are not counted as False Positive (FP), only as False Negative (FN). This is to avoid penalizing partial matches more than missed matches.

The measures used to evaluate the systems are Precision (P), Recall (R) and F-score (F1). In the proposed evaluation, FN are counted either by the system not identifying negation cues 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. FP are counted when the system produces a negation cue not present in the gold annotations and TP are counted when the system produces negation cues exactly as they are in the gold annotations.

For evaluating the "Role of negation in sentiment analysis" task, the traditional measures used in text classification are applied: P, R, F1 and Accuracy (Acc). P, R and F1-score are measured per class and averaged using macro-average method.

$$P = \frac{TP}{TP + FP} \tag{7.1}$$

$$R = \frac{TP}{TP + FN} \tag{7.2}$$

$$F1 = \frac{2PR}{P + R} \tag{7.3}$$

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{7.4}$$

7.5 Participants and results

This section presents participants and results for NEGES 2018 and NEGES 2019.

As it has been previously mentioned, in NEGES 2017 no task is organized because the objective is to bring together the scientific community to see the interest aroused by the phenomenon of negation and to decide how to propose future editions. For this, participants make an oral presentation related to how negation is being addressed, what are the main problems encountered and resources and tools available.

7.5.1 NEGES 2018

In the 2018 edition of NEGES 10 teams show interest and 4 teams submit results: 2 for “Annotation guidelines” task and 2 for “Negation cues detection” task. Task 3: “Role of negation in sentiment analysis” has no participants.

7.5.1.1 Annotation guidelines

Task 1: "Annotation guidelines" has two participants: the CLiC team from the University of Barcelona, and Lucia Donatelli from the Georgetown University.

Martí & Taulé (2018) carry out an analysis of 5 fundamental aspects of the corpora analyzed: i) the negation cue, ii) the scope and the inclusion of the subject in the scope, iii) the coordinated structures, iv) the negative locutions and v) the lexical and morphological negation. Taking into account the differences observed in the annotation of the corpora, they proposed the following guidelines:

- Annotate the negation cue whenever possible, as it will allow to use it whenever necessary or to ignore it otherwise. Moreover, they consider that it should be distinguished between simple cues (e.g. "no" *[no/not]*, "sin" *[without]*) and complex cues (e.g. "no...nadie" *[no...nobody]*), where one implies the presence of the other. They propose to make use of the typology defined for us in the annotation of the SFU Review$_{\mathrm{SP}}$-NEG corpus.
- Annotate the scope including the subject within it. They mention that in many cases the focus of negation corresponds to the subject and this would facilitate future annotations of the corpus.
- Perform coordinated negation treatment. They propose to distinguish between coordinated structures affected by the same predicate and negation cue (Example 176) and coordinated structures with independent negation cues and predicates (Example 177), so that in the first case a single negation cue is considered and the rest of the negation structure as scope and, in the second case, a separate scope is annotated for each coordinated negation cue.
- Annotate negative locutions (e.g. "en absoluto" *[not at all]*), even if they do not contain explicit negation cues.
- Annotate lexical and morphological negation, which have only been addressed restrictively in the UHU-HUVR and IULA Spanish Clinical Record corpora.

- Annotate the focus of negation, which is not deal with in any of the guidelines analyzed.

176. [**No** es ni muy pesado ni muy ligero]. (SFU Review$_{SP}$-NEG)

    It is neither too heavy nor too light.

177. [**No** soy muy alta] [**tampoco** un pitufo]. (SFU Review$_{SP}$-NEG)

    I am not tall, but I am not a smurf either.

Donatelli (2018) describes each corpus individually and indicates which elements are missing in the annotation of each of them and those aspects that should have been taken into account. She considers that some components of the different guidelines can be combined in order to set linguistically precise guidelines and neutral guidelines with regard to the domain. She indicates that in order to represent the semantic of negation, the following elements must be annotated:

- The negation cue: lexical item that expresses negation.
- The scope: part of the text that is negated.
- The focus: part of the scope that is prominently or explicitly negated.
- The reinforcement (if exists): auxiliary negation or element of negative polarity, known as NPI (Negative Polarity Item) (Altuna et al., 2017).

Below we can see, in an example provided by the author (Example 178), the different elements explained above. The negation cue appears in bold, the scope in brackets, the focus in italics, and the reinforcement underlined.

178. John **no** [come *carne* <u>sino</u> verduras].

    John does not eat meat, he eats vegetables.

Donatelli considers that the scheme proposed by us for the annotation of the SFU ReviewSP-NEG corpus (Jiménez-Zafra et al., 2018) is suitable for capturing the layers of negation complexity and proposes to combine it with the use of the label *NegPolItem* used by Marimon et al. (2017) in the annotation of the IULA Spanish Clinical Record corpus to annotate items of negative polarity (NPI) or auxiliary negations.

7.5.1.2 Negation cues detection

Task 2: "Negation cues detection" has also two participants: the UPC team from the Universitat Politècnica de Catalunya (Loharja et al., 2018), and the UNED team from the National Distance Education University of Spain (UNED) (Fabregat et al., 2018). The official results by domain are shown in Table 7.4, and overall results are presented in Table 7.5, both evaluated in terms of P, R and F1.

Table 7.4: Official results by domain for NEGES 2018 - Task 2: "Negation cues detection".

	UNED			UPC		
Domain	P	R	F1	P	R	F1
Books	79.52	66.27	**72.29**	84.19	84.52	**84.35**
Cars	94.23	72.06	81.67	95.08	85.29	89.92
Cell phones	93.33	73.68	**82.35**	89.80	77.19	83.02
Computers	-	-	-	91.36	91.36	**91.36**
Hotels	97.67	71.19	**82.35**	94.00	79.66	86.24
Movies	86.26	69.33	76.87	89.68	85.28	87.42
Music	92.59	57.47	**70.92**	92.96	75.86	**83.54**
Washing machines	92.00	66.67	77.31	94.74	78.26	85.72

Table 7.5: Overall official results for NEGES 2018 - Task 2: "Negation cues detection".

Team	P	R	F1
UNED	79.45 (90.80)	59.58 (68.10)	67.97 (77.68)
UPC	91.48 (91.49)	82.18 (80.87)	86.45 (85.74)

The results by domain (Table 7.4) show that there are sub-collections such as books and music in which both systems obtain worse results compared to the rest of the sub-collections. The system developed by the UNED team obtains the highest performance in cell phones and hotels sub-collections, while the UPC system shows a better detection of negation cues in the computers sub-collection, in particular, it obtains an F1 of 91.36%.

Table 7.5 presents overall performances. The results in parentheses correspond to the performances without considering the computers subset, since the UNED team could not submit the results for computers due to technical problems. However, if we look at the other domains (Table 7.4), the UPC system clearly achieve better results.

In terms of the approaches applied, both proposals use the standard labelling scheme *BIO* where the first word of a negation structure denotes by *B* and the remaining words by *I*. The label *O* indicates that the word does not correspond with a negation cue.

The UNED team applies a model of deep learning inspired by named entity recognition architectures and negation detection models. Specifically, this system is focused on the use of several neural networks together with a bidirectional LSTM (Long Short-Term Memory). This supervised approach is based on pretrained word embeddings for Spanish. For its part, the UPC team uses Conditional Random Fields with a set of features such as the part-of-speech of the word and information about how the words are written.

Finally, the resources used by the participants are diverse. The UNED team uses Keras (Chollet et al., 2015) and TensorFlow (Abadi et al., 2016) libraries, as well as pretrained word embeddings for Spanish (Cardellino, 2016), and the UPC team uses NLTK (Loper & Bird, 2002).

7.5.1.3 Role of negation in sentiment analysis

Task 3: "Role of negation in sentiment analysis" has no participants. Some of the teams registered for the workshop show interest in the task, but expressed that they do not participate due to lack of time.

7.5.2 NEGES 2019

In the 2019 edition of NEGES 13 teams show interest and 5 teams submit results: 4 for Sub-task A: "Negation cues detection" and 1 for Sub-task B: "Role of negation in sentiment analysis".

7.5.2.1 Negation cues detection

Sub-task A: "Negation cues detection" has 4 participants: Aspie96 from the University of Turin (Giudice, 2019b), the CLiC team from Universitat de Barcelona (Beltrán & González, 2019), the IBI team from Integrative Biomedical Informatics group of Universitat Pompeu Fabra (Domínguez-Mas et al., 2019), and the UNED team from Universidad Nacional de Eduación a Distancia (UNED) and Instituto Mixto de Investigación-Escuela Nacional de Sanidad (IMIENS) (Fabregat et al., 2019). The official results by domain are shown in Table 7.6, and overall results are presented in Table 7.7, both evaluated in terms of P, R and F1. For IBI and UNED teams the domain in which it was most difficult to detect the negation cues was that of cell phones reviews, while for Aspie96 and CLiC it was the domain of hotels and books reviews, respectively. In terms of overall performance, the results of Aspie96 were quite low compared to the other teams. CLiC, IBI and UNED team obtained similar precision. However, the CLiC team achieved the highest recall, reaching the first rank position.

Table 7.6: Official results by domain for NEGES 2019 - Sub-task A: "Negation cues detection".

	Aspie96			CLiC			IBI			UNED		
Domain	P	R	F1	P	R	F1	P	R	F1	P	R	F1
Books	16.00	28.57	20.51	80.59	75.79	78.12	80.97	72.62	76.57	84.02	81.35	**82.66**
Cars	19.42	29.41	23.39	94.92	82.35	**88.19**	92.73	75.00	82.93	94.83	80.88	87.30
Cell phones	18.07	26.32	21.53	87.76	75.44	**81.13**	90.48	66.67	76.77	88.37	66.67	76.00
Computers	17.36	25.93	20.80	90.48	93.83	**92.12**	89.06	70.37	78.62	94.12	79.01	85.91
Hotels	10.59	15.25	12.50	87.50	71.19	78.51	97.67	71.19	82.35	93.62	74.58	**83.02**
Movies	20.53	33.13	25.35	88.67	81.60	84.99	90.30	74.23	81.48	89.86	81.60	**85.53**
Music	24.17	33.33	28.02	94.44	78.16	**85.53**	94.20	74.71	83.33	95.38	71.26	81.57
Washing machines	24.24	34.78	28.57	92.98	76.81	**84.13**	94.34	72.46	81.96	94.34	72.46	81.96

Aspie96 (Giudice, 2019b) presents a model based in a convolutional Recurrent Neural Network (RNN) previously used for irony detection in Italian tweets (Giudice, 2018) at IronITA shared task (Cignarella et al., 2018). In order to address the task at NEGES, the system is modified to

Table 7.7: Overall official results for NEGES 2019 - Sub-task A: “Negation cues detection”.

Team	P	R	F1
CLiC	89.67	79.40	84.09
UNED	91.82	75.98	82.99
IBI	91.22	72.16	80.50
Aspie96	18.80	28.34	22.58

take tokens and Spanish spelling into account. Each word is represented using a 50-character window in which non-word tokens are also considered. The words are then fed into a GRU layer to expand the context. The GRU layer’s output is fed to a classifier that classifies each word as not part of a negation cue, the first word of a negation cue or part of the latest started negation cue. A similar model is shown to be suitable for the classification of irony (Giudice, 2018) and factuality (Giudice, 2019a), but for negation it is not. The results of the task are quite low compared to other competing systems.

The CLiC team (Beltrán & González, 2019) develops a system based on the Conditional Random Field (CRF) algorithm, inspired in the system of Loharja et al. (2018) presented in NEGES 2018 (Jiménez-Zafra et al., 2019a), which achieves the best results. They use as features the word forms and PoS-tags of the actual word, the posterior word and the previous 6 words. They also conduct experiments including two post-processing methods: a set of rules and a vocabulary list composed of candidate cues extracted from an annotated corpus (NewsCom). Neither the rules nor the list of candidates boost basic CRF’s results during the development phase. Therefore, they present to the competition the CRF model without post-processing, achieving the first position in the rank.

The IBI team (Domínguez-Mas et al., 2019) experiments with four supervised learning approaches (CRF, Random Forest, Support Vector Machine with linear kernel and XGBoost) using shallow textual, lemma, PoS-tags and dependency tree features to characterize each token. For Random Forest, Support Vector Machine with linear kernel and XGBoost they also use the same set of features for the three previous and three posterior tokens in order to model the context of the token in focus. The highest performance during the development phase is the one grounded by the CRF approach. Therefore, they choose it to support their participation, reaching the third rank position in the competition.

The UNED team (Fabregat et al., 2019) participates in the sub-task with a system based on Deep Learning, which is an evolution of the system presented in the previous edition of this workshop (Fabregat et al., 2018). Specifically, they propose a BILSTM-based model using words, PoS-tags and characters embedding features, and a one-hot vector to represent casing information. Moreover, they include in the system a post-processing phase in which some rules are used to correct frequent errors made by the network. The results obtained represent an improvement in relation to those of the 2018 edition of NEGES and place them in the second position in the 2019 edition.

7.5.2.2 Role of negation in sentiment analysis

Sub-task B: "Role of negation in sentiment analysis" has 1 participant: LTG-Oslo from University of Oslo (Barnes, 2019). The official results per sentiment class (positive and negative) and overall results are shown in Table 7.8. The results for the positive class are better than those of the negative class and, overall, they do not give a strong performance in absolute numbers, but the proposed approach is very interesting. LTG-Oslo (Barnes, 2019) addresses the task using a multi-task learning approach where a single model is trained simultaneously to negation detection and sentiment analysis. Specifically, shared lower-layers in a deep Bidirectional Long Short-Term Memory network (BiLSTM) are used to predict negation, while the higher layers are dedicated to predicting sentiment at document-level.

Table 7.8: Official results for NEGES 2019 - Sub-task B: "Role of negation in sentiment analysis".

	LTG-Oslo			
	P	R	F1	Acc
Positive class	68.90	70.50	69.70	-
Negative class	62.90	61.10	62.00	-
Overall	65.90	65.80	65.85	66.20

7.6 Conclusion

In this chapter it has been presented NEGES, the first initiative promoting negation research in Spanish. Up to now, three editions of NEGES have been held: the first two in 2017 and 2018 as a workshop (NEGES: Workshop on Negation in Spanish) and the third in 2019 as a task (NEGES task: Negation in Spanish) in the evaluation forum IberLEF. All of the in the context of the International Conference of the Spanish Society for Natural Language Processing (SEPLN). In the course of the editions, three tasks have been proposed: i) "Annotation guidelines" to reach an agreement on the guidelines to follow for the annotation of negation in Spanish, ii) "Negation cues detection" to promote the identification of negation cues in Spanish, and iii) "Role of negation in sentiment analysis" in order to evaluate the role of negation in Spanish sentiment analysis. The SFU Review$_{\mathrm{SP}}$-NEG corpus is the collection of documents used to train and test the systems in the evaluation tasks.

In the 2017 edition of NEGES a total of 7 teams participate. No task is organized, they have to make an oral presentation related to how negation is being addressed, what are the main problems encountered and resources and tools available. The objective of this edition is to bring together the scientific community interested in negation to determine future direction of the workshop.

In the 2018 edition of NEGES, a total of 4 teams participate in the workshop, 2 for developing annotation guidelines and 2 for negation cues detection. The task of studying the role of negation in sentiment analysis has no participants. In the 2019 edition, 5 teams submit results, 4 for identifying negation cues and 1 for studying the role of negation in sentiment analysis. The low number of submissions in the "Role of negation in sentiment analysis" task in both editions may be due to the fact that in order to study the impact of accurate negation detection in sentiment analysis it is necessary to determine how to efficiently represent negation, in the case of machine learning systems, or how to modify the polarity of the words within the scope of negation in the case of lexicon-based systems.

Regarding the approaches followed to detect negation cues, the teams opt for traditional machine learning approaches and deep learning algorithms, confirming that the use of Conditional Random Fields obtains the best results in "Negation cues detection" task.

Concerning the system errors and difficulties encountered in the identification of negation cues, we can say the following. Aspie96 reports that the low results of its system could be due to the fact that only the text of the documents have been taken into account, without incorporating features such as the lemma and the PoS-tags of the words, which could be of help. In fact, the other teams use them and obtain good results. The CLiC team reports several types of errors: errors in identifying negation cues that do not express negation (e.g. "Ya estaba casi, no (B)?" *[It was almost there, wasn't it?]*); not correctly identifying continuous cues (e.g. "a no ser que" *[unless]*, "a excepción de" *[with the exception of]*, "a falta de" *[in the absence of]*); tagging elements such as "tan" *[so]*, "tanto" *[so much]*, "muy" *[very]* or "mucho" *[much]* in discontinuous cues; and not detecting discontinuous cues. The IBI team detects that the performance of the approaches tested drastically decreases when they deal with multi-token negation cues. The UNED and UPC teams also find it more difficult to identify multiple-term negation cues.

As for the difficulties and errors in the evaluation of the role of negation in sentiment analysis, LTG-Oslo states that given the fact that the task is performed at the document level, it is difficult to determine them exactly. However, it is concluded that the multi-task model (MTL) is better than the single-task sentiment model (STL) for this task and that the training size and different domains complicate the use of deep neural architectures.

Chapter 8

Conclusions

Nowadays, there is a vast amount of information on the Internet. The large number of sources and the high volume of texts make it difficult for users to select information of interest. Natural Language Processing has as a goal the automatic processing of natural language in order to facilitate access to information. In order to extract fine-grained information, systems need to be able to process a diversity of linguistic phenomena such as negation, irony or sarcasm that are used to give words a different meaning. This doctoral thesis deals with the phenomenon of negation.

Negation is a universal linguistic phenomenon with a great qualitative impact on Natural Language Processing applications. All languages possess different types of resources (morphological, lexical, syntactic), which allow to speak about properties that people or things do not hold or events that do not happen. The presence of a negation cue in a sentence modifying a proposition can describe a completely different situation.

Negation is a complex phenomenon that has been widely studied from a theoretical perspective (L. R. Horn, 1989; L. Horn, 2010; Morante & Sporleder, 2012a), and less from an applied point of view. However, interest in the computational treatment of this phenomenon is growing. Processing negation is relevant for a wide range of applications such as information retrieval (Liddy et al., 2000), information extraction (Savova et al., 2010), machine translation (Baker et al., 2012) or sentiment analysis (Liu, 2015). Information retrieval systems aim to provide relevant documents from a collection, given a user query. Negation has an important role because it is not the same to make a search (*"recetas con tomate"*) than to make the negated version of the search (*"recetas sin tomate"*). The information retrieval system must return completely different documents for each query. In other tasks, such as information extraction, negation analysis is also beneficial. Clinical texts often refer to negative findings, that is, conditions that are not present in the patient. Processing negation in these documents is crucial because the health of patients is at stake. For example, a diagnosis of a patient will be totally different if negation is not detected in the sentence *"No signs of DVT"*. Translating a negative sentence from one language into another is also challenging because negation is not used in the same way. For example, the Spanish sentence *"No tiene ninguna pretensión en la vida"* is equivalent to the English sentence *"He has no pretense in life"*, but in the first case two negation cues are used while in the second only one is used. Sentiment analysis is also another task in which the presence of negation has a great impact. A sentiment analysis system that does not process negation can extract a completely different opinion than the one expressed by the opinion holder. For example, the polarity of the sentence *"A fascinating film, I would repeat"* should be the opposite of its negation *"A film nothing fascinating, I would not repeat"*. Notwithstanding, negation does not always imply polarity reversal, it can also increment, reduce or have no effect on sentiment expressions, which makes the task even more difficult.

However, as we can see in some of the systems we use regularly, this phenomenon is not being processed effectively. For example, if we do the Google search in Spanish *"películas que no sean de aventuras"* (*non-adventure movies*), we obtain adventure movies, which reflects that the engine is not taking into account negation. Other examples can be found in on-line systems for sentiment analysis. If we analyze the Spanish sentence *"Jamás recomendaría comprar este producto."* (*I would never recommend buying this product.*) with Mr. Tuit system[1], we can see that the output returned by the system is positive but the text clearly expresses a negative opinion. In the meaning cloud system[2] we can find another example. If we write the Spanish sentence *"Este producto tiene fiabilidad cero."* (*This product has zero reliability.*), the system

[1] http://www.mrtuit.com/

[2] https://www.meaningcloud.com/es/productos/analisis-de-sentimiento

indicates that it is a positive text, while in fact it is negative.

Most of the research on negation has been done for English. Therefore, this doctoral thesis aims to advance the study of this phenomenon in Spanish, the second most widely spoken language in the world and the third most widely used on the Internet.

Below are the main contributions of this doctoral thesis, the conclusions of the same and the future work that can continue this research. Finally, the publications derived from this work and the research awards and recognitions obtained are presented.

8.1 Contributions

The work reported in this doctoral thesis contributes to advance in the processing of negation in Spanish and to show the importance of the computational treatment of this phenomenon for Natural Language Processing systems. The main contributions are:

1. Definition of negation and sentiment analysis from the computational point of view (Chapter 2).
2. Compilation of state-of-the-art negation processing systems in English and Spanish (Chapter 2).
3. Compilation of corpora annotated with negation in all languages. To the best of our knowledge there are corpora annotated for English, Spanish, Swedish, Dutch, Japanese, Chinese, German and Italian (Chapter 2).
4. Compilation of state-of-the-art systems applying negation for sentiment analysis in English and Spanish (Chapter 2).
5. Polarity classification system of Spanish tweets that incorporates a set of syntactic rules for determining the scope of negation. This rule-based approach has been proved to be better than the method most used to determine the scope of negation in English tweets (Chapter 3).

6. Typology of negation patterns for Spanish (Chapter 4).

7. Annotation scheme for the annotation of negation and sentiment (Chapter 4).

8. Problematic cases in the annotation of negation in Spanish (Chapter 4).

9. SFU Review$_{\text{SP}}$-NEG corpus, the first corpus annotated with negation in the review domain for Spanish. Each review is automatically annotated at the token level with PoS-tags and lemmas using Freeling (Padró, 1998)), 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. It is the first Spanish corpus that includes the event in the annotation of negation and that takes into account discontinuous negation cues. Moreover, it is the first corpus in which it is annotated how negation affects the words that are within its scope (Chapter 4).

10. Analysis of the corpora annotated with negation based on the following criteria: language, domain, availability, guidelines, size in number of sentences, annotated elements, elements with negation, and types and components of negation that have been annotated. This analysis discusses the possibility of merging corpora to create a larger data set to train a negation processing system. Moreover, it also shows overall negation processing tasks for which the corpora could be used, and specific tasks for which the corpora could be used to evaluate the impact of processing negation (Chapter 5).

11. A machine learning system to process negation in Spanish. The system outperforms state-of-the-art results for negation cue detection, whereas for scope identification it is the first system that performs the task for Spanish (Chapter 5).

12. Study of the effect of the negation processing system developed on sentiment analysis. The negation processing system presented in Chapter 5 is integrated into a well-know sentiment polarity classifier, the SO-CAL system (Taboada et al., 2011). The performance of our system is compared to the search heuristic implemented in SO-CAL for negation

detection, showing that accurate negation detection is of paramount importance to the sentiment classification task (Chapter 6).

13. NEGES group and NEGES workshop, the first initiative promoting negation research in Spanish (Chapter 7).

8.2 Conclusions and future work

Negation is a complex linguistic phenomenon and the issue of its computational treatment has not been resolved yet due to its complexity, the multiple linguistic forms in which it can appear and the different ways it can act on the words within its scope. If we want to develop systems that approach human understanding, it is necessary to incorporate the treatment of one of the main linguistic phenomena used by people in their daily communication. This doctoral thesis aims to advance the study of this phenomenon in Spanish and the conclusions obtained as a results of the same are exposed below.

As mentioned in the introduction of this chapter, negation has a great qualitative impact on Natural Language Processing applications such us sentiment analysis, information retrieval, information extraction or machine translation. When applying negation to a specific task, the first step should be to process negation with an accurate system and, second, to use the output to improve the task. Some of the Spanish sentiment analysis developed so far apply negation for a better classification of opinions, but they do not assess the processing of negation, probably due to the non-existence of a corpus annotated with negation and sentiment. In order to determine the strengths and weaknesses of sentiment analysis systems that incorporate a module for negation processing, it is necessary a corpus annotated at both levels, sentiment and negation. In this way, an error analysis could be carried out to check whether the system correctly determines the negation cues and their corresponding scopes or if some of the errors are caused by the polarity classifier used. For this reason, one of the contributions of this doctoral thesis is the SFU Review$_{\mathrm{SP}}$-NEG corpus, the first Spanish corpus annotated with negation and sentiment in the review domain.

For the annotation of the SFU Review$_{SP}$-NEG corpus we develop a typology of negation patterns for Spanish, taking into account the basic principles contained in the standard descriptive and normative grammars of Spanish language (Demonte & Bosque, 1999; Española, 2009). Most existing annotation schemes for Spanish do not account for the complexity of the linguistic structures used to express negation, maybe because corpora have been created for specific purposes such as extracting negated clinical events, and not with the intention of accounting for all the linguistic complexity of negation.

The SFU Review$_{SP}$-NEG corpus is annotated to contribute to the study of negation in Spanish. We use it to develop a Spanish negation processing. The results for negation cue detection outperform state-of-the-art, whereas for scope identification this is the first system that performs the task for Spanish. The error analysis of the experiments carried out shows that the most frequent negation cue, the simple cue “no”, is also one of the most difficult to classify due to its ambiguity. It is also a frequent cue in comparative and contrasting constructions. In addition, the identification of discontinuous and infrequent cues remains a challenge. Regarding scopes, it is difficult to determine when the subject and adverbial complements of the verb are included within the scope, as well as the elements of coordinated structure. Moreover, some inconsistent annotations are detected in the training set.

For the development of the machine learning negation processing system we review all existing corpora in order to see whether it is possible to merge them to create a larger training corpus. This analysis reveals that most of the corpora have been annotated in the last 5 years, which shows that negation is a phenomenon whose processing has not yet been resolved and which is generating interest. Concerning the domains, those that have mainly attracted the attention of researchers are the medical domain and reviews/opinion articles. The conclusion of the analysis is that corpora can not be merged in their actual form. There are differences in the annotation schemes used, and most importantly, in the annotation guidelines, the way in which each corpus is tokenized and the negation elements that have been annotated. The annotation formats are different for each corpus, there is no standard annotation scheme. Moreover, the criteria used during the annotation process are different, especially with regard to three aspects: the inclusion or not of the subject and the cue in the scope, the annotations of the scope as the largest or shortest syntactic unit, and the annotation of all the negation cues or a subset of them according to a predefined set. Another important finding is that, in most of the corpora it is not specified how they are tokenized, which is essential for negation processing systems because the identification of negated elements (cue, scope, event and focus) is carried out at token level.

At a later stage, the negation processing system is applied to improve sentiment analysis task. It is integrated into a well-known sentiment analysis system, SO-CAL (Taboada et al., 2011), which includes a rule-based heuristic for negation detection. The performance of our system is compared to that of a baseline model which does not take negation into account (SO-CAL without negation) and the SO-CAL system with built-in negation. The results confirm that systems considering negation outperform the baseline, being the configuration of SO-CAL with our negation processing system the one that achieves the best performance. In this study it is shown the importance of developing accurate negation processing systems for sentiment analysis. Moreover, error analysis reveals possible ways to improve the SO-CAL system. Spanish dictionaries and the negation weighting factor of the sentiment heuristic should be reviewed.

Finally, we have launched NEGES, the first initiative promoting negation research in Spanish. It is a group made up of researchers from the Computational Linguistics and Natural Language Processing fields which provides a means of exchanging news of recent research developments as well as to share resources relevant to negation detection in Spanish such as corpora, annotation guidelines, evaluation scripts, etc. In addition, each year we organized a workshop in the context of the International Conference of the Spanish Society for Natural Language Processing (SEPLN) in which evaluation taks are proposed using the SFU Review$_{SP}$-NEG corpus. When resources are developed, the most ideal is to give them visibility in order to contribute to the advancement of the phenomenon studied. The conclusions obtained after 3 editions are that although a considerable number of teams show interest in the tasks, finally only a few send results, probably due to the difficulty of studying this phenomenon. In the course of the editions, three tasks have been proposed: i) “Annotation guidelines” to reach an agreement on the guidelines to follow for the annotation of negation in Spanish, ii) “Negation cues detection” to promote the identification of negation cues in Spanish, and iii) “Role of negation in sentiment analysis" in order to evaluate the role of negation in Spanish sentiment analysis.

As for future work, we plan to integrate the negation processing system in other Natural Language Processing tasks. Besides the sentiment analysis task, accurate negation detection is useful for other tasks, such as information retrieval, information extraction, or machine translation. The system presented in this doctoral thesis could also be tested on them. Moreover, we will review the SFU Review$_{SP}$-NEG corpus to resolve some inconsistent annotations detected and we will experiment with other features and algorithms in order to improve the accuracy of our negation processing system. In addition, we will work on the standardization of negation as has been done for other well established tasks like semantic role labelling and parsing. A robust and precise annotation scheme should be defined for the different elements that represent the phenomenon of negation (cue, scope, negated event and focus) and researchers should work together to define common annotation guidelines. Finally, we will continue organizing tasks in NEGES in order to advance in the study of negation in Spanish.

8.3 Publications

This section presents the publications derived from the work of this doctoral thesis. They are divided into main publications and other publications. Main publications are a direct result of this doctoral thesis, and other publications are those in which we have applied techniques and tools related to Natural Language Processing that have also contributed to the development of the thesis.

8.3.1 Main publications

8.3.1.1 JCR-indexed journals

1. **Jiménez-Zafra, S. M.**, Cruz-Díaz, N. P., Taboada, M., & Martín-Valdivia, M. T. (2019). Negation Detection for Sentiment Analysis: A Case Study in English and Spanish. Journal of Natural Language Engineering, Special Issue on Processing Negation (Under re-

view).

Impact factor: 1.130, **Quartile:** Q2, **Number of citations (Google Scholar):** -

2. **Jiménez-Zafra, S. M.**, Morante, R., Martín-Valdivia, M. T., & Ureña-López, L. A. (2019). Corpora Annotated with Negation: An Overview. Computational Linguistics (Under review - Second round)

   Impact factor: 2.130, **Quartile:** Q1, **Number of citations (Google Scholar):** -

3. **Jiménez-Zafra, S. M.**, Martín-Valdivia, M. T., Molina-González, M. D., & Ureña-López, L. A. (2018). Relevance of the SFU Review$_{SP}$-NEG corpus annotated with the scope of negation for supervised polarity classification in Spanish. Information Processing & Management, 54(2), 240-251. DOI: 10.1016/j.ipm.2017.11.007

   Impact factor: 3.892, **Quartile:** Q1, **Number of citations (Google Scholar):** 2

4. **Jiménez-Zafra, S. M.**, Taulé, M., Martín-Valdivia, M. T., Ureña-López, L. A., & Martí, M. A. (2018). SFU Review SP-NEG: a Spanish corpus annotated with negation for sentiment analysis. a typology of negation patterns. Language Resources and Evaluation, 52(2), 533-569. DOI: 10.1007/s10579-017-9391-x

   Impact factor: 1.029, **Quartile:** Q4, **Number of citations (Google Scholar):** 11

5. **Jiménez-Zafra, S. M.**, Martín-Valdivia, M. T., Martínez-Cámara, E., & Ureña-López, L. A. (2017). Studying the scope of negation for Spanish sentiment analysis on Twitter. IEEE Transactions on Affective Computing, 10(1), 129-141.
   DOI: 10.1109/TAFFC.2017.2693968

   Impact factor: 4.585, **Quartile:** Q1, **Number of citations (Google Scholar):** 9

8.3.1.2 Peer-reviewed journals

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

SCImago Journal Rankings (SJR): 0.235, **Quartile:** Q2, **Number of citations (Google Scholar):** -

2. Martí, M. A., Martín-Valdivia, M. T., Taulé Delor, M., **Jiménez-Zafra, S. M.**, Nofre, M., & Marsó, L. (2016). La negación en español: análisis y tipología de patrones de negación. Procesamiento del Lenguaje Natural, (57), 41-48.
   SCImago Journal Rankings (SJR): 0.199, **Quartile:** Q2, **Number of citations (Google Scholar):** 14

3. **Jiménez-Zafra, S. M.**, Martínez-Cámara, E. M., Martín-Valdivia, M. T.,& Molina-González, M. D. (2015). Tratamiento de la Negación en el Análisis de Opiniones en Español. Procesamiento del Lenguaje Natural, (54), 37-44.
   SCImago Journal Rankings (SJR): 0.191, **Quartile:** Q2, **Number of citations (Google Scholar):** 11

8.3.1.3 International conferences

1. **Jiménez-Zafra, S. M.**, Morante, R., Martín-Valdivia, M. T., & Ureña-López, L. A. (2018, August). A review of Spanish corpora annotated with negation. In Proceedings of the 27th International Conference on Computational Linguistics (pp. 915-924).
   Number of citations (Google Scholar): 1

8.3.1.4 National conferences

1. **Jiménez-Zafra, S. M.** (2017, September). Detección de la negación en textos en español y aplicación al análisis de sentimientos. In Proceedings of Doctoral Symposium of the 33rd Conference of the Spanish Society for Natural Language Processing (SEPLN 2017) (pp. 1-6)
   Number of citations (Google Scholar): 1

2. **Jiménez-Zafra, S. M.** (2015, September). Análisis de Sentimientos a nivel de aspecto y estudio de la negación en opiniones escritas en español. Actas del XXXI Congreso de la

Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN 2015) (pp. 1-6)
Number of citations (Google Scholar): -

8.3.1.5 International workshops

1. **Jiménez-Zafra, S. M.**, Martin, M., Lopez, L. A. U., Marti, T., & Taulé, M. (2016, December). Problematic cases in the annotation of negation in Spanish. In Proceedings of the Workshop on Extra-Propositional Aspects of Meaning in Computational Linguistics (ExProM) (pp. 42-48).
Number of citations (Google Scholar): 5

8.3.1.6 National workshops

1. **Jiménez-Zafra, S. M.**, Díaz, N. P. C., Morante, R., & Martín-Valdivia, M. T. (2019). NEGES 2019 Task: Negation in Spanish. Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2019), CEUR Workshop Proceedings, Bilbao, Spain, CEUR-WS (2019, September), (To be published)
Number of citations (Google Scholar): -

2. **Jiménez-Zafra, S. M.**, Díaz, N. P. C., Morante, R., & Martín-Valdivia, M. T. (2018, September). Tarea 1 del Taller NEGES 2018: Guías de Anotación. In Proceedings of NEGES 2018: Workshop on Negation in Spanish (Vol. 2174, pp. 15-21).
Number of citations (Google Scholar): 2

3. **Jiménez-Zafra, S. M.**, Díaz, N. P. C., Morante, R., & Martín-Valdivia, M. T. (2018, September). Tarea 2 del Taller NEGES 2018: Detección de Claves de Negación. In Proceedings of NEGES 2018: Workshop on Negation in Spanish (Vol. 2174, pp. 35-41).
Number of citations (Google Scholar): 2

4. **Jiménez-Zafra, S. M.**, Martín-Valdivia, M. T., & Ureña-López, L. A. (2014, November). Nuevos retos en el Análisis de Sentimientos. I Congreso Internacional de Jóvenes Investigadores en Patrimonio. II Jornadas Doctorales de la Universidad de Jaén (pp.

1-3).

Number of citations (Google Scholar): -

5. **Jiménez Zafra, S. M.**, Martínez-Cámara, E. , Martín-Valdivia, M. T., & Ureña-López, L. A. (2014, September). SINAI-ESMA: An unsupervised approach for Sentiment Analysis in Twitter. In Proceedings of the TASS Workshop at SEPLN Conference (Girona, Spain, September 16-19, 2014).

   Number of citations (Google Scholar): 7

6. **Jiménez Zafra, S. M.**, Martínez-Cámara, E. , Martín-Valdivia, M. T., & Ureña-López, L. A. (2014, July).Desafíos del Análisis de Sentimientos. V Jornadas TIMM (pp. 15-18).

   Number of citations (Google Scholar): -

8.3.2 Other publications

8.3.2.1 JCR-indexed journals

1. **Jiménez-Zafra, S. M.**, Sáez-Castillo, A. J., Conde-Sánchez, A., & Martín-Valdivia, M. T. (2019). How Do Sentiments Affect Virality on Twitter? ACM Transactions on Internet Technology (Under review)

   Impact factor: 2.382, **Quartile:** Q2, **Number of citations (Google Scholar):** -

2. **Jiménez-Zafra, S. M.**, Martín-Valdivia, M. T., Molina-González, M. D., & Ureña-López, L. A. (2019). How do we talk about doctors and drugs? Sentiment analysis in forums expressing opinions for medical domain. Artificial Intelligence in Medicine, 93, 50-57. DOI: 10.1016/j.artmed.2018.03.007

   Impact factor: 3.574, **Quartile:** Q1, **Number of citations (Google Scholar):** 1

3. **Jiménez-Zafra, S. M.**, Martín-Valdivia, M. T., Martínez-Cámara, E., & Ureña-López, L. A. (2016). Combining resources to improve unsupervised sentiment analysis at aspect-level. Journal of Information Science, 42(2), 213-229. DOI: 10.1177/0165551515593686

   Impact factor: 1.372, **Quartile:** Q3, **Number of citations (Google Scholar):** 19

8.3.2.2 Peer-reviewed journals

1. **Jiménez-Zafra, S. M.**, Plaza-del-Arco, F. M., García-Cumbreras, M. A., Molina-González, M. D., Ureña-López, L. A. & Martín-Valdivia, M. T. (2018). Monge: Geographic Monitor of Diseases. Procesamiento del Lenguaje Natural, 61, 193-196.
   SCImago Journal Rankings (SJR): 0.235, **Quartile:** Q3, **Number of citations (Google Scholar):** -

2. Plaza-del-Arco, F. M., Molina-González, M. D., **Jiménez-Zafra, S. M.**, & Martín-Valdivia, M. T. (2018). Lexicon Adaptation for Spanish Emotion Mining. Procesamiento del Lenguaje Natural, 61, 117-124.
   SCImago Journal Rankings (SJR): 0.235, **Quartile:** Q3, **Number of citations (Google Scholar):** -

3. **Jiménez-Zafra, S. M.**, Martín-Valdivia, M. T., Maks, I., & Izquierdo, R. (2017). Analysis of patient satisfaction in Dutch and Spanish online reviews. Procesamiento del Lenguaje Natural, 58, 101-108.
   SCImago Journal Rankings (SJR): 0.210, **Quartile:** Q3, **Number of citations (Google Scholar):** 9

4. Plaza-del-Arco, F. M., Martín-Valdivia, M. T., **Jiménez-Zafra, S. M.**, Molina-González, M. D., & Martínez-Cámara, E. (2016). COPOS: corpus of patient opinions in spanish. application of sentiment analysis techniques. Procesamiento del Lenguaje Natural, (57), 83-90.
   SCImago Journal Rankings (SJR): 0.199, **Quartile:** Q3, **Number of citations (Google Scholar):** 20

5. Villena-Román, J., Martínez-Cámara, E. M., García-Morera, J., & **Jiménez-Zafra, S. M.** (2015). TASS 2014 – The Challenge of Aspect-based Sentiment Analysis. Procesamiento del Lenguaje Natural, 54, 61-68.
   SCImago Journal Rankings (SJR): 0.191, **Quartile:** Q3, **Number of citations (Google Scholar):** 23

6. Molina-González, M. D., Martínez-Cámara, E. M., Martín-Valdivia, M. T., & **Jiménez-Zafra, S. M.** (2015). eSOLHotel: Generación de un lexicón de opinión en español adaptado al dominio turístico. Procesamiento del Lenguaje Natural, 54, 21-28.
   SCImago Journal Rankings (SJR): 0.191, **Quartile:** Q3, **Number of citations (Google Scholar):** 5

8.3.2.3 International conferences

1. **Jiménez-Zafra, S. M.**, Berardi, G., Esuli, A., Marcheggiani, D., Martín-Valdivia, M. T., & Fernández, A. M. (2015, September). A multi-lingual annotated dataset for aspect-oriented opinion mining. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (pp. 2533-2538).
   Number of citations (Google Scholar): 4

8.3.2.4 International workshops

1. García-Cumbreras, M. A., **Jiménez-Zafra, S. M.**, Montejo-Ráez, A., Díaz-Galiano, M. C., & Saquete, E. (2019, June). SINAI-DL at SemEval-2019 Task 7: Data Augmentation and Temporal Expressions. In Proceedings of the 13th International Workshop on Semantic Evaluation (pp. 1120-1124).
   Number of citations (Google Scholar): -

2. Montejo-Ráez, A., **Jiménez-Zafra, S. M.**, García-Cumbreras, M. A., & Díaz-Galiano, M. C. (2019, June). SINAI-DL at SemEval-2019 Task 5: Recurrent networks and data augmentation by paraphrasing. In Proceedings of the 13th International Workshop on Semantic Evaluation (pp. 480-483).
   Number of citations (Google Scholar): -

3. Plaza-del-Arco, F. M, **Jiménez-Zafra, S. M.**, Martín-Valdivia, M. T., & Ureña-López, L. A (2018, June). SINAI at SemEval-2018 Task 1: Emotion Recognition in Tweets. In Proceedings of The 12th International Workshop on Semantic Evaluation (pp. 128-132).
   Number of citations (Google Scholar): -

4. Díaz-Galiano, M. C., Martín-Valdivia, M. T., **Jiménez-Zafra, S. M.**, & Ureña-López, L. A. (2017, September). Sinai at clef ehealth 2017 task 3. CLEF. CEUR Workshop Proceedings (pp. 1-5).

   Number of citations (Google Scholar): 1

5. **Jiménez-Zafra, S. M.**, Montejo-Ráez, A., Martín-Valdivia, M. T., & Ureña-López, L. A. (2017, August). SINAI at SemEval-2017 Task 4: User based classification. In Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017) (pp. 634-639).

   Number of citations (Google Scholar): 1

6. **Jiménez-Zafra, S. M.**, Martín-Valdivia, M. T., & Molina-González, M. D. (2017, June). Corpus annotation for aspect based sentiment analysis in medical domain. Knowledge Representation for Health Care Process-Oriented Information Systems in Health Care Extraction & Processing of Rich Semantics from Medical Texts (pp. 115-125).

   Number of citations (Google Scholar): 2

7. Gutiérrez, Y., **Jiménez-Zafra, S. M.**, Moreno, I., Martín-Valdivia, M. T., Tomás, D., Montejo-Ráez, A., Montoyo, A., Ureña-López, L. A., Martínez-Barco, P., Martínez-Santiago, F., Muñoz, R., & Blanco-López, E. (2016, November). REDES at TAC Knowledge Base Population 2016: EDL and BeSt tracks. In Proceedings of TAC Knowledge Base Population (pp. 1-11).

   Number of citations (Google Scholar): 1

8. Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., Mohammad, A. S., Al-Ayyoub, M., Zhao, Y., Qin, B., De Clercq, O., Hoste, V., Apidianaki, M., Tannier, X., Loukachevitch, N., Kotelnikov, E., Bel, N., & **Jiménez-Zafra, S. M.** (2016, June). SemEval-2016 Task 5: Aspect Based Sentiment Analysis. In Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016) (pp. 19-30).

   Number of citations (Google Scholar): 491

9. **Jiménez-Zafra, S. M.**, Martín-Valdivia, M. T., Molina-González, M. D., & Ureña-López, L. A. (2016, June). Domain adaptation of polarity lexicon combining term fre-

quency and bootstrapping. In Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (pp. 137-146).
Number of citations (Google Scholar): 6

10. **Jiménez-Zafra, S. M.**, Martínez-Cámara, E., Martín-Valdivia, M. T., & Ureña-López, L. A. (2015, June). SINAI: Syntactic Approach for Aspect-Based Sentiment Analysis. In Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015) (pp. 730-735).
    Number of citations (Google Scholar): 3

11. **Jiménez-Zafra, S. M.**, Martínez-Cámara, E., Martin, M., & Lopez, L. A. U. (2014, August). SINAI: Voting System for Aspect Based Sentiment Analysis. In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014) (pp. 566-571).
    Number of citations (Google Scholar): 1

12. Martínez-Cámara, E., **Jiménez-Zafra, S. M.**, Martin, M., & Lopez, L. A. U. (2014, August). Sinai: Voting system for twitter sentiment analysis. In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014) (pp. 572-577).
    Number of citations (Google Scholar): 7

8.3.2.5 National workshops

1. Plaza-del-Arco, F. M., **Jiménez-Zafra, S. M.**, Martín-Valdivia, M. T., & Urena-López, L. A. (2018, October). Using Facebook Reactions to Recognize Emotion in Political Domain. I Workshop en Ciencia de Datos en Redes Sociales (CIDRES), XVIII Conferencia de la Asociación Española para la Inteligencia Artificial, CAEPIA 2018 (pp. 955-960)
   Number of citations (Google Scholar): -

2. López-Úbeda, P., Díaz-Galiano, M. C., Martín-Valdivia, M. T., & **Jiménez-Zafra, S. M.** (2018, September). SINAI at DIANN-IberEval 2018. Annotating Disabilities in Multi-language Systems with UMLS. In IberEval@ SEPLN (pp. 37-43).
   Number of citations (Google Scholar): -

3. García-Vega, M., Montejo-Ráez, M. C., Díaz-Galiano, M. C., & **Jiménez-Zafra, S. M.** (2017, September). SINAI en TASS 2017: clasificación de la polaridad de tweets integrando información de usuario. In Proceedings of TASS 2017: Workshop on Sentiment Analysis at SEPLN (pp. 91-96)
Number of citations (Google Scholar): -

8.3.2.6 Books

1. Montejo-Ráez, A., & **Jiménez-Zafra, S.M.** (2019). Curso de programación Python. Anaya Multimedia
Number of citations (Google Scholar): -

8.4 Research awards and recognitions

The research awards and recognitions obtained in the course of this doctoral thesis are presented below:

1. **Description:** Research award to the best research work at "IV Jornadas Doctorales TIC".
Concessionaire entity: Centro de Estudios Avanzados en Tecnologías de la Información y la Comunicación (CEATIC), Universidad de Jaén.
Date: May 25, 2018

2. **Description:** Second prize at "II Hackathon de Tecnologías del Lenguaje" in the modality "Corpus general" within Four Years From Now (4YFN) of the Mobile World Congress (MWC).
Concessionaire entity: Red.es, in collaboration with Secretaría de Estado para la Sociedad de la Información y la Agenda Digital (SESIAD).

Date: February 26, 2018

3. **Description:** Research recognition at "Jornadas Doctorales para jóvenes investigadores de la Universidad de Jaén 2017".
   Concessionaire entity: Vicerrectorado de Enseñanzas de Grado, Posgrado y Formación Permanente, Universidad de Jaén.
   Date: November 24, 2017

4. **Description:** Research recognition at "III Jornadas Doctorales TIC".
   Concessionaire entity: Centro de Estudios Avanzados en Tecnologías de la Información y la Comunicación (CEATIC), Universidad de Jaén.
   Date: May 20, 2017

5. **Description:** Research award to the best work of research initiation at "II Premios Ada Lovelace de Tecnologías de la Información y la Comunicación".
   Concessionaire entity: Centro de Estudios Avanzados en Tecnologías de la Información y la Comunicación (CEATIC), Universidad de Jaén.
   Date: July 8, 2016

6. **Description:** Research award to the best research work at "II Jornadas Doctorales TIC".
   Concessionaire entity: Centro de Estudios Avanzados en Tecnologías de la Información y la Comunicación (CEATIC), Universidad de Jaén.
   Date: June 4, 2016

Appendix A

Comparative tables of corpora analysis

Table A.1: Language and year of publication of the corpora.

Corpus	Language	Year
BioInfer (Pyysalo et al., 2007)	English	2007
Genia Event (Kim et al., 2008)	English	2008
BioScope (Vincze et al., 2008)	English	2008
Product Review (Councill et al., 2010)	English	2010
Stockholm Electronic Patient Record (Dalianis & Velupillai, 2010)	Swedish	2010
PropBank Focus (PB-FOC) (Blanco & Moldovan, 2011a)	English	2011
ConanDoyle-neg (Morante & Daelemans, 2012)	English	2012
SFU Review$_{EN}$ (Konstantinova et al., 2012)	English	2012
NEG-DrugDDI (Bokharaeian et al., 2013)	English	2013
UAM Spanish Treebank (Sandoval & Salazar, 2013)	Spanish	2013
NegDDI-DrugBank (Bokharaeian et al., 2014)	English	2014
EMC Dutch (Afzal et al., 2014)	Dutch	2014
Review and Newspaper Japanese (Matsuyoshi et al., 2014)	Japanese	2014
IxaMed-GS (Oronoz et al., 2015)	Spanish	2015
Deep Tutor Negation (Banjade & Rus, 2016)	English	2016
CNeSp (Zou et al., 2016)	Chinese	2016
German negation and speculation (Cotik, Roller, et al., 2016)	German	2016
Fact-Ita Bank Negation (Altuna et al., 2017)	Italian	2016
SFU Review$_{SP}$-NEG (Jiménez-Zafra et al., 2018)	Spanish	2017
UHU-HUVR (Cruz Díaz et al., 2017)	Spanish	2017
IULA Spanish Clinical Record (Marimon et al., 2017)	Spanish	2017
SOCC (Kolhatkar et al., 2018)	English	2018

Table A.2: Availability of the corpora.

Corpus	Links to the data
BioInfer (Pyysalo et al., 2007)	http://mars.cs.utu.fi/BioInfer/
Genia Event (Kim et al., 2008)	http://www.geniaproject.org/genia-corpus/event-corpus
BioScope (Vincze et al., 2008)	http://rgai.inf.u-szeged.hu/index.php?lang=en&page=bioscope
Product Review (Councill et al., 2010)	-
Stockholm Electronic Patient Record (Dalianis & Velupillai, 2010)	-
PropBank Focus (PB-FOC) (Blanco & Moldovan, 2011a)	http://www.clips.ua.ac.be/sem2012-st-neg/data.html
ConanDoyle-neg (Morante & Daelemans, 2012)	http://www.clips.ua.ac.be/sem2012-st-neg/data.html
SFU Review$_{EN}$ (Konstantinova et al., 2012)	https://www.sfu.ca/~mtaboada/SFU_Review_Corpus.html
NEG-DrugDDI (Bokharaeian et al., 2013)	http://nil.fdi.ucm.es/sites/default/files/NegDrugDDI.zip
UAM Spanish Treebank (Sandoval & Salazar, 2013)	http://www.lllf.uam.es/ESP/Treebank.html
NegDDI-DrugBank (Bokharaeian et al., 2014)	http://nil.fdi.ucm.es/sites/default/files/NegDDI_DrugBank.zip
EMC Dutch (Afzal et al., 2014)	-
Review and Newspaper Japanese (Matsuyoshi et al., 2014)	http://cl.cs.yamanashi.ac.jp/nldata/negation/
IxaMed-GS (Oronoz et al., 2015)	-
Deep Tutor Negation (Banjade & Rus, 2016)	http://deeptutor.memphis.edu/resources.htm
CNeSp (Zou et al., 2016)	http://nlp.suda.edu.cn/corpus/CNeSp/
German negation and speculation (Cotik, Roller, et al., 2016)	-
Fact-Ita Bank Negation (Altuna et al., 2017)	https://hlt-nlp.fbk.eu/technologies/fact-ita-bank
SFU Review$_{SP}$-NEG (Jiménez-Zafra et al., 2018)	http://sinai.ujaen.es/sfu-review-sp-neg-2/
UHU-HUVR (Cruz Díaz et al., 2017)	-
IULA Spanish Clinical Record (Marimon et al., 2017)	http://eines.iula.upf.edu/brat//#/NegationOnCR_IULA/
SOCC (Kolhatkar et al., 2018)	https://researchdata.sfu.ca/islandora/object/islandora%3A9109

NOTE: Link to the Review and Japanese corpus is currently not available (Accessed by June 27, 2019). However, authors say that they plan to freely distribute it in the provided link.

Table A.3: Corpora size.

Corpus	Language	Domain	Sentences	Elements	Elements with negation
BioInfer (Pyysalo et al., 2007)	English	Biomedical	1,100	2,662 relations	163 relations (6.12%)
Genia Event (Kim et al., 2008)	English	Biomedical	9,372	36,858 events	2,351 events (6.38%)
BioScope (Vincze et al., 2008)	English	Biomedical	20,924	20,924 sentences	2,720 sentences (13%)
Product Review (Councill et al., 2010)	English	Reviews	2,111	2,111 sentences	679 sentences (32.16%)
Stockholm Electronic Patient Record (Dalianis & Velupillai, 2010)	Swedish	Clinical reports	6,740	6,966 expressions	1,008 expressions (10.67%)
PropBank Focus (PB-FOC) (Blanco & Moldovan, 2011a)	English	Journal stories	3,779	NA	3,993 verbal negations
ConanDoyle-neg (Morante & Daelemans, 2012)	English	Literary	4,423	4,423 sentences	995 sentences (22.5%)
SFU Review$_{EN}$ (Konstantinova et al., 2012)	English	Reviews	17,263	17,263 sentences	3,017 sentences (17.48%)
NEG-DrugDDI (Bokharaeian et al., 2013)	English	Biomedical	5,806	5,806 sentences	1,399 sentences (24.10%)
UAM Spanish Treebank (Sandoval & Salazar, 2013)	Spanish	Newspaper articles	1,500	1,500 sentences	160 sentences (10.67%)
NegDDI-DrugBank (Bokharaeian et al., 2014)	English	Biomedical	6,648	6,648 sentences	1,448 sentences (21.78%)
EMC Dutch (Afzal et al., 2014)	Dutch	Clinical reports	NA	12,852 medical terms	1,804 medical terms (14.04%)
Review and Newspaper Japanese (Matsuyoshi et al., 2014)	Japanese	Reviews and newspaper articles	10,760	10,760 sentences	1,785 sentences (16.59%)
IxaMed-GS (Oronoz et al., 2015)	Spanish	Clinical reports	NA	2,766 entities	763 entities (27.58%)
Deep Tutor Negation (Banjade & Rus, 2016)	English	Tutorial dialogues	NA	27,785 student responses	2,603 student responses (9.37%)
CNeSp (Zou et al., 2016)	Chinese	Scientific literature, product reviews and financial articles	16,841	16,841 sentences	4,517 sentences (26.82%)
German negation and speculation (Cotik, Roller, et al., 2016)	German	Clinical reports	NA	1,114 medical terms	443 medical terms (39.77%)
Fact-Ita Bank Negation (Altuna et al., 2017)	Italian	News articles	1,290	1,290 sentences	278 sentences (21.55%)
SFU Review$_{SP}$-NEG (Jiménez-Zafra et al., 2018)	Spanish	Reviews	9,446	9,446 sentences	2,825 sentences (29.91%)
UHU-HUVR (Cruz Díaz et al., 2017)	Spanish	Clinical reports	8,412	8,412 sentences	2,298 sentences (27.32%)
IULA Spanish Clinical Record (Marimon et al., 2017)	Spanish	Clinical reports	3,194	3,194 sentences	1,093 sentences (34.22%)
SOCC (Kolhatkar et al., 2018)	English	Opinion articles	3,612	3,612 sentences	1,130 sentences (31.28%)

Table A.4: Annotation guidelines.

Corpus	Annotation guidelines
BioInfer (Pyysalo et al., 2007)	http://tucs.fi/publications/view/?pub_id=tGiPyBjHeSa07a
Genia Event (Kim et al., 2008)	http://www.nactem.ac.uk/meta-knowledge/Annotation_Guidelines.pdf
BioScope (Vincze et al., 2008)	http://rgai.inf.u-szeged.hu/project/nlp/bioscope/Annotation%20guidelines2.1.pdf
Product Review (Councill et al., 2010)	(Councill et al., 2010)
Stockholm Electronic Patient Record (Dalianis & Velupillai, 2010)	-
PropBank Focus (PB-FOC) (Blanco & Moldovan, 2011a)	(Blanco & Moldovan, 2011a)
ConanDoyle-neg (Morante & Daelemans, 2012)	(Morante et al., 2011)
SFU Review$_{EN}$ (Konstantinova et al., 2012)	(Konstantinova et al., 2011)
NEG-DrugDDI (Bokharaeian et al., 2013)	-
UAM Spanish Treebank (Sandoval & Salazar, 2013)	(Sandoval & Salazar, 2013)
NegDDI-DrugBank (Bokharaeian et al., 2014)	-
EMC Dutch (Afzal et al., 2014)	-
Review and Newspaper Japanese (Matsuyoshi et al., 2014)	(Matsuyoshi et al., 2014)
IxaMed-GS (Oronoz et al., 2015)	-
Deep Tutor Negation (Banjade & Rus, 2016)	-
CNeSp (Zou et al., 2016)	(Zou et al., 2016)
German negation and speculation (Cotik, Roller, et al., 2016)	-
Fact-Ita Bank Negation (Altuna et al., 2017)	(Altuna et al., 2017)
SFU Review$_{SP}$-NEG (Jiménez-Zafra et al., 2018)	(Martí et al., 2016; Jiménez-Zafra et al., 2018)
UHU-HUVR (Cruz Díaz et al., 2017)	(Cruz Díaz et al., 2017)
IULA Spanish Clinical Record (Marimon et al., 2017)	(Marimon et al., 2017)
SOCC (Kolhatkar et al., 2018)	https://github.com/sfu-discourse-lab/SOCC/tree/master/guidelines

Table A.5: Negation elements (NA: Non-Available, -: Absent, ✓: Present).

Corpus	Negation	Cue	Scope	Event	Focus
BioInfer (Pyysalo et al., 2007)	PS, PM, PL	-	✓	-	-
Genia Event (Kim et al., 2008)	PS, PM, PL	-	-	✓	-
BioScope (Vincze et al., 2008)	CS, CL	✓	✓	-	-
Product Review (Councill et al., 2010)	CS	-	✓	-	-
Stockholm Electronic Patient Record (Dalianis & Velupillai, 2010)	CS	✓	-	-	-
PropBank Focus (PB-FOC) (Blanco & Moldovan, 2011a)	PS	✓	-	-	✓
ConanDoyle-neg (Morante & Daelemans, 2012)	CS	✓	✓	✓	-
SFU Review$_{EN}$ (Konstantinova et al., 2012)	CS	✓	✓	-	-
NEG-DrugDDI (Bokharaeian et al., 2013)	CS, CM, CL	✓	✓	-	-
UAM Spanish Treebank (Sandoval & Salazar, 2013)	CS	✓	✓	-	-
NegDDI-DrugBank (Bokharaeian et al., 2014)	CS, CM, CL	✓	✓	-	-
EMC Dutch (Afzal et al., 2014)	NA	-	-	✓	-
Review and Newspaper Japanese (Matsuyoshi et al., 2014)	CS,CM, CL	✓	-	-	✓
IxaMed-GS (Oronoz et al., 2015)	PS, PM, PL	-	-	✓	-
Deep Tutor Negation (Banjade & Rus, 2016)	CS, CL	✓	✓	-	✓
CNeSp (Zou et al., 2016)	NA	✓	✓	-	-
German negation and speculation (Cotik, Roller, et al., 2016)	NA	-	-	✓	-
Fact-Ita Bank Negation (Altuna et al., 2017)	CS	✓	✓	✓	-
SFU Review$_{SP}$-NEG (Jiménez-Zafra et al., 2018)	CS	✓	✓	✓	-
UHU-HUVR (Cruz Díaz et al., 2017)	CS, CM, CL	✓	✓	✓	-
IULA Spanish Clinical Record (Marimon et al., 2017)	CS, CL	✓	✓	-	-
SOCC (Kolhatkar et al., 2018)	CS, PM, PL	✓	✓	-	✓

NOTE: PS, PM and PL are used when syntactic, morphological and lexical negations are annotated partially. CS, CM and CL represents that all syntactic, morphological and lexical negations have been annotated.

Table A.6: English Corpora annotated with negation (NA: Non-Available, -: Absent, ✓:Present).

Corpus	Availability	Language	Domain	Sentences	Elements	Elements with negation	Negation	Cue	Scope	Event	Focus	Agreement	Format	Is negation the main phenomenon?
BioInfer (Pyysalo et al., 2007)	✓	English	Biomedical	1,100	2,662 relations	163 relations (6.12%)	PS, PM, PL	-	✓	-	-	-	XML	NO
Genia Event (Kim et al., 2008)	✓	English	Biomedical	9,372	36,858 events	2,351 events (6.38%)	PS, PM, PL	-	-	✓	-	-	XML	NO
BioScope (Vincze et al., 2008)	✓	English	Biomedical	20,924	20,924 sentences	2,720 sentences (13%)	CS, CL	✓	✓	-	-	✓	XML	YES
Product Review (Councill et al., 2010)	NA	English	Reviews	2,111	2,111 sentences	679 sentences (32.16%)	CS	-	✓	-	-	✓	NA	YES
PropBank Focus (PB-FOC) (Blanco & Moldovan, 2011a)	✓	English	Journal stories	3,779	NA	3,993 verbal negations	PS	✓	-	-	✓	✓	TXT	YES
ConanDoyle-neg (Morante & Daelemans, 2012)	✓	English	Literary	4,423	4,423 sentences	995 sentences (22.5%)	CS	✓	✓	✓	-	✓	TXT	YES
SFU Review$_{EN}$ (Konstantinova et al., 2012)	✓	English	Reviews	17,263	17,263 sentences	3,017 sentences (17.48%)	CS	✓	✓	-	-	✓	XML	YES
NEG-DrugDDI (Bokharaeian et al., 2013)	✓	English	Biomedical	5,806	5,806 sentences	1,399 sentences (24.10%)	CS, CM, CL	✓	✓	-	-	-	XML	YES
NegDDI-DrugBank (Bokharaeian et al., 2014)	✓	English	Biomedical	6,648	6,648 sentences	1,448 sentences (21.78%)	CS, CM, CL	✓	✓	-	-	-	XML	YES
Deep Tutor Negation (Banjade & Rus, 2016)	✓	English	Tutorial dialogues	NA	27,785 student responses	2,603 student responses (9.37%)	CS, CL	✓	✓	-	✓	✓	TXT	YES
SOCC (Kolhatkar et al., 2018)	✓	English	Opinion articles	3,612	3,612 sentences	1,130 sentences (31.28%)	CS, PL, PM	✓	✓	-	✓	✓	TSV	NO

NOTE: PS, PM and PL are used when syntactic, morphological and lexical negations are annotated partially. CS, CM and CL represents that all syntactic, morphological and lexical negations have been annotated.

Table A.7: Spanish Corpora annotated with negation (NA: Non-Available, -: Absent, ✓:Present).

Corpus	Availability	Language	Domain	Sentences	Elements	Elements with negation	Negation	Cue	Scope	Event	Focus	Agreement	Format	Is negation the main phenomenon?
UAM Spanish Treebank (Sandoval & Salazar, 2013)	✓	Spanish	Newspaper articles	1,500	1,500 sentences	160 sentences (10.67%)	CS	✓	✓	-	-	✓	XML	YES
IxaMed-GS (Oronoz et al., 2015)	NA	Spanish	Clinical reports	NA	2,766 entities	763 entities (27.58%)	PS, PL, PM	-	-	✓	-	✓	NA	NO
SFU Review$_{SP}$-NEG (Jiménez-Zafra et al., 2018)	✓	Spanish	Movies, books, product reviews	9,446	9,446 sentences	2,825 sentences (29.91%)	CS	✓	✓	✓	-	✓	XML	YES
UHU-HUVR (Cruz Díaz et al., 2017)	NA	Spanish	Clinical reports	8,412	8,412 sentences	2,298 sentences (27.32%)	CS, CL, CM	✓	✓	✓	-	✓	NA	YES
IULA Spanish Clinical Record (Marimon et al., 2017)	✓	Spanish	Clinical reports	3,194	3,194 sentences	1,093 sentences (34.22%)	CS, CL	✓	✓	-	-	✓	ANN, TXT	YES

NOTE: PS, PM and PL are used when syntactic, morphological and lexical negations are annotated partially. CS, CM and CL represents that all syntactic, morphological and lexical negations have been annotated.

Table A.8: Other Corpora annotated with negation (NA: Non-Available, -: Absent, ✓:Present).

Corpus	Availability	Language	Domain	Sentences	Elements	Elements with negation	Negation	Cue	Scope	Event	Focus	Agreement	Format	Is negation the main phenomenon?
Stockholm Electronic Patient Record (Dalianis & Velupillai, 2010)	NA	Swedish	Clinical reports	6,740	6,966 expressions	1,008 expressions (10.67%)	CS	✓	-	-	-	✓	NA	YES
EMC Dutch (Afzal et al., 2014)	NA	Dutch	Clinical reports	NA	12,852 medical terms	1,804 medical terms (14.04%)	NA	-	-	✓	-	✓	NA	NO
Review and Newspaper Japanese (Matsuyoshi et al., 2014)	NA	Japanese	Reviews and news articles	10,760	10,760 sentences	1,785 sentences (16.59%)	CS, CM, CL	✓	-	-	✓	✓	NA	YES
CNeSp (Zou et al., 2016)	✓	Chinese	Scientific literature, product reviews, financial articles	16,841	16,841 sentences	4,517 sentences (26.82%)	NA	✓	✓	-	-	✓	XML	YES
German negation and speculation (Cotik, Roller, et al., 2016)	NA	German	Clinical reports	NA	1,114 medical terms	443 medical terms (39.77%)	NA	-	-	✓	-	✓	NA	YES
Facti-Ita Bank Negation (Cotik, Roller, et al., 2016)	✓	Italian	News articles	1,290	1,290 sentences	278 sentences (21.55%)	CS	✓	✓	✓	-	✓	NA	YES

NOTE: PS, PM and PL are used when syntactic, morphological and lexical negations are annotated partially. CS, CM and CL represents that all syntactic, morphological and lexical negations have been annotated.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... others (2016). Tensorflow: a system for large-scale machine learning. In *Prooceedings of osdi 2016* (Vol. 16, pp. 265–283).

Afzal, Z., Pons, E., Kang, N., Sturkenboom, M. C., Schuemie, M. J., & Kors, J. A. (2014). ContextD: an algorithm to identify contextual properties of medical terms in a Dutch clinical corpus. *BMC bioinformatics*, *15*(1), 1.

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

Agresti, A. (2007). *An Introduction to Categorical Data Analysis, Second Edition* (2nd ed.). JohnWileySons. doi: 10.1002/0470114754

Altuna, B., Minard, A.-L., & Speranza, M. (2017). The Scope and Focus of Negation: A Complete Annotation Framework for Italian. In *Proceedings of the workshop computational semantics beyond events and roles* (pp. 34–42).

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

Amores Fernández, M. A. (2016). *Detección de la polaridad de las opiniones basada en nuevos recursos léxicos* (Unpublished doctoral dissertation). Universidad Central "Marta Abreu" de Las Villas.

Ananiadou, S., & McNaught, J. (2006). *Text mining for biology and biomedicine.* Artech House London.

Anta, A. F., Chiroque, L. N., Morere, P., & Santos, A. (2013). Sentiment analysis and topic detection of spanish tweets: A comparative study of of nlp techniques. *Procesamiento del lenguaje natural*, *50*, 45–52.

Appel, O., Chiclana, F., Carter, J., & Fujita, H. (2016). A hybrid approach to the sentiment analysis problem at the sentence level. *Knowledge-Based Systems*, *108*, 110–124.

Asmi, A., & Ishaya, T. (2012). Negation identification and calculation in sentiment analysis. In *The second international conference on advances in information mining and management* (pp. 1–7).

Atserias Batalla, J., Comelles Pujadas, E., & Mayor, A. (2005). TXALA un analizador libre de dependencias para el castellano. *Procesamiento del Lenguaje Natural*, *35*(0). Retrieved from http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/3006

Baccianella, S., Esuli, A., & Sebastiani, F. (2010, may). SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In *Proceedings of the seventh international conference on language resources and evaluation (lrec'10)* (pp. 2200–2204). Valletta, Malta.

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

Ballesteros, M., Francisco, V., Díaz, A., Herrera, J., & Gervás, P. (2012). Inferring the scope of negation in biomedical documents. *Computational Linguistics and Intelligent Text Processing*, 363–375.

Banjade, R., & Rus, V. (2016, may). DT-Neg: Tutorial Dialogues Annotated for Negation Scope and Focus in Context. In N. C. C. Chair) et al. (Eds.), *Proceedings of the tenth international conference on language resources and evaluation (lrec 2016).* Paris, France: European Language Resources Association (ELRA).

Barnes, J. (2019). LTG-Oslo Hierarchical Multi-task Network: The importance of negation for document-level sentiment in Spanish. In *Proceedings of the iberian languages evaluation forum (iberlef 2019).* Bilbao, Spain: CEUR-WS.

Basile, V., Bos, J., Evang, K., & Venhuizen, N. (2012). Developing a large semantically annotated corpus. In *Lrec 2012, eighth international conference on language resources and evaluation* (pp. 3196–3200).

Beltrán, J., & González, M. (2019). Detection of Negation Cues in Spanish: The CLiC-Neg System. In *Proceedings of the iberian languages evaluation forum (iberlef 2019).* Bilbao, Spain: CEUR-WS.

Benamara, F., Chardon, B., Mathieu, Y. Y., Popescu, V., & Asher, N. (2012). 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)* (p. 10-18). Jeju, Korea.

Blanco, E., & Moldovan, D. (2011a, June). Semantic Representation of Negation Using Focus Detection. In *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies* (pp. 581–589). Portland, Oregon, USA: Association for Computational Linguistics. Retrieved from http://www.aclweb.org/anthology/P11-1059

Blanco, E., & Moldovan, D. (2011b). Some issues on detecting negation from text. In *Twenty-fourth international flairs conference* (pp. 228–233).

Blanco, E., & Moldovan, D. (2014). Retrieving implicit positive meaning from negated statements. *Natural Language Engineering*, *20*(4), 501-535.

Blanco, E., & Morante, R. (2018, June). *Proceedings of the workshop on computational semantics beyond events and roles.* New Orleans, Louisiana: Association for Computational Linguistics. Retrieved from https://www.aclweb.org/anthology/W18-1300 doi: 10.18653/v1/W18-13

Blanco, E., Morante, R., & Saurí, R. (2016). Proceedings of the workshop on extra-propositional aspects of meaning in computational linguistics (exprom). In *Proceedings of the workshop on extra-propositional aspects of meaning in computational linguistics (exprom).*

Blanco, E., Morante, R., & Saurí, R. (2017). Proceedings of the workshop computational semantics beyond events and roles. In *Proceedings of the workshop computational semantics beyond events and roles.*

Blanco, E., Morante, R., & Sporleder, C. (2015). Proceedings of the second workshop on extra-propositional aspects of meaning in computational semantics (exprom 2015). In *Proceedings of the second workshop on extra-propositional aspects of meaning in computational semantics (exprom 2015).*

Bokharaeian, B., Diaz, A., Neves, M., & Francisco, V. (2014). Exploring negation annotations in the DrugDDI Corpus. In *Fourth workshop on building and evaluating resources for health and biomedical text processing (biotxtm 2014).*

Bokharaeian, B., Díaz Esteban, A., & Ballesteros Martínez, M. (2013). Extracting Drug-Drug interaction from text using negation features.

Bollegala, D., Mu, T., & Goulermas, J. Y. (2016). Cross-domain sentiment classification using sentiment sensitive embeddings. *IEEE Transactions on Knowledge and Data Engineering*, *28*(2), 398–410.

Boucher, J. D., & Osgood, C. E. (1969). The pollyanna hypothesis. *Journal of Verbal Learning and Verbal Behaviour*, *8*, 1-8.

Brants, T. (2000). TnT: a statistical part-of-speech tagger. In *Proceedings of the sixth conference on applied natural language processing* (pp. 224–231).

Brooke, J., Tofiloski, M., & Taboada, M. (2009). Cross-linguistic sentiment analysis: From english to spanish. In *Proceedings of the international conference ranlp-2009* (pp. 50–54).

Buchholz, S., & Marsi, E. (2006). CoNLL-X shared task on multilingual dependency parsing. In *Proceedings of the tenth conference on computational natural language learning* (pp. 149–164).

Camarero, M. G. (2008). Cuestiones pragmáticas sobre la negación. *redELE: Revista Electrónica de Didáctica ELE*(12), 3.

Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New avenues in opinion mining and sentiment analysis. *IEEE Intelligent systems*, *28*(2), 15–21.

Cardellino, C. (2016). *Spanish Billion Words Corpus and Embeddings.* https://crscardellino.github.io/SBWCE/. Retrieved from https://crscardellino.github.io/SBWCE/ ([Accessed 29-January-2019])

Caselli, T., Lenzi, V. B., Sprugnoli, R., Pianta, E., & Prodanof, I. (2011). Annotating events, temporal expressions and relations in Italian: the It-TimeML experience for the Ita-TimeBank. In *Proceedings of the 5th linguistic annotation workshop* (pp. 143–151).

ChandraKala, S., & Sindhu, C. (2012). Opinion mining and sentiment classification: A survey. *ICTACT journal on soft computing*, *3*(1), 420–425.

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

Chollet, F., et al. (2015). *Keras.* https://keras.io. ([Accessed 29-January-2019])

Cignarella, A. T., Frenda, S., Basile, V., Bosco, C., Patti, V., Rosso, P., et al. (2018). Overview of the evalita 2018 task on irony detection in italian tweets (ironita). In *Proceedings of the 6th evaluation campaign of natural language processing and speech tools for italian (evalita'18)* (pp. 26–34). Retrieved from http://ceur-ws.org/Vol-2263/paper005.pdf

Costumero, R., López, F., Gonzalo-Martín, C., Millan, M., & Menasalvas, E. (2014). An approach to detect negation on medical documents in Spanish. In *International conference on brain informatics and health* (pp. 366–375).

Cotik, V., Roller, R., Xu, F., Uszkoreit, H., BuddeO, K., & SchmidtO, D. (2016). Negation Detection in Clinical Reports Written in German. *BioTxtM 2016*, 115.

Cotik, V., Stricker, V., Vivaldi, J., & 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* (pp. 156–165).

Councill, I. G., McDonald, R., & 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* (pp. 51–59).

Cruz, N. P., Taboada, M., & 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., Morante Vallejo, R., Maña López, M. J., Mata Vázquez, J., & Parra Calderón, C. L. (2017). Annotating Negation in Spanish Clinical Texts. *SemBEaR 2017*, 53.

Cruz Díaz, N. P., Maña López, M. J., Vázquez, J. M., & 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.

Curran, J. R., Clark, S., & Bos, J. (2007). Linguistically motivated large-scale NLP with C&C and Boxer. In *Proceedings of the 45th annual meeting of the acl on interactive poster and demonstration sessions* (pp. 33–36).

Dadvar, M., Hauff, C., & De Jong, F. M. (2011). Scope of negation detection in sentiment analysis. In *Proceedings of the dutch-belgian information retrieval workshop, dir 2011* (pp. 16–19). Twente.

Dalianis, H., Hassel, M., & Velupillai, S. (2009). The Stockholm EPR Corpus–characteristics and some initial findings. *Women*, *219*(906), 54.

Dalianis, H., & Velupillai, S. (2010). How Certain are Clinical Assessments?: Annotating Swedish Clinical Text for (Un) certainties, Speculations and Negations. In *the seventh international conference on language resources and evaluation, lrec 2010* (pp. 3071–3075).

Das, S., & Chen, M. (2001). Yahoo! for Amazon: Extracting Market Sentiment from Stock Message Boards. In *Proceedings of the asia pacific finance association annual conference (apfa).*

de Castilho, R. E., Mujdricza-Maydt, E., Yimam, S. M., Hartmann, S., Gurevych, I., Frank, A., & Biemann, C. (2016). A web-based tool for the integrated annotation of semantic and syntactic structures. In *Proceedings of the workshop on language technology resources and tools for digital humanities (lt4dh)* (pp. 76–84).

Demonte, V., & Bosque, I. (1999). *Gramática descriptiva de la lengua española.* Espasa Calpe.

Diamantini, C., Mircoli, A., & Potena, D. (2016). A negation handling technique for sentiment analysis. In *2016 international conference on collaboration technologies and systems (cts)* (pp. 188–195).

Díaz, N. P. C., & López, M. J. M. (2019). *Negation and speculation detection* (Vol. 13). John Benjamins Publishing Company.

Domínguez-Mas, L., Ronzano, F., & Furlong, L. I. (2019). Supervised learning approaches to detect negation cues in Spanish reviews. In *Proceedings of the iberian languages evaluation forum (iberlef 2019).* Bilbao, Spain: CEUR-WS.

Donatelli, L. (2018). Cues, Scope, and Focus: Annotating Negation in Spanish Corpora. In *Proceedings of NEGES 2018: Workshop on Negation in Spanish, ceur workshop proceedings* (Vol. 2174, pp. 29–34).

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.

Española, R. A. (2009). Nueva gramática de la lengua española.

Fabregat, H., Duque, A., Martínez-Romo, J., & Araujo, L. (2019). Extending a Deep Learning approach for Negation Cues Detection in Spanish. In *Proceedings of the iberian languages evaluation forum (iberlef 2019).* Bilbao, Spain: CEUR-WS.

Fabregat, H., Martínez-Romo, J., & 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* (Vol. 2174, pp. 43–48).

Fancellu, F., Reddy, S., Lopez, A., & Webber, B. (2017). Universal Dependencies to Logical Form with Negation Scope. In *Proceedings of the workshop computational semantics beyond events and roles* (pp. 22–32).

Farkas, R., Vincze, V., Móra, G., Csirik, J., & 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—shared task* (pp. 1–12).

Farooq, U., Mansoor, H., Nongaillard, A., Ouzrout, Y., & Qadir, M. A. (2017). Negation handling in sentiment analysis at sentence level. *JCP*, *12*(5), 470–478.

Flickinger, D. (2000). On building a more effcient grammar by exploiting types. *Natural Language Engineering*, *6*(1), 15–28.

Flickinger, D., Zhang, Y., & Kordoni, V. (2012). Deepbank. A dynamically annotated treebank of the Wall Street Journal. In *Proceedings of the 11th international workshop on treebanks and linguistic theories* (pp. 85–96).

Gamallo, P., Garcia, M., & Fernández-Lanza, S. (2013). Tass: A naive-bayes strategy for sentiment analysis on spanish tweets. In *Workshop on sentiment analysis at sepln (tass2013)* (pp. 126–132).

Ghiassi, M., Skinner, J., & Zimbra, D. (2013). Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Systems with applications*, *40*(16), 6266–6282.

Giudice, V. (2018). Aspie96 at IronITA (EVALITA 2018): Irony Detection in Italian Tweets with Character-Level Convolutional RNN. In *Proceedings of the 6th evaluation campaign of natural language processing and speech tools for italian (evalita'18)* (pp. 160–165). Retrieved from http://ceur-ws.org/Vol-2263/paper026.pdf

Giudice, V. (2019a). Aspie96 at FACT (IberLEF 2019): Factuality Classification in Spanish Texts with Character-Level Convolutional RNN and Tokenization. In *Proceedings of the iberian languages evaluation forum (iberlef 2019).* Bilbao, Spain: CEUR-WS.

Giudice, V. (2019b). 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).* Bilbao, Spain: CEUR-WS.

Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford*, *1*(12).

Harkema, H., Dowling, J. N., Thornblade, T., & 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.

Herrero Zazo, M., Segura Bedmar, I., Martínez, P., & Declerck, T. (2013). The DDI corpus: An annotated corpus with pharmacological substances and drug–drug interactions. *Journal of biomedical informatics*, *46*(5), 914–920.

Hirschberg, J., & Manning, C. D. (2015). Advances in natural language processing. *Science*, *349*(6245), 261–266.

Hogenboom, A., Van Iterson, P., Heerschop, B., Frasincar, F., & Kaymak, U. (2011). Determining negation scope and strength in sentiment analysis. In *2011 ieee international conference on systems, man, and cybernetics* (pp. 2589–2594).

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

Horn, L. R. (1989). *A natural history of negation.* CSLI Publications.

Horn, L. R., & Wansing, H. (2017). Negation. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2017 ed.). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2017/entries/negation/.

Hu, M., & Liu, B. (2004). Mining and Summarizing Customer Reviews. In *Proceedings of the tenth acm sigkdd international conference on knowledge discovery and data mining* (pp. 168–177). New York, NY, USA: ACM. Retrieved from http://doi.acm.org/10.1145/1014052.1014073 doi: 10.1145/1014052.1014073

Hussein, D. M. E.-D. M. (2018). A survey on sentiment analysis challenges. *Journal of King Saud University-Engineering Sciences*, *30*(4), 330–338.

Ide, N. (2017). Introduction: The handbook of linguistic annotation. In *Handbook of linguistic annotation* (pp. 1–18). Springer.

Ide, N., Fellbaum, C., Baker, C., & Passonneau, R. (2010). The manually annotated sub-corpus: A community resource for and by the people. In *Proceedings of the acl 2010 conference short papers* (pp. 68–73).

Israel, M. (2004). The pragmatics of polarity. In L. Horn & G. Ward (Eds.), *The handbook of pragmatics* (p. 701-723). Malden, MA: Blackwell.

Jia, L., Yu, C., & 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* (pp. 1827–1830).

Jiang, L., Yu, M., Zhou, M., Liu, X., & Zhao, T. (2011). Target-dependent twitter sentiment classification. In *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies-volume 1* (pp. 151–160).

Jiménez-Zafra, S., Martín-Valdivia, M., Martínez-Cámara, E., & Ureña-López, L. (2016). Combining resources to improve unsupervised sentiment analysis at aspect-level. *Journal of Information Science*, *42*(2), 213–229.

Jiménez-Zafra, S., Martín-Valdivia, M., Martínez-Cámara, E., & Ureña-López, L. (2019, Jan). Studying the scope of negation for spanish sentiment analysis on twitter. *IEEE Transactions on Affective Computing*, *10*(1), 129-141. (First published online on April 12, 2017) doi: 10.1109/TAFFC.2017.2693968

Jiménez-Zafra, S., Taulé, M., Martín-Valdivia, M., Ureña-López, L., & 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. Retrieved from http://dx.doi.org/10.1007/s10579-017-9391-x doi: 10.1007/s10579-017-9391-x

Jiménez-Zafra, S. M., Cruz Díaz, N. P., Morante, R., & Martín-Valdivia, M. T. (2018a). Tarea 1 del taller neges 2018: Guías de anotación. In *Proceedings of neges 2018: Workshop on negation in spanish* (Vol. 2174, pp. 15–21).

Jiménez-Zafra, S. M., Cruz Díaz, N. P., Morante, R., & Martín-Valdivia, M. T. (2018b). Tarea 2 del Taller NEGES 2018: Detección de Claves de Negación. In *Proceedings of neges 2018: Workshop on negation in spanish* (Vol. 2174, pp. 35–41).

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

Jiménez-Zafra, S. M., Cruz Díaz, N. P., Morante, R., & Martín-Valdivia, M. T. (2019b). NEGES 2019 task: Negation in Spanish. In *Proceedings of the iberian languages evaluation forum (iberlef 2019).* Bilbao, Spain: CEUR-WS.

Jiménez-Zafra, S. M., Martínez-Cámara, E., Martín-Valdivia, M. T., & 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., Martín-Valdivia, M., Ureña-López, L., Marti, T., & Taulé, M. (2016). Problematic cases in the annotation of negation in Spanish. In *Proceedings of the workshop on extra-propositional aspects of meaning in computational linguistics (exprom)* (pp. 42–48).

Jiménez-Zafra, S. M., Morante, R., Blanco, E., Martín-Valdivia, M., & Ureña-López, L. (2019). Detecting Negation Cues and Scopes in Spanish. In *Proceedings of the 23nd Conference on Computational Natural Language Learning - (Under review)* (pp. XXX–XXX).

Kennedy, A., & Inkpen, D. (2006). Sentiment Classification Of Movie Reviews Using Contextual Valence Shifters. *Computational Intelligence*, *22*(2), 110–125. Retrieved from http://dx.doi.org/10.1111/j.1467-8640.2006.00277.x doi: 10.1111/j.1467-8640.2006.00277.x

Khan, A. Z., Atique, M., & Thakare, V. (2015). Combining lexicon-based and learning-based methods for twitter sentiment analysis. *International Journal of Electronics, Communication and Soft Computing Science & Engineering (IJECSCSE)*, 89-91.

Kim, J.-D., Ohta, T., Pyysalo, S., Kano, Y., & 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* (pp. 1–9).

Kim, J.-D., Ohta, T., Tateisi, Y., & Tsujii, J. (2003). GENIA corpus—a semantically annotated corpus for bio-textmining. *Bioinformatics*, *19*(suppl_1), i180–i182.

Kim, J.-D., Ohta, T., & Tsujii, J. (2008). Corpus annotation for mining biomedical events from literature. *BMC bioinformatics*, *9*(1), 1.

Kolhatkar, V., Wu, H., Cavasso, L., Francis, E., Shukla, K., & Taboada, M. (2018). The SFU Opinion and Comments Corpus: A corpus for the analysis of online news comments. *Under review*.

Kong, L., Schneider, N., Swayamdipta, S., Bhatia, A., Dyer, C., & Smith, N. A. (2014). A dependency parser for tweets.

Konstantinova, N., De Sousa, S. C., Díaz Cruz, N., López, M. J. M., Taboada, M., & Mitkov, R. (2012). A review corpus annotated for negation, speculation and their scope. In *Lrec* (pp. 3190–3195).

Konstantinova, N., De Sousa, S. C., & Sheila, J. (2011). Annotating Negation and Speculation: the Case of the Review Domain. In *Ranlp student research workshop* (pp. 139–144).

Kouloumpis, E., Wilson, T., & Moore, J. D. (2011). Twitter sentiment analysis: The good the bad and the omg! *Icwsm*, *11*(538-541), 164.

Lafferty, J. D., McCallum, A., & 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* (pp. 282–289). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc. Retrieved from http://dl.acm.org/citation.cfm?id=645530.655813

Lapponi, E., Read, J., & Øvrelid, L. (2012). Representing and resolving negation for sentiment analysis. In *2012 ieee 12th international conference on data mining workshops* (pp. 687–692).

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

Li, F., Pan, S. J., Jin, O., Yang, Q., & Zhu, X. (2012). Cross-domain co-extraction of sentiment and topic lexicons. In *Proceedings of the 50th annual meeting of the association for computational linguistics: Long papers-volume 1* (pp. 410–419).

Li, J., Zhou, G., Wang, H., & Zhu, Q. (2010). Learning the scope of negation via shallow semantic parsing. In *Proceedings of the 23rd international conference on computational linguistics* (pp. 671–679).

Liddy, E. D., Paik, W., McKenna, M. E., Weiner, M. L., Edmund, S. Y., Diamond, T. G., ... Snyder, D. L. (2000, February 15). *User interface and other enhancements for natural language information retrieval system and method.* Google Patents. (US Patent 6,026,388)

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

Loharja, H., Padró, L., & 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* (Vol. 2174, pp. 49–54).

Loper, E., & Bird, S. (2002). NLTK: The natural language toolkit. In *Proceedings of the acl-02 workshop on effective tools and methodologies for teaching natural language processing and computational linguistics-volume 1* (pp. 63–70).

Marimon, M., Vivaldi, J., Bel, N., & Boronat, R. (2017). Annotation of negation in the IULA Spanish Clinical Record Corpus. *SemBEaR 2017*, *5*(36.41), 43.

Martí, M. A., Martín-Valdivia, M. T., Taulé, M., Jiménez-Zafra, S. M., Nofre, M., & Marsó, L. (2016). La negación en español: análisis y tipología de patrones de negación. *Procesamiento del Lenguaje Natural*, *57*, 41–48.

Martí, M. A., & Taulé, M. (2018). Análisis Comparativo de los Sistemas de Anotación de la Negación en Español. In *Proceedings of NEGES 2018: Workshop on Negation in Spanish, ceur workshop proceedings* (Vol. 2174, pp. 23–28).

Martínez-Cámara, E., Martín-Valdivia, M. T., Molina-González, M. D., & Perea-Ortega, J. M. (2014). Integrating spanish lexical resources by meta-classifiers for polarity classification. *Journal of Information Science*, *40*(4), 538–554.

Martínez-Cámara, E., Martín-Valdivia, M. T., Ureña López, L. A., & Mitkov, R. (2015). Polarity classification for Spanish tweets using the COST corpus. *Journal of Information Science*. Retrieved from http://jis.sagepub.com/content/early/2015/02/02/0165551514566564.abstract doi: 10.1177/0165551514566564

Matsuyoshi, S., Otsuki, R., & Fukumoto, F. (2014). Annotating the Focus of Negation in Japanese Text. In *Lrec* (pp. 1743–1750).

Maynard, D., & Funk, A. (2011). Automatic detection of political opinions in tweets. In *Extended semantic web conference* (pp. 88–99).

McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, *12*(2), 153-157. Retrieved from http://dx.doi.org/10.1007/BF02295996 doi: 10.1007/BF02295996

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

Minard, A.-L., Marchetti, A., & Speranza, M. (2014). Event factuality in italian: Annotation of news stories from the ita-timebank. In *First italian conference on computational linguistics* (pp. 260–264).

Minard, A.-L., Speranza, M., & Caselli, T. (2016). The EVALITA 2016 Event Factuality Annotation Task (FactA). In *Clic-it/evalita.*

Miranda, C. H., Guzmán, J., & 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.

Mitchell, K. J., Becich, M. J., Berman, J. J., Chapman, W. W., Gilbertson, J. R., Gupta, D., . . . Crowley, R. S. (2004). Implementation and evaluation of a negation tagger in a pipeline-based system for information extraction from pathology reports. In *Medinfo* (pp. 663–667).

Mohammad, S., Kiritchenko, S., & Zhu, X. (2013, June). NRC-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets. In *Second joint conference on lexical and computational semantics (*sem), volume 2: Proceedings of the seventh international workshop on semantic evaluation (semeval 2013)* (pp. 321–327). Atlanta, Georgia, USA: Association for Computational Linguistics. Retrieved from http://www.aclweb.org/anthology/S13-2053

Mohammad, S. M. (2012). #Emotional Tweets. In *Proceedings of 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* (pp. 246–255). Stroudsburg, PA, USA: Association for Computational Linguistics. Retrieved from http://dl.acm.org/citation.cfm?id=2387636.2387676

Molina-González, M. D., Martínez-Cámara, E., Martín-Valdivia, M.-T., & Perea-Ortega, J. M. (2013). Semantic orientation for polarity classification in spanish reviews. *Expert Systems with Applications*, *40*(18), 7250–7257.

Molina-González, M. D., Martínez-Cámara, E., Martín-Valdivia, M. T., & Ureña-López, L. A. (2015). A spanish semantic orientation approach to domain adaptation for polarity classification. *Information Processing & Management*, *51*(4), 520–531.

Montejo-Ráez, A., Díaz-Galiano, M. C., Martínez-Santiago, F., & Ureña López, L. A. (2014, October). Crowd explicit sentiment analysis. *Know.-Based Syst.*, *69*(1), 134–139. doi: 10.1016/j.knosys.2014.05.007

Montejo-Ráez, A., Martínez-Cámara, E., Martín-Valdivia, M. T., & Ureña López, L. A. (2014). A knowledge-based approach for polarity classification in Twitter. *Journal of the Association for Information Science and Technology*, *65*(2), 414–425. Retrieved from http://dx.doi.org/10.1002/asi.22984 doi: 10.1002/asi.22984

Morante, R. (2010). Descriptive analysis of negation cues in biomedical texts. In *Proceedings of the seventh conference on international language resources and evaluation (lrec'10).*

Morante, R., & Blanco, E. (2012a). * 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)* (Vol. 1, pp. 265–274).

Morante, R., & Blanco, E. (2012b, 7-8 June). *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)* (pp. 265–274). Montréal, Canada: Association for Computational Linguistics. Retrieved from http://www.aclweb.org/anthology/S12-1035

Morante, R., & Daelemans, W. (2012). ConanDoyle-neg: Annotation of negation in Conan Doyle stories. In *Proceedings of the eighth international conference on language resources and evaluation, istanbul.*

Morante, R., Liekens, A., & Daelemans, W. (2008). Learning the scope of negation in biomedical texts. In *Proceedings of the conference on empirical methods in natural language processing* (pp. 715–724).

Morante, R., Schrauwen, S., & Daelemans, W. (2011). Annotation of negation cues and their scope: Guidelines v1. *Computational linguistics and psycholinguistics technical report series, CTRS-003*.

Morante, R., & Sporleder, C. (2010). Proceedings of the workshop on negation and speculation in natural language processing. In *Proceedings of the workshop on negation and speculation in natural language processing* (pp. 1–109).

Morante, R., & Sporleder, C. (2012a). Modality and negation: An introduction to the special issue. *Computational linguistics*, *38*(2), 223–260.

Morante, R., & Sporleder, C. (2012b). Proceedings of the workshop on extra-propositional aspects of meaning in computational linguistics. In *Proceedings of the workshop on extra-propositional aspects of meaning in computational linguistics.*

Moreno, A., López, S., Sánchez, F., & Grishman, R. (2003). Developing a syntactic annotation scheme and tools for a Spanish treebank. In *Treebanks* (pp. 149–163). Springer.

Mowery, D. L., Velupillai, S., South, B. R., Christensen, L., Martinez, D., Kelly, L., ... others (2014). Task 2: Share/clef ehealth evaluation lab 2014. In *Proceedings of clef 2014.*

Mutalik, P. G., Deshpande, A., & 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.

Nakov, P., Kozareva, Z., Ritter, A., Rosenthal, S., Stoyanov, V., & Wilson, T. (2013). Semeval-2013 task 2: Sentiment analysis in twitter. In *Second joint conference on lexical and compu-*

*tational semantics (*sem), volume 2: Proceedings of the seventh international workshop on semantic evaluation (semeval 2013)* (pp. 312–320).

Ohana, B., Tierney, B., & Delany, S. J. (2016). Sentiment classification using negation as a proxy for negative sentiment. In *The twenty-ninth international flairs conference.*

Ohta, T., Tateisi, Y., & Kim, J.-D. (2002). The GENIA corpus: An annotated research abstract corpus in molecular biology domain. In *Proceedings of the second international conference on human language technology research* (pp. 82–86).

Okazaki, N. (2007). *CRFsuite: a fast implementation of Conditional Random Fields (CRFs).* Retrieved from http://www.chokkan.org/software/crfsuite/

Oronoz, M., Gojenola, K., Pérez, A., D'ıaz de Ilarraza, A., & Casillas, A. (2015). On the creation of a clinical gold standard corpus in Spanish: Mining adverse drug reactions. *Journal of biomedical informatics*, *56*, 318–332.

Padró, L. (1998). A Hybrid Environment for Syntax-Semantic Tagging. *CoRR*, *cmp-lg/9802002*. Retrieved from http://arxiv.org/abs/cmp-lg/9802002

Padró, L., & Stanilovsky, E. (2012, May). Freeling 3.0: Towards Wider Multilinguality. In *Proceedings of the language resources and evaluation conference (lrec 2012).* Istanbul, Turkey.

Palmer, M., Gildea, D., & Kingsbury, P. (2005). The proposition bank: An annotated corpus of semantic roles. *Computational linguistics*, *31*(1), 71–106.

Pang, B., Lee, L., et al. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, *2*(1–2), 1–135.

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the acl-02 conference on empirical methods in natural language processing-volume 10* (pp. 79–86).

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Polanyi, L., & Zaenen, A. (2006). Contextual valence shifters. In J. G. Shanahan, Y. Qu, & J. Wiebe (Eds.), *Computing attitude and affect in text: Theory and applications* (p. 1-10). Dordrecht: Springer.

Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., AL-Smadi, M., ... Eryigit, G. (2016). Semeval-2016 task 5: Aspect based sentiment analysis. In (pp. 19–30).

Potts, C. (2011a). On the negativity of negation. In *Proceedings of salt 20: Semantics and linguistic theory* (p. 636-659). Vancouver.

Potts, C. (2011b). Sentiment symposium tutorial. In *Sentiment analysis symposium.* San Francisco, California, November: Alta Plana Corporation. Retrieved from http://sentiment.christopherpotss.net/

Pröllochs, N., Feuerriegel, S., & Neumann, D. (2016). Negation scope detection in sentiment analysis: Decision support for news-driven trading. *Decision Support Systems*, *88*, 67–75.

Pyysalo, S., Ginter, F., Heimonen, J., Björne, J., Boberg, J., Järvinen, J., & Salakoski, T. (2007). BioInfer: a corpus for information extraction in the biomedical domain. *BMC bioinformatics*, *8*(1), 1.

Qian, Z., Li, P., Zhu, Q., Zhou, G., Luo, Z., & 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* (pp. 815–825).

Ramshaw, L., & Marcus, M. (1995). Text chunking using transformation-based learning. In *Proceedings of the 3rd workshop on very large corpora (acl 1995)* (pp. 82–94).

Read, J. (2005). Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In *Proceedings of the acl student research workshop* (pp. 43–48).

Redondo, J., Fraga, I., Padrón, I., & Comesaña, M. (2007). The spanish adaptation of anew (affective norms for english words). *Behavior research methods*, *39*(3), 600–605.

Reitan, J., Faret, J., Gambäck, B., & 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* (pp. 99–108).

Ren, Y., Fei, H., & 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)* (pp. 739–742).

Ribeiro, F. N., Araújo, M., Gonçalves, P., Gonçalves, M. A., & Benevenuto, F. (2016). Sentibench: A benchmark comparison of state-of-the-practice sentiment analysis methods. *EPJ Data Science*, *5*(23), 1-29.

Rosenthal, S., Farra, N., & Nakov, P. (2017). Semeval-2017 task 4: Sentiment analysis in twitter. In *Proceedings of the 11th international workshop on semantic evaluation (semeval-2017)* (pp. 502–518).

Sandoval, A. M., & Salazar, M. G. (2013). La anotación de la negación en un corpus escrito etiquetado sintácticamente. Annotation of negation in a written treebank. *Revista Iberoamericana de Linguistica*, *8*.

Saurí, R., & Pustejovsky, J. (2009). FactBank: A corpus annotated with event factuality. *Language resources and evaluation*, *43*(3), 227–268.

Saurí, R. (2008). *A Factuality Profiler for Eventualities in Text* (Ph.D. dissertation). Brandeis University.

Savova, G. K., Masanz, J. J., Ogren, P. V., Zheng, J., Sohn, S., Kipper-Schuler, K. C., & 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.

Sebastiani, F. (2002, March). Machine learning in automated text categorization. *ACM Computing Surveys*, *34*(1), 1–47. Retrieved from http://doi.acm.org/10.1145/505282.505283 doi: 10.1145/505282.505283

Segura Bedmar, I., Martinez, P., & de Pablo Sánchez, C. (2011). Using a shallow linguistic kernel for drug–drug interaction extraction. *Journal of biomedical informatics*, *44*(5), 789–804.

Sharif, W., Samsudin, N. A., Deris, M. M., & Naseem, R. (2016). Effect of negation in sentiment analysis. In *2016 sixth international conference on innovative computing technology (intech)* (pp. 718–723).

Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A., & Potts, C. (2013). Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 conference on empirical methods in natural language processing* (pp. 1631–1642).

Stavrianou, A., Brun, C., Silander, T., & Roux, C. (2014). Nlp-based feature extraction for automated tweet classification. *Interactions between Data Mining and Natural Language Processing*, 145.

Stone, P. J., Dunphy, D. C., & Smith, M. S. (1966). *The General Inquirer: A Computer Approach to Content Analysis.* M.I.T. press.

Stricker, V., Iacobacci, I., & 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.*

Styler IV, W. F., Bethard, S., Finan, S., Palmer, M., Pradhan, S., de Groen, P., ... others (2014). Temporal annotation in the clinical domain. *Transactions of the Association for Computational Linguistics*, *2*, 143–154.

Szarvas, G., Vincze, V., Farkas, R., & Csirik, J. (2008). The BioScope corpus: annotation for negation, uncertainty and their scope in biomedical texts. In *Proceedings of the workshop on current trends in biomedical natural language processing* (pp. 38–45).

Szarvas, G., Vincze, V., Farkas, R., Móra, G., & Gurevych, I. (2012). Cross-genre and cross-domain detection of semantic uncertainty. *Computational Linguistics*, *38*(2), 335–367.

Taboada, M. (2016). Sentiment analysis: An overview from Linguistics. *Annual Review of Linguistics*, *2*, 325-347.

Taboada, M., Anthony, C., & Voll, K. (2006). Methods for creating semantic orientation dictionaries. In *Proceedings of the 5th conference on language resources and evaluation (lrec'06)* (pp. 427–432).

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

Tang, D., Wei, F., Yang, N., Zhou, M., Liu, T., & Qin, B. (2014). Learning sentiment-specific word embedding for twitter sentiment classification. In *Proceedings of the 52nd annual meeting of the association for computational linguistics (volume 1: Long papers)* (Vol. 1, pp. 1555–1565).

Taylor, A., Marcus, M., & Santorini, B. (2003). The Penn treebank: an overview. In *Treebanks* (pp. 5–22). Springer.

Thet, T. T., Na, J.-C., & Khoo, C. S. (2010). Aspect-based sentiment analysis of movie reviews on discussion boards. *Journal of information science*, *36*(6), 823–848.

Tottie, G. (1993). Negation in English Speech and Writing: A Study in Variation. *Language*, *69*(3), 590–593.

Turney, P. D. (2002). Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th annual meeting on association for computational linguistics* (pp. 417–424).

Uzuner, Ö., South, B. R., Shen, S., & 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., & Gómez-Rodríguez, C. (2013). Clasificación de polaridad en textos con opiniones en español mediante análisis sintáctico de dependencias. *Procesamiento del lenguaje natural*, *50*, 13–20.

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

Villena-Román, J., Lana-Serrano, S., Martínez-Cámara, E., & González-Cristóbal, J. C. (2013). Tass - workshop on sentiment analysis at sepln. *Procesamiento del Lenguaje Natural*, *50*(0), 37–44. Retrieved from http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/4657

Vincze, V. (2010). Speculation and negation annotation in natural language texts: what the case of BioScope might (not) reveal. In *Proceedings of the workshop on negation and speculation in natural language processing* (pp. 28–31).

Vincze, V., Szarvas, G., Farkas, R., Móra, G., & Csirik, J. (2008). The BioScope corpus: biomedical texts annotated for uncertainty, negation and their scopes. *BMC bioinformatics*, *9*(11), 1.

Wang, H., Lu, Y., & Zhai, C. (2011). Latent aspect rating analysis without aspect keyword supervision. In *Proceedings of the 17th acm sigkdd international conference on knowledge discovery and data mining* (pp. 618–626).

Wiegand, M., Balahur, A., Roth, B., Klakow, D., & Montoyo, A. (2010). A survey on the role of negation in sentiment analysis. In *Proceedings of the workshop on negation and speculation in natural language processing* (p. 60-68). Uppsala, Sweden.

Wilson, T., Wiebe, J., & Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of human language technology conference and conference on empirical methods in natural language processing.*

Yu, H., & Hatzivassiloglou, V. (2003). Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of the 2003 conference on empirical methods in natural language processing* (pp. 129–136).

Zou, B., Zhou, G., & Zhu, Q. (2016). Research on Chinese negation and speculation: corpus annotation and identification. *Frontiers of Computer Science*, *10*(6), 1039–1051.