Corpora Annotated with Negation: An Overview

Salud María Jiménez-Zafra
SINAI, Computer Science Department
CEATIC, Universidad de Jaén
sjzafra@ujaen.es

Roser Morante
CLTL Lab, Computational Linguistics
VU University Amsterdam
r.morantevallejo@vu.nl

María Teresa Martín-Valdivia
SINAI, Computer Science Department
CEATIC, Universidad de Jaén
maite@ujaen.es

L. Alfonso Ureña-López
SINAI, Computer Science Department
CEATIC, Universidad de Jaén
laurena@ujaen.es

Negation is a universal linguistic phenomenon with a great qualitative impact on natural language processing applications. The availability of corpora annotated with negation is essential to training negation processing systems. Currently, most 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. In this study, we present a review of the corpora annotated with negation information in several languages with the goal of evaluating what aspects of negation have been annotated and how compatible the corpora are. We conclude that it is very difficult to merge the existing corpora because we found differences in the annotation schemes used, and most importantly, in the annotation guidelines: the way in which each corpus was tokenized and the negation elements that have been annotated. Differently than for other well established tasks like semantic role labeling or parsing, for negation there is no standard annotation scheme nor guidelines, which hampers progress in its treatment.

Submission received: 4 December 2018; revised version received: 23 October 2019; accepted for publication: 17 November 2019.

https://doi.org/10.1162/COLI_a_00371

1. Introduction

Negation is a key universal phenomenon in language. All languages possess different types of resources (morphological, lexical, syntactic) that allow speakers to speak about properties that people or things do not hold or events that do not happen. The presence of a negation in a sentence can have enormous consequences in many real world situations: A world in which Donald Trump was elected as president would be very different from a world in which Donald Trump was not elected as president, for example. Thus, the presence of a single particle modifying a proposition describes a completely different situation. Negation is a main 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.

Natural language processing (NLP) 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 and 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 NLP 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 and, as a matter fact, there are no publicly available off-the-shelf tools that can be easily incorporated into applications to detect negations.

Work on negation started in 2001 with the aim of processing clinical records (Chapman et al. 2001a; Mutalik, Deshpande, and Nadkarni 2001; Goldin and Chapman 2003). Some rule-based systems were developed based on lists of negations and stop words (Mitchell et al. 2004; Harkema et al. 2009; Mykowiecka, Marciniak, and Kupść 2009; Uzuner, Zhang, and Sibanda 2009; Sohn, Wu, and Chute 2012). With the surge of opinion mining, negation was studied as a marker of polarity change (Das and Chen 2001; Wilson, Wiebe, and Hoffmann 2005; Polanyi and Zaenen 2006; Taboada et al. 2011; Jiménez-Zafra et al. 2017). Only with the release of the BioScope corpus (Vincze et al. 2008) did the work on negation receive a boost. But even so, despite the existence of several publications that focus on negation, it is difficult to find a negation processor for languages other than English. For English, some systems are available for processing clinical documents (NegEx [Chapman et al. 2001b], ConText [Harkema et al. 2009], Deepen [Mehrabi et al. 2015]) and, recently, a tool for detecting negation cues and scopes in natural language texts has been published (Enger, Velldal, and Øvrelid 2017).

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 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. 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 cue detection. In this article we focus on reviewing existing corpora annotated with negation, without entering in the realm of reviewing negation processing systems.

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. 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 modeling factuality for its computational treatment (Saurí and 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%, whereas 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.

Processing 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 (*"recipes with milk and cheese"*) than to make the negated version of the search (*"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 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 whereas 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 online 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 indicates that it is a positive text, although in fact it is negative.

One of the first steps when attempting to develop 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 labeling 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 characterize 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, Zhang, and Kordoni 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 and Daelemans 2012]). It should be noted that we focus on corpora that deal with *string-level* negation.

The rest of the article is organized as follows: In Section 2 previous overviews that focus on negation are presented; in Section 3 the criteria used to review the existing corpora annotated with negation are described; in Sections 4, 5, and 6 the existing corpora for English, Spanish, and other languages are reviewed; in Section 7 we briefly describe negation processing systems that have been developed using the corpora; in Sections 8 and 9 the corpora are analyzed showing features of interest, applications for which they can be used, and problems found for the development of negation processing systems; and finally, conclusions are drawn in Section 10.

2. Related Work

To the best of our knowledge, there are currently no extensive reviews of corpora annotated with negation, but there are overviews that focus on the role of negation. An interesting overview on how modality and negation have been modeled in computational linguistics was presented by Morante and Sporleder (2012). The authors emphasize that most research in NLP 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,

1 `http://www.mrtuit.com/`.
2 `https://www.meaningcloud.com/es/productos/analisis-de-sentimiento`.

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 and 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.

In relation to the modeling of negation, we can reference one survey about the role of negation in sentiment analysis (Wiegand et al. 2010). In this survey, several papers with novel approaches to modeling negation in sentiment analysis are presented. Sentiment analysis focuses on the automatic detection and classification of opinions expressed in texts; and negation can affect the polarity of a word (usually positive, negative, or neutral) because it can change, increment, or reduce the polarity value, hence the importance of dealing with this phenomenon in this area. The authors study the level of representation used for sentiment analysis, negation word detection, and scope of negation. In relation to the representation of negation, the usual way to incorporate negation in supervised machine learning is to use a bag-of-words model adding a new feature *NOT_x*. Thus, if a word *x* is preceded by a negation marker (e.g., not, never), it would be represented as *NOT_x* and as *x* in any other case. Pang, Lee, and Vaithyanathan (2002) followed a similar approach but they added the tag *NOT* to every word between a negation cue and the first punctuation mark. They found that the effect of adding negation was relatively small, probably because the introduction of the feature *NOT_x* increased the feature space. Later, negation was modeled as a polarity shifter and not only negation was considered, but also intensifiers and diminishers. Negation was incorporated into models including knowledge of polar expressions by changing the polarity of an expression (Polanyi and Zaenen 2004; Kennedy and Inkpen 2006) or encoding negation as features using polar expressions (negation features, shifter features, and polarity modification features) (Wilson, Wiebe, and Hoffmann 2005). The results obtained with these models led to a significant improvement over the bag-of-words model. The conclusion drawn by the authors of this survey is that negation is highly relevant to sentiment analysis and that for a negation model to be effective in this area, knowledge of polar expressions is required. Moreover, they state that negation markers do not always function as negators and, consequently, need to be disambiguated. Another interesting remark is that, despite the existence of several approaches to modeling negation for sentiment analysis, to make affirmations of the effectiveness of the methods it is necessary to carry out comparative analysis with regard to classification type, text granularity, target domain, language, and so forth. The papers presented in this study are the pioneering studies of negation modeling in sentiment analysis for English texts. In recent studies researchers have been developing rule-based systems using syntactic dependency trees (Jia, Yu, and Meng 2009), applying more complex calculations in order to obtain polarity (Taboada et al. 2011), using deep-learning (Socher et al. 2013), and using machine-learning with lexical and syntactic features (Cruz, Taboada, and Mitkov 2016a).

The studies analyzed above were carried out on English texts, but interest in processing negation in languages other than English has been increasing in recent years. Jiménez-Zafra et al. (2018a) recently presented a review of Spanish corpora annotated with negation. The authors consulted the main catalogs and platforms that provide information about resources and/or access to them (LDC catalog,[3] ELRA catalog,[4] LRE Map,[5] META-SHARE,[6] and ReTeLe[7]) with the aim of developing a negation processing system for Spanish. Because of the difficulty in finding corpora annotated with negation in Spanish, they conducted an exhaustive search of these resources. As a result, they provided a description of the corpora found as well as the direct links for accessing the data where possible. Moreover, the main features of the corpora were analyzed in order to determine whether the existing annotation schemes account for the complexity of negation in Spanish, that is, whether the typology of negation patterns in this language (Marti et al. 2016) was taken into account in the existing annotation guidelines. The conclusions drawn from this analysis were that the Spanish corpora are very different in several aspects: the genres, the annotation guidelines, and the aspects of negation that have been annotated. As a consequence, it would not be possible to merge all of them to training a negation processing system.

3. Criteria for Corpus Review

An essential requirement for developing machine learning systems is the availability of annotated corpora, and also that the corpora be large enough and the annotations consistent. In order to gain insight into the available data sets, we reviewed all existing corpora annotated with negation, based on several criteria of analysis that we present in this section. To the best of our knowledge, there are corpora annotated for English, Spanish, Swedish, Chinese, Dutch, German, and Italian. For each corpus we collected information about the source of the texts, and 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. Next, we define the criteria that have been applied to review the corpora:

- **Language**: The language(s) of the texts included in the corpus. This characteristic should always be specified in the description of any corpus, as it conditions its use.

3 https://catalog.ldc.upenn.edu/.
4 http://catalog.elra.info/en-us/.
5 http://lremap.elra.info/.
6 http://www.meta-share.org/.
7 http://linguistic.linkeddata.es/retele-share/sparql-editor/.

- **Domain**: Field to which the texts belong. Although cross-domain methodologies are being used for many tasks (Li et al. 2012; Szarvas et al. 2012; Bollegala, Mu, and Goulermas 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 not the number of sentences but 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, and so forth.
- **Elements with negation**: Total number of elements that have been annotated with negation. As has been mentioned before, the number of annotated sentences is not important, but rather the information annotated in them. 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. There are different types of negation depending on the type of negation cue used (Jiménez-Zafra et al. 2018b):
  - **Syntactic negation**, if a syntactically independent negation marker 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.
- **Negation components**: Components of negation that have been annotated:
  - **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., *I have **never** been to Los Angeles*), pronouns (e.g., *His decisions have **nothing***

to do with me), verbs (e.g., *The magazine* ***desisted*** *from published false stories about the celebrity*), and words with negative prefixes (e.g., *What you've done is* ***illegal***). They may consist of a single token (e.g., *I do* ***not*** *like the food of this restaurant*), a sequence of two or more contiguous tokens (e.g., *He has* ***not even*** *tried it*), or two or more non-contiguous tokens (e.g., *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 and Moldovan 2011b). For example, consider the sentence (a) *My children do not like meat* and its positive counterpart (b) *My children like meat*. In order for (b) to be true the following conditions must be satisfied: (i) somebody likes something, (ii) my children are the ones who like it, and (iii) meat is what is liked. The falsity of any of them would make (a) true. Therefore, all these elements are the scope of negation: *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, Ohta, and Tsujii 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 *"Technical assistance did not arrive on time,"* the event is the verbal form *"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 and 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 *"I'm not going to the concert with you,"* the focus is *"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 and Blanco 2012).

Example (1) shows a sentence with the last four elements, which have been explained above. The negation cue appears in **bold**, the event in *italics*, the focus <u>underlined</u>, 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[8] 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 (2) we show a sentence in which the scope of negation is the whole sentence and, in Example (3), a sentence with two coordinated structures with independent negation cues and predicates in which a scope is annotated for each coordinated negation marker.

1. Es una persona que [**no** *tiene* límites], aunque a veces puede controlarse.
   He is a person who has no limits, although sometimes he can control himself.
2. [El objetivo de la cámara **nunca** *ha funcionado* bien].
   The camera lens has never worked well.
3. [**No** *soy* alta] aunque [**tampoco** *soy* un pitufo].
   I'm not tall, but I'm not a smurf either.

In this section we have presented the aspects that we have described for each corpus. In Sections 4, 5, and 6, we present the existing corpora annotated with negation grouped by language. In Section 9 we provide an analysis of all the factors and we summarize them in different tables than can be found in Appendix A.

4. English Corpora

As we already indicated, our analysis 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, Zhang, and Kordoni 2012). The Groningen Meaning Bank[9] 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, Clark, and Bos 2007) and then manually corrected. The DeepBank corpus[10] contains rich syntactic and semantic annotations for the 25 Wall Street Journal sections included in the Penn Treebank (Taylor, Marcus, and Santorini 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.

4.1 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

8 There are authors that do not include the negation cue within the scope.
9 The Groningen Meaning Bank is available at `http://gmb.let.rug.nl`.
10 DeepBank is available at `http://moin.delph-in.net/DeepBank`.

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 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* or *unable to*, was annotated with the predicate NOT:

4. Abundance of actin is not affected by calreticulin expression. (See Figure 1.)
   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 1
Annotated example from the BioInfer corpus (not affected by).

5. N-WASP mutant unable to interact with profilin. (See Figure 2.)
   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
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/.

4.2 Genia Event

The **Genia Event** corpus (Kim, Ohta, and Tsujii 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, Tateisi, and Kim 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, using the label *non-exists* or *exists*, respectively. The three types of negation were considered, but linguistic cues were not annotated.

6. This pathway involves the Rac1 and Cdc42 GTPases, two enzymes that are not required for NF-kappaB activation by IL-1beta in epithelial cells. (See Figure 3.)

```
<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 3
Annotated example from the Genia Event corpus.

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.[11] 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.

4.3 BioScope

The **BioScope** corpus (Vincze et al. 2008) is one of the largest corpora and is the first in which negation and speculation markers 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 (Collier et al. 1999). In total, it has 20,924 sentences, out of which 2,720 contains negations.

Negation is understood 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). The largest syntactic unit possible

11 `http://www.nactem.ac.uk/meta-knowledge/Annotation_Guidelines.pdf`.

should be annotated as scope. Moreover, negation cues were also included within the scope.

7. PMA treatment, and not retinoic acid treatment of the U937 cells, acts in inducing NF-KB expression in the nuclei. (See Figure 4.)

```
<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 4
Annotated example from the BioScope corpus.

The corpus was annotated by two independent linguist annotators and a chief linguist following annotation guidelines.[12] 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 used in the CoNLL-2010 Shared Task: *Learning to detect hedges and their scope in natural language text* (Farkas et al. 2010).

4.4 Product Review Corpus

In 2010, the **Product Review corpus** was presented (Councill, McDonald, and Velikovich 2010b). 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 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:

8. I am a soft seller, If you don't want or need the services offered that's cool with me. (See Figure 5.)

12 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).

```
<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 5
Annotated example from the Product Review corpus.

4.5 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 and 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, Gildea, and Kingsbury 2005), and performed annotations of negation focus. They reduced the task to selecting the semantic role most likely to be the focus.

Fifty percent 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 and Blanco 2012). 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 6 shows the annotations for Example (4.5). 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.

9. Marketers believe most Americans won't make the convenience trade-off. (See Figure 6.)

4.6 ConanDoyle-neg

The **ConanDoyle-neg** (Morante and 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

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 6
Annotated example from the PropBank Focus (PB-FOC) corpus.

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 7
Annotated example from the ConanDoyle-neg corpus.

annotation guidelines[13] 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.

The ConanDoyle-neg corpus was prepared with the aim of using it at the *SEM 2012 Shared Task[14] (Morante and Blanco 2012), 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 7 it can be seen how Example (4.6) 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 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.

10. After his habit he said nothing, and after mine I asked no questions. (See Figure 7.)

No license is needed to download the corpus.

13 The annotation guidelines are described in Morante, Schrauwen, and Daelemans (2011).
14 `www.clips.ua.ac.be/sem2012-st-neg/`.

4.7 SFU Review$_{EN}$

Konstantinova et al. (2012) annotated the **SFU Review**$_{EN}$ corpus (Taboada, Anthony, and Voll 2006) with information about negation and speculation. This corpus is composed of 400 reviews extracted from the Web site *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 17,263 sentences, 18% contain negation cues (3,017 sentences). In this corpus syntactic negation was annotated, but not lexical nor morphological negation.

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 a value 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, De Sousa, and Sheila (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. Figure 8 shows how Example (4.7) is annotated in the corpus:

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

```
<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 8
Annotated example from the SFU Review$_{EN}$ corpus.

4.8 NEG-DrugDDI

In the biomedical domain, the DrugDDI 2011 corpus (Segura Bedmar, Martinez, and de Pablo Sánchez 2011) was also tagged with negation cues and their scopes, producing the **NEG-DrugDDI corpus** (Bokharaeian, Díaz Esteban, and Ballesteros Martínez 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. Figure 9 expands Example (12), a corpus sentence containing two negations.

12. 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 9
Annotated example from the NEG-Drug DDI corpus.

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 three different sets that were assigned to three 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 first three 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.

The NEG-DrugDDI corpus is in XML format and can be downloaded at `http://nil.fdi.ucm.es/sites/default/files/NegDrugDDI.zip`.

4.9 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.

This corpus is in XML format and is freely available at http://nil.fdi.ucm.es/sites/default/files/NegDDI_DrugBank.zip. In Figure 10, we show the annotations from Example (13). It can be seen that the annotation scheme is the same as the one used in the corpus NEG-DrugDDI.

13. 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 10
Annotated example from the NEGDDI-DrugBank corpus.

4.10 Deep Tutor Negation

The **Deep Tutor Negation corpus (DT-Neg)** (Banjade and 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 marker. In this corpus, syntactic and lexical negation were taken into account but not morphological negation.

In relation to the annotation process, the corpus was first automatically annotated based on a list of cue words that the authors compiled from different research reports (Morante, Schrauwen, and Daelemans 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 five graduate students and researchers following an annotation manual that was inspired by the guidelines of Morante, Schrauwen, and Daelemans (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. Figure 11 is an example of an annotated response.

14. They will not hit the water at the same time. (See Figure 11.)

```
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 11
Annotated example from the Deep Tutor Negation corpus.

4.11 SOCC

Finally, the last English corpus we are aware of is the **SFU Opinion and Comments Corpus (SOCC)** (Kolhatkar et al. 2019) 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.[15]

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 toward 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.

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.[16] 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.

15 https://github.com/sfu-discourse-lab/SOCC.
16 https://github.com/sfu-discourse-lab/SOCC/tree/master/guidelines.

Regarding the agreement, two annotators performed the annotation, a graduate student in 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-NonCommercial-ShareAlike 4.0 International License. Next, we show an annotated example in Figure 12.

15. 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 _
2-8 228-232 this _
2-9 233-235 is _
2-10 236-240 just _
2-11 241-248 another _
2-12 249-255 murder _
2-13 256-261 where _
2-14 262-269 someone _
2-15 270-273 was _
2-16 274-276 at _
2-17 277-280 the _
2-18 281-286 WRONG _
2-19 287-292 PLACE _
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 12
Annotated example from the SOCC corpus.

5. Spanish Corpora

In this section 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.

5.1 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 and 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).

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 and 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. In Figure 13, we present an example extracted from the corpus in which negation affects the complete sentence.

16. 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 13
Annotated example from the UAM Spanish Treebank corpus.

5.2 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 and 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[17] following a procedure of some conditions that include a confidentiality agreement, and its format is not specified.

5.3 SFU Review$_{SP}$-NEG

The **SFU Review$_{SP}$-NEG**[18] (Jiménez-Zafra et al. 2018b) 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, Anthony, and Voll 2006) and it could be considered the counterpart of the SFU Review Corpus with negation and speculation annotations[19] (Konstantinova et al. 2012).

The Spanish SFU Review corpus consists of 400 reviews extracted from the Web site *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 (Padro and 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.

17 `http://ixa.si.ehu.eus/extrecm`.
18 First Online: 22 May 2017 `https://doi.org/10.1007/s10579-017-9391-x`.
19 `https://www.sfu.ca/~mtaboada/SFU_Review_Corpus.html`.

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.[20] 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 (Marti 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. In Figure 14, we present an example of a sentence containing negation annotated in this corpus:

17. El 307 es muy bonito, pero no os lo recomiendo.
    The 307 is very nice, but I don't recommend it.

```
<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 14
Annotated example from the SFU Review$_{SP}$-NEG corpus.

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

18. 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 15
Annotated example from the SFU ReviewSP-NEG corpus **for negation cue detection** in CoNLL format.

The annotations of this corpus were used in *NEGES 2018: Workshop on Negation in Spanish* (Jiménez-Zafra et al. 2019) for Task 2: "Negation cues detection" (Jiménez-Zafra et al. 2018). The corpus was converted to CoNLL format (Farkas et al. 2010) as in the *SEM 2012 Shared Task (Morante and Blanco 2012). This format of the corpus can be downloaded from the Web site of the workshop `http://www.sepln.org/workshops/neges/index.php?lang=en` or by sending an email to the organizers. In Figure 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 columns contain "-", because the proposed task was only negation cue detection. Figure 15 shows an annotated example.

5.4 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). A total of 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, namely, 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 ethics issues.

5.5 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, Deshpande, and Nadkarni 2001; Szarvas et al. 2008; Morante and 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. Figure 16 is an example of the annotation of a sentence in this corpus is presented:

19. 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 16
Annotated example from the IULA Spanish Clinical Record corpus.

6. Other Corpora

Some corpora have been created for languages other than Spanish and English. We present them in this section.

6.1 Swedish Uncertainty, Speculation, and Negation Corpus

Dalianis and Velupillai (2010) annotated a subset of the **Stockholm Electronic Patient Record corpus** (Dalianis, Hassel, and Velupillai 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, 6,740 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 (Figure 17). 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 6,996 expressions, out of which 1,008 were negative keywords.

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

20. 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 17
Annotated example from the Stockholm Electronic Patient Record corpus.

6.2 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 3,626 medical terms from general practitioners, 12% were annotated as negated (435).

- Out of a total of 2,748 medical terms from specialists' letters, 15% were annotated as negated (412).
- Out of a total of 3,684 medical terms from radiology reports, 16% were annotated as negated (589).
- Out of a total of 2,830 medical terms from discharge letters, 13% were annotated as negated (368).

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

6.3 Japanese Negation Corpus

Matsuyoshi, Otsuki, and Fukumoto (2014) proposed an annotation scheme for the focus of negation in Japanese and annotated a corpus of reviews from "Rakuten Travel: User review data"[21] and the newspaper subcorpus of the "Balanced Corpus of Contemporary Written Japanese (BCCWJ)"[22] 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[23] 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 (6.3) 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://cl.cs.yamanashi.ac.jp/nldata/negation/`, although it is not available yet.[24]

21. 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.

6.4 Chinese Negation and Speculation Corpus

Zou, Zhou, and Zhu (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

21 `http://rit.rakuten.co.jp/rdr/index_en.html`.
22 `http://www.ninjal.ac.jp/english/products/bccwj/`.
23 `http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html`.
24 Accessed March 19, 2019.

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 a value of 0.96, 0.96, and 0.93 for negation cue detection and 0.90, 0.91, and 0.88 for scope identification, scientific literature, and 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/`. In Figure 18 we show an annotation example of a hotel review sentence.

22. 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 18
Annotated example from the CNeSp corpus.

6.5 German Negation and Speculation Corpus

The **German negation and speculation corpus** (Cotik et al. 2016a) 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.

6.6 Italian Negation Corpus

Altuna, Minard, and Speranza (2017) proposed an annotation framework for negation in Italian based on the guidelines proposed by Morante, Schrauwen, and Daelemans (2011) and Blanco and 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") (Figure 19).

23. 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 19
Annotated example from the Fact-Ita Bank Negation corpus.

The corpus is composed of 71 documents from the Fact-Ita Bank corpus (Minard, Marchetti, and Speranza 2014 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, Speranza, and Caselli 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 a value of 0.98, 0.67, and 0.58, respectively.

The corpus is in XML format and it can be downloaded at `https://hlt-nlp.fbk.eu/technologies/fact-ita-bank` under a Creative Commons Attribution-NonCommercial 4.0 International License. It should be mentioned that only news annotations are available. Tweets are not available because they are from another corpus that has copyright. In Figure 19, a negation sentence of the corpus is shown.

7. Negation Processing

Some of the corpora described in the previous sections have been used to develop negation processing systems. The tasks that are performed by the systems are directly related to how negation has been modeled in the annotations. Four tasks are usually performed in relation to processing negation:

Negation cue detection that aims at finding the words that express negation.

Scope identification that consists in determining which parts of the sentence are affected by the negation cues. The task was introduced in 2008, when the BioScope corpus was released as a machine learning sequence labeling task (Morante, Liekens, and Daelemans 2008).

Negated event recognition that focuses on detecting whether events are affected by the negation cues; this task was motivated by the release of biomedical corpora annotated with negated events, such as BioInfer and Genia Event.

Focus detection consisting of finding the part of the scope that is most prominently negated. This task was introduced by Blanco and Moldovan (2011b), who argued that the *scope* and *focus* of negation are crucial for a correct interpretation of negated statements. The authors released the PropBank Focus corpus, on which all focus detection systems have been trained. The corpus was used in the first edition of the *SEM Shared Task, which was dedicated to resolving the scope (Task 1) and focus (Task 2) of negation (Morante and Blanco 2012). Both rule-based (Rosenberg and Bergler 2012) and machine learning approaches (Blanco and Moldovan 2013; Zou, Zhu, and Guodong 2015) have been applied to solve this task.

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 cue detection.

The interest in processing negation originated from the need to extract information from clinical records (Chapman et al. 2001a; Mutalik, Deshpande, and Nadkarni 2001; Goldin and Chapman 2003). Despite the fact that many studies have focused on negation in clinical texts, the problem is not yet solved (Wu et al. 2014), due to several reasons, among which is the lack of consistent annotation guidelines.

Three main types of approaches have been applied to processing negation: (i) rule-based systems have been developed based on lists of negations and stop words (Mitchell et al. 2004; Harkema et al. 2009; Mykowiecka, Marciniak, and Kupść 2009; Uzuner, Zhang, and Sibanda 2009; Sohn, Wu, and Chute 2012). The first system was the

NegEx algorithm (Chapman et al. 2001a), which was then improved resulting in systems such as ConText (Harkema et al. 2009), DEEPEN (Mehrabi et al. 2015), and NegMiner (Elazhary 2017); (ii) machine learning techniques (Agarwal and Yu 2010; Li et al. 2010; Cruz Díaz et al. 2012; Velldal et al. 2012; Cotik et al. 2016b; Li and Lu 2018); and (iii) deep learning approaches (Fancellu, Lopez, and Webber 2016; Qian et al. 2016; Ren, Fei, and Peng 2018; Lazib et al. 2018). Although the interest in processing negation has only increased, negation resolvers are not yet a standard component of the natural language processing pipeline. Recently, a tool for detecting negation cues and scopes in English natural language texts has been released (Enger, Velldal, and Øvrelid 2017).

Later on, with the developments in opinion mining, negation was studied as a marker of polarity change (Das and Chen 2001; Wilson, Wiebe, and Hoffmann 2005; Polanyi and Zaenen 2006; Taboada et al. 2011; Jiménez-Zafra et al. 2017) and was incorporated in sentiment analysis systems. Some systems use rules to detect negation, without evaluating their impact (Das and Chen 2001; Polanyi and Zaenen 2006; Kennedy and Inkpen 2006; Jia, Yu, and Meng 2009), whereas other systems use a lexicon of negation cues and predict the scope with machine learning algorithms (Councill, McDonald, and Velikovich 2010a; Lapponi, Read, and 2012; Cruz, Taboada, and Mitkov 2016b). Most systems are tested on the SFU Review corpus.

Several shared tasks have addressed negation processing for English: the BioNLP'09 Shared Task 3 (Kim et al. 2009), the i2b2 NLP Challenge (Uzuner et al. 2011), the *SEM 2012 Shared Task (Morante and Blanco 2012), and the ShARe/CLEF eHealth Evaluation Lab 2014 Task 2 (Mowery et al. 2014).

Although most of the work on processing negation focused on English texts, recently, negation in Spanish texts has attracted the attention of researchers. Costumero et al. (2014), Stricker, Iacobacci, and Cotik (2015), and Cotik et al. (2016b) develop systems for the identification of negation in clinical texts by adapting the NegEx algorithm (Chapman et al. 2001b). Regarding product reviews, there are some works that treat negation as a subtask of sentiment analysis (Taboada et al. 2011; Vilares, Alonso, and Gómez-Rodríguez 2013, 2015; Jiménez-Zafra et al. 2015; Amores, Arco, and Barrera 2016; Miranda, Guzmán, and Salcedo 2016; Jiménez-Zafra et al. 2019). The first systems that detect negation cues were developed in the framework of the NEGES workshop 2018 (Jiménez-Zafra et al. 2019) and were trained on the SFU Corpus (Jiménez-Zafra et al. 2018). Fabregat, Martínez-Romo, and Araujo (2018) applied a deep learning model based on the combination of some dense neural networks and one Bidirectional Long Short-Term Memory network and Loharja, Padró, and Turmo (2018) used a CRF model. Additionally there is also work for other languages such as Swedish (Skeppstedt 2011), German (Cotik et al. 2016a), or Chinese (Kang et al. 2017).

8. Analysis

Negation is an important phenomenon to deal with in NLP tasks if we want to develop accurate systems. Work on processing negation has started relatively late as compared to work on processing other linguistic phenomena, and there are no publicly available off-the-shelf tools for detecting negations that can be easily incorporated into applications. In this overview, the corpora annotated with negation so far are presented with the aim of promoting the development of such tools. For the development of a negation processing system it is not only important that corpora exist, but also that they be publicly available, well documented, and have annotations of quality. Moreover, to train robust machine learning systems it is necessary to have large enough data covering all possible cases of the phenomenon under study. Therefore, in this section we perform an analysis of the features of the corpora we have described and in the next section we discuss the possibility of merging the existing negation corpora in order to create a larger training corpus. In Appendix A, the information analyzed is summarized in Tables 8, 9, and 10.

8.1 Language and Year of Publication

The years of publication of the corpora (Table 3, 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 five corpora have been compiled, three 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.

8.2 Domain

If we look at Tables 8–10 (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 it is 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. 2001b, page 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 (24) we can find a positive opinion that changes to negative if negation is present as in Example (25), or by contrast, in Example (26) there is a positive opinion in which negation is present whose meaning changes if it does not have negation as in Example (27).

24. The camera works well.
25. The camera does not work well.
26. I have not found a camera that works better.
27. 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.

8.3 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 seven non-available corpora, five 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 4 (Appendix A).

8.4 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 5 (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, whereas it appears more frequently in Spanish, Italian, Chinese, and German. 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%, whereas for Spanish it is 29.13%, for Italian 21.55%, and for Chinese 26.82%.[25] 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 with 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 Section 8.5.

8.5 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?

8.5.1 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.

25 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.

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 6 (Appendix A), we show the link or reference to the annotation guidelines of the different corpora.

As Table 6 (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.

8.5.2 Negation Elements. Another important aspect to be analyzed from the corpora is what elements of negation have been annotated. As mentioned in Section 3, 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 (see Section 3). 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 7 (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 (e.g., because only negation that 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 (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 7 (Appendix A), this information is represented using ✓ if the corpus contains annotations about the scope, event, and focus, respectively, or – otherwise.

8.5.3 Tokenization. The way in which each corpus was tokenized is also important and 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.

8.5.4 Annotation Scheme and Guidelines. In the previous sections an example of each corpus has been provided whenever possible. If we take a 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[26]:

- 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 (28)). On the contrary, in the IULA Spanish Clinical Record corpus the subject is included within the scope (Example (29)) only when it is located after the verb (Example 30), or when there is an unaccusative verb (Example (31)).

  28. 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.
  29. *MVC*$_{SUBJ}$ **sin** [ruidos sobreañadidos].
      NBS no additional sounds.
  30. Se **descarta** [*enolismo*$_{SUBJ}$].
      Oenolism discarded.
  31. [*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 (32)), whereas in the BioScope corpus it was included (Example (33)).

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

32. I **cannot** [believe that one of the suicide bombers was deported back to Belgium.]
33. 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 (34)), whereas in ConanDoyle-neg corpus the longest relevant scope of the negation cue was marked (Example (35)).

34. Long live ambitious filmmakers with **no** [talent]
35. [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.

9. Discussion

The perspective that we have taken in this article when analyzing the corpora annotated with negation is computational, because 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, analyzed in Section 8.1, and the domain, reviewed in Section 8.2. Next, we discuss the possibility of merging corpora according to each of these aspects.

On the one hand, it can 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 1 we show for which of these tasks each corpus could 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 the 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 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	BioInfer	Genia Event	PropBank Focus (PB-FOC)
	PropBank Focus (PB-FOC)	BioScope	ConanDoyle-neg	Deep Tutor Negation
	ConanDoyle-neg	ConanDoyle-neg		SOCC
	SFU Review$_{EN}$	SFU Review$_{EN}$		
	NEG-DrugDDI	NEG-DrugDDI		
	NegDDI-DrugBank	NegDDI-DrugBank		
	Deep Tutor Negation	Deep Tutor Negation		
	SOCC	SOCC		
Spanish	UAM Spanish Treebank	UAM Spanish Treebank	IxaMed-GS	
	SFU Review$_{SP}$-NEG	SFU Review$_{SP}$-NEG	SFU Review$_{SP}$-NEG	
	UHU-HUVR	UHU-HUVR	UHU-HUVR	
	IULA Spanish Clinical Record	IULA Spanish Clinical Record		
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 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 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. As we have indicated in Section 8.3, 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

Table 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	

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 in Section 8.5:

1. As we showed in Section 8.5.1, 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. As has been mentioned in Section 8.5.2, 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. As indicated in Section 8.5.3, 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. As we have shown in Section 8.5.4, 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, as indicated in Sub-subsection 8.5.4, 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.

As our analysis shows, the main problem 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, which 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 1).

For **negation cue detection**, we could merge the corpora that have been completely annotated for the same type of negation (Table 7). 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 merged for the identification of lexical cues. However, we cannot 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 obtain negation cues information and convert that 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 (36), corresponding 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 a 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 (37)), with non-contiguous cues (Example (37)), or with multi-words expressions (Example (38)). 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, contiguous, and non-contiguous negation cues. The annotation of these structures needs to be unified.

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

37. El final del libro **no** te aporta **nada**, **no** aade **nada** nuevo, no crees?
The end of the book doesn't give you anything, it doesn't add anything new, didn't you?

38. 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 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 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 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.

10. Conclusions

In this article, we have reviewed the existing corpora annotated with negation information in several languages. Processing negation is a very important task in NLP because negation is a linguistic phenomenon that can change the truth value of a proposition, and so it is crucial in some tasks such as sentiment analysis, information extraction, summarization, machine translation, and question answering. Most corpora have been annotated for English, but it is also necessary to focus on other languages whose presence on the Internet is growing, such as Chinese or Spanish.

We have conducted an exhaustive search of corpora annotated with negation, finding corpora for the following languages: English, Spanish, Swedish, Dutch, Japanese, Chinese, German, and Italian. We have described the main features of the corpora based on the following criteria: the language, year of publication, domain, the availability, size, types of negation taken into account (syntactic and/or lexical and/or morphological), negation elements annotated (cue and/or scope and/or negated event and/or focus) and the way in which each corpus was tokenized, the annotation guidelines, and annotation scheme used. In addition, we have included an appendix with tables summarizing all this information in order to facilitate analysis.

In sum, our analysis demonstrates that the language and year of publication of the corpora show that interest in the annotation of negation started in 2007 with English texts followed by Swedish in 2010, whereas for the other languages (Spanish, Dutch, Chinese, German and Italian) it is a task of recent interest. Most of the corpora have been documented 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. Another important fact that we have analyzed is the availability of the corpora. Most of them are publicly available and most of the non-available corpora contain clinical reports, with legal and ethical issues probably affecting their status. The length of the corpora shows that existing corpora are not very large, which hinders the development of machine learning systems, since the frequency of negations is low. Finally, in relation to the annotation guidelines, most of the annotators define guidelines, but some of them are not complete and others are not available. In addition, we found differences in the annotation schemes used, and, most importantly, in the annotation guidelines: the way in which each corpus was 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 were tokenized— this being essential for processing negation systems because the identification of negated elements (cue, scope, event, and focus) is carried out at token level.

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 labeling 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.

Appendix A: Comparative Tables

Table 3
Language and year of publication of the corpora.

Corpus	Language	Year
BioInfer (Pyysalo et al. 2007)	English	2007
Genia Event (Kim, Ohta, and Tsujii 2008)	English	2008
BioScope (Vincze et al. 2008)	English	2008
Product Review (Councill, McDonald, and Velikovich 2010b)	English	2010
Stockholm Electronic Patient Record (Dalianis and Velupillai 2010)	Swedish	2010
PropBank Focus (PB-FOC) (Blanco and Moldovan 2011a)	English	2011
ConanDoyle-neg (Morante and Daelemans 2012)	English	2012
SFU Review$_{EN}$ (Konstantinova et al. 2012)	English	2012
NEG-DrugDDI (Bokharaeian, Díaz Esteban, and Ballesteros Martínez 2013)	English	2013
UAM Spanish Treebank (Sandoval and 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, Otsuki, and Fukumoto 2014)	Japanese	2014
IxaMed-GS (Oronoz et al. 2015)	Spanish	2015
Deep Tutor Negation (Banjade and Rus 2016)	English	2016
CNeSp (Zou, Zhou, and Zhu 2016)	Chinese	2016
German negation and speculation (Cotik et al. 2016a)	German	2016
Fact-Ita Bank Negation (Altuna, Minard, and Speranza 2017)	Italian	2016
SFU Review$_{SP}$-NEG (Jiménez-Zafra et al. 2018b)	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. 2019)	English	2018

Table 4
Availability of the corpora.

Corpus	Links to the data
BioInfer (Pyysalo et al. 2007)	http://mars.cs.utu.fi/BioInfer/
Genia Event (Kim, Ohta, and Tsujii 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, McDonald, and Velikovich 2010b)	–
Stockholm Electronic Patient Record (Dalianis and Velupillai 2010)	–
PropBank Focus (PB-FOC) (Blanco and Moldovan 2011a)	http://www.clips.ua.ac.be/sem2012-st-neg/data.html
ConanDoyle-neg (Morante and 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, Díaz Esteban, and Ballesteros Martínez 2013)	http://nil.fdi.ucm.es/sites/default/files/NegDrugDDI.zip
UAM Spanish Treebank (Sandoval and 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, Otsuki, and Fukumoto 2014)	http://cl.cs.yamanashi.ac.jp/nldata/negation/
IxaMed-GS (Oronoz et al. 2015)	–
Deep Tutor Negation (Banjade and Rus 2016)	http://deeptutor.memphis.edu/resources.htm
CNeSp (Zou, Zhou, and Zhu 2016)	http://nlp.suda.edu.cn/corpus/CNeSp/
German negation and speculation (Cotik et al. 2016a)	–
Fact-Ita Bank Negation (Altuna, Minard, and Speranza 2017)	https://hlt-nlp.fbk.eu/technologies/fact-ita-bank
SFU Review$_{SP}$-NEG (Jiménez-Zafra et al. 2018b)	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. 2019)	https://researchdata.sfu.ca/islandora/object/islandora%3A9109

Note: Link to the Review and Japanese corpus is currently not available (Accessed March 19, 2019). However, authors say that they plan to freely distribute it in the provided link.

Table 5
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, Ohta, and Tsujii 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, McDonald, and Velikovich 2010b)	English	Reviews	2,111	2,111 sentences	679 sentences (32.16%)
Stockholm Electronic Patient Record (Dalianis and Velupillai 2010)	Swedish	Clinical reports	6,740	6,966 expressions	1,008 expressions (10.67%)
PropBank Focus (PB-FOC) (Blanco and Moldovan 2011a)	English	Journal stories	3,779	NA	3,993 verbal negations
ConanDoyle-neg (Morante and 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, Díaz Esteban, and Ballesteros Martínez 2013)	English	Biomedical	5,806	5,806 sentences	1,399 sentences (24.10%)
UAM Spanish Treebank (Sandoval and 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, Otsuki, and Fukumoto 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 and Rus 2016)	English	Tutorial dialogues	NA	27,785 student responses	2,603 student responses (9.37%)
CNeSp (Zou, Zhou, and Zhu 2016)	Chinese	Scientific literature, product reviews and financial articles	16,841	16,841 sentences	4,517 sentences (26.82%)
German negation and speculation (Cotik et al. 2016a)	German	Clinical reports	NA	1,114 medical terms	443 medical terms (39.77%)
Fact-Ita Bank Negation (Altuna, Minard, and Speranza 2017)	Italian	News articles	1,290	1,290 sentences	278 sentences (21.55%)
SFU Review$_{SP}$-NEG (Jiménez-Zafra et al. 2018b)	Spanish	Reviews	9,455	9,455 sentences	3,022 sentences (31.97%)
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. 2019)	English	Opinion articles	3,612	3,612 sentences	1,130 sentences (31.28%)

Table 6
Annotation guidelines.

Corpus	Annotation guidelines
BioInfer (Pyysalo et al. 2007)	`http://tucs.fi/publications/view/?pub_id=tGiPyBjHeSa07a`
Genia Event (Kim, Ohta, and Tsujii 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, McDonald, and Velikovich 2010b)	(Councill, McDonald, and Velikovich 2010b)
Stockholm Electronic Patient Record (Dalianis and Velupillai 2010)	—
PropBank Focus (PB-FOC) (Blanco and Moldovan 2011a)	(Blanco and Moldovan 2011a)
ConanDoyle-neg (Morante and Daelemans 2012)	(Morante, Schrauwen, and Daelemans 2011)
SFU Review$_{EN}$ (Konstantinova et al. 2012)	(Konstantinova, De Sousa, and Sheila 2011)
NEG-DrugDDI Bokharaeian, Díaz Esteban, and Ballesteros Martínez 2013	—
UAM Spanish Treebank (Sandoval and Salazar 2013)	(Sandoval and Salazar 2013)
NegDDI-DrugBank (Bokharaeian et al. 2014)	—
EMC Dutch (Afzal et al. 2014)	—
Review and Newspaper Japanese (Matsuyoshi, Otsuki, and Fukumoto 2014)	(Matsuyoshi, Otsuki, and Fukumoto 2014)
IxaMed-GS (Oronoz et al. 2015)	—
Deep Tutor Negation (Banjade and Rus 2016)	—
CNeSp (Zou, Zhou, and Zhu 2016)	(Zou, Zhou, and Zhu 2016)
German negation and speculation (Cotik et al. 2016a)	—
Fact-Ita Bank Negation (Altuna, Minard, and Speranza 2017)	(Altuna, Minard, and Speranza 2017)
SFU Review$_{SP}$-NEG (Jiménez-Zafra et al. 2018b)	(Marti et al. 2016; Jiménez-Zafra et al. 2018b)
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. 2019)	`https://github.com/sfu-discourse-lab/SOCC/tree/master/guidelines`

Table 7
Negation elements (NA: Non-Available, −: Absent, ✓: Present).

Corpus	Negation	Cue	Scope	Event	Focus
BioInfer (Pyysalo et al. 2007)	PS, PM, PL	–	✓	–	–
Genia Event (Kim, Ohta, and Tsujii 2008)	PS, PM, PL	–	–	✓	–
BioScope (Vincze et al. 2008)	CS, CL	✓	✓	–	–
Product Review (Councill, McDonald, and Velikovich 2010b)	CS	–	✓	–	–
Stockholm Electronic Patient Record (Dalianis and Velupillai 2010)	CS	✓	–	–	–
PropBank Focus (PB-FOC) (Blanco and Moldovan 2011a)	PS	✓	–	–	✓
ConanDoyle-neg (Morante and Daelemans 2012)	CS	✓	✓	✓	–
SFU Review$_{EN}$ (Konstantinova et al. 2012)	CS	✓	✓	–	–
NEG-DrugDDI (Bokharaeian, Díaz Esteban, and Ballesteros Martínez 2013)	CS, CM, CL	✓	✓	–	–
UAM Spanish Treebank (Sandoval and Salazar 2013)	CS	✓	✓	–	–
NegDDI-DrugBank (Bokharaeian et al. 2014)	CS, CM, CL	✓	✓	–	–
EMC Dutch (Afzal et al. 2014)	NA	–	–	✓	–
Review and Newspaper Japanese (Matsuyoshi, Otsuki, and Fukumoto 2014)	CS,CM, CL	✓	–	–	✓
IxaMed-GS (Oronoz et al. 2015)	PS, PM, PL	–	–	✓	–
Deep Tutor Negation (Banjade and Rus 2016)	CS, CL	✓	✓	–	✓
CNeSp (Zou, Zhou, and Zhu 2016)	NA	✓	✓	–	–
German negation and speculation (Cotik et al. 2016a)	NA	–	–	✓	–
Fact-Ita Bank Negation (Altuna, Minard, and Speranza 2017)	CS	✓	✓	✓	–
SFU Review$_{SP}$-NEG (Jiménez-Zafra et al. 2018b)	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. 2019)	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 8
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, Ohta, and Tsujii 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, McDonald, and Velikovich 2010b)	NA	English	Reviews	2,111	2,111 sentences	679 sentences (32.16%)	CS	—	✓	—	—	✓	NA	YES
PropBank Focus (PB-FOC) (Blanco and Moldovan 2011a)	✓	English	Journal stories	3,779	NA	3,993 verbal negations	PS	✓	—	—	✓	✓	TXT	YES
ConanDoyle-neg (Morante and 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, Díaz Esteban, and Ballesteros Martínez 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 and Rus 2016)	✓	English	Tutorial dialogues	NA	27,785 student responses	2,603 student responses (9.37%)	CS, CL	✓	✓	—	✓	✓	TXT	YES
SOCC (Kolhatkar et al. 2019)	✓	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 9
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 and 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 (Jimenez-Zafra et al. 2018b)	✓	Spanish	Movies, books, product reviews	9,455	9,455 sentences	3,022 sentences (31.97%)	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 10
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 and 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, Otsuki, and Fukumoto 2014)	NA	Japanese	Reviews and news articles	10,760	10,760 sentences	1,785 sentences (16.59%)	CS, CM, CL	✓	-	-	✓	✓	NA	YES
CNeSp (Zou, Zhou, and Zhu 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 et al. 2016a)	NA	German	Clinical reports	NA	1,114 medical terms	443 medical terms (39.77%)	NA	-	-	✓	-	✓	NA	YES
Facti-Ita Bank Negation (Cotik et al. 2016a)	✓	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.

Acknowledgments

This work has been partially supported by a grant from the Ministerio de Educación Cultura y Deporte (MECD - scholarship FPU014/00983), LIVING-LANG project (RTI2018-094653-B-C21), Fondo Europeo de Desarrollo Regional (FEDER), and REDES project (TIN2015-65136-C2-1-R) from the Spanish Government. R.M. was supported by the Netherlands Organization for Scientific Research (NWO) via the Spinoza-prize awarded to Piek Vossen (SPI 30-673, 2014-2019). We are thankful to the authors of the corpora who kindly answered our questions.

References

Afzal, Zubair, Ewoud Pons, Ning Kang, Miriam C. J. M. Sturkenboom, Martijn J. Schuemie, and Jan A. Kors. 2014. ContextD: An algorithm to identify contextual properties of medical terms in a Dutch clinical corpus. *BMC Bioinformatics*, 15(1):1–12.

Agarwal, Shashank and Hong Yu. 2010. Biomedical negation scope detection with conditional random fields. *Journal of the American Medical Informatics Association*, 17(6):696–701.

Altuna, Begoña, Anne-Lyse Minard, and Manuela Speranza. 2017. The scope and focus of negation: A complete annotation framework for Italian. In *Proceedings of the Workshop Computational Semantics Beyond Events and Roles*, pages 34–42, Valencia.

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

Ananiadou, Sophia and John McNaught. 2006. *Text Mining for Biology and Biomedicine*. Artech House London.

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

Ballesteros, Miguel, Virginia Francisco, Alberto Díaz, Jesús Herrera, and Pablo Gervás. 2012. Inferring the scope of negation in biomedical documents. *Computational Linguistics and Intelligent Text Processing*, pages 363–375.

Banjade, Rajendra and Vasile Rus. 2016. DT-Neg: Tutorial dialogues annotated for negation scope and focus in context. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, pages 3768–3771. Paris.

Basile, Valerio, Johan Bos, Kilian Evang, and Noortje Venhuizen. 2012. Developing a large semantically annotated corpus. In *LREC 2012, Eighth International Conference on Language Resources and Evaluation*, pages 3196–3200.

Blanco, Eduardo and Dan Moldovan 2011a. Semantic representation of negation using focus detection. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 581–589, Portland, OR.

Blanco, Eduardo and Dan Moldovan. 2011b. Some issues on detecting negation from text. In *Proceedings of the Twenty-Fourth International FLAIRS Conference*, AAAI, pages 228–233, Florida.

Blanco, Eduardo and Dan Moldovan. 2013. Retrieving implicit positive meaning from negated statements. *Natural Language Engineering*, 20(4):501–535.

Bokharaeian, Behrouz, Alberto Diaz, Mariana Neves, and Virginia Francisco. 2014. Exploring negation annotations in the DrugDDI Corpus. In *Fourth Workshop on Building and Evaluating Resources for Health and Biomedical Text Processing (BIOTxtM 2014)*, pages 1–8, Citeseer.

Bokharaeian, Behrouz, Alberto Díaz Esteban, and Miguel Ballesteros Martínez. 2013. Extracting drug-drug interaction from text using negation features. Procesamiento del Lenguaje Natural, 51:49–56.

Bollegala, Danushka, Tingting Mu, and John Yannis Goulermas. 2016. Cross-domain sentiment classification using sentiment sensitive embeddings. *IEEE Transactions on Knowledge and Data Engineering*, 28(2):398–410.

Caselli, Tommaso, Valentina Bartalesi Lenzi, Rachele Sprugnoli, Emanuele Pianta, and Irina Prodanof. 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*, pages 143–151, Portland, OR.

de Castilho, Richard Eckart, Eva Mujdricza-Maydt, Seid Muhie Yimam, Silvana Hartmann, Iryna Gurevych, Anette Frank, and Chris Biemann. 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), pages 76–84, Osaka.

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

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

Collier, Nigel, Hyun Seok Park, Norihiro Ogata, Yuka Tateishi, Chikashi Nobata, Tomoko Ohta, Tateshi Sekimizu, Hisao Imai, Katsutoshi Ibushi, and Jun-ichi Tsujii. 1999. The Genia project: Corpus-based knowledge acquisition and information extraction from genome research papers. In *Proceedings of the Ninth Conference on European chapter of the Association for Computational Linguistics*, pages 271–272, Athens.

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

Cotik, Viviana, Roland Roller, Feiyu Xu, Hans Uszkoreit, Klemens Budde, and Danilo Schmidt 2016a. Negation detection in clinical reports written in German. In *Proceedings of the Fifth Workshop on Building and Evaluating Resources for Biomedical Text Mining (BioTxtM2016)*, pages 115–124, Osaka.

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

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

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

Cruz, Noa P., Maite Taboada, and Ruslan Mitkov. 2016a. 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, Noa P., Maite Taboada, and Ruslan Mitkov. 2016b. 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, Noa P., Manuel J. Maña López, Jacinto Mata Vázquez, and Victoria Pachón Álvarez. 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.

Cruz Díaz, Noa P., Roser Morante Vallejo, Manuel J. Maña López, Jacinto Mata Vázquez, and Carlos L. Parra Calderón. 2017. Annotating negation in Spanish clinical texts. *SemBEaR 2017*, pages 53–58.

Curran, James R, Stephen Clark, and Johan Bos. 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*, pages 33–36, Prague.

Dalianis, Hercules, Martin Hassel, and Sumithra Velupillai. 2009. The Stockholm EPR corpus–characteristics and some initial findings. *Women*, 219(906):1–7.

Dalianis, Hercules and Sumithra Velupillai. 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*, pages 3071–3075.

Das, Sanjiv and Mike Chen. 2001. Yahoo! for Amazon: Extracting market sentiment from stock message boards. In *Proceedings of the Asia Pacific Finance Association Annual Conference (APFA)*, volume 35, pages 1–16. Bangkok.

Elazhary, Hanan. 2017. Negminer: An automated tool for mining negations from electronic narrative medical documents. *International Journal of Intelligent Systems and Applications*, 9(4):14–22.

Enger, Martine, Erik Velldal, and Lilja Øvrelid. 2017. An open-source tool for negation detection: A maximum-margin approach. In *Proceedings of the Workshop Computational Semantics Beyond Events and Roles*, pages 64–69, Valencia.

Fabregat, Hermenegildo, Juan Martínez-Romo, and Lourdes Araujo. 2018. Deep learning approach for negation cues detection in Spanish at NEGES 2018. In *Proceedings of NEGES 2018: Workshop on Negation in Spanish, CEUR Workshop Proceedings*, volume 2174, pages 43–48, Seville.

Fancellu, Federico, Adam Lopez, and Bonnie Webber. 2016. Neural networks for negation scope detection. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 495–504, Berlin.

Fancellu, Federico, Siva Reddy, Adam Lopez, and Bonnie Webber. 2017. Universal dependencies to logical form with negation scope. In *Proceedings of the Workshop Computational Semantics Beyond Events and Roles*, pages 22–32, Valencia.

Farkas, Richárd, Veronika Vincze, György Móra, János Csirik, and György Szarvas. 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*, pages 1–12, Uppsala.

Flickinger, Dan. 2000. On building a more effcient grammar by exploiting types. *Natural Language Engineering*, 6(1):15–28.

Flickinger, Dan, Yi Zhang, and Valia Kordoni. 2012. DeepBank: A dynamically annotated treebank of the Wall Street Journal. In *Proceedings of the 11th International Workshop on Treebanks and Linguistic Theories*, pages 85–96, Lisbon.

Goldin, I. M. and W. W. Chapman. 2003. Learning to detect negation with 'Not' in medical texts. In *Proceedings of ACM-SIGIR 2003*, pages 1–7, Toronto.

Harkema, Henk, John N. Dowling, Tyler Thornblade, and Wendy W. Chapman. 2009. Context: An algorithm for determining negation, experiencer, and temporal status from clinical reports. *Journal of Biomedical Informatics*, 42(5):839–851.

Herrero Zazo, María, Isabel Segura Bedmar, Paloma Martínez, and Thierry Declerck 2013. The DDI corpus: An annotated corpus with pharmacological substances and drug–drug interactions. *Journal of Biomedical Informatics*, 46(5):914–920.

Hirschberg, Julia and Christopher D. Manning. 2015. Advances in natural language processing. *Science*, 349(6245):261–266.

Ide, Nancy. 2017 Introduction: The handbook of linguistic annotation. In *Handbook of Linguistic Annotation*. Springer, pages 1–18.

Ide, Nancy, Christiane Fellbaum, Collin Baker, and Rebecca Passonneau. 2010. The manually annotated sub-corpus: A community resource for and by the people. In *Proceedings of the ACL 2010 Conference Short Papers*, pages 68–73, Uppsala.

Jia, Lifeng, Clement Yu, and Weiyi Meng. 2009. The effect of negation on sentiment analysis and retrieval effectiveness. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, CIKM '09, pages 1827–1830, New York.

Jiménez-Zafra, Salud María, Noa P. Cruz-Díaz, Roser Morante, and María Teresa Martín-Valdivia. 2018. Tarea 2 del Taller NEGES 2018: Detección de Claves de Negación. In *Proceedings of NEGES 2018: Workshop on Negation in Spanish*, volume 2174, pages 35–41, Seville.

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

Jiménez-Zafra, Salud María, M. Teresa Martín-Valdivia, L. Alfonso Ureña-López, M. Antonia Martí, and Mariona Taulé. 2016. Problematic cases in the annotation of negation in Spanish. *ExProM 2016*, pages 42–48.

Jiménez-Zafra, Salud María, María Teresa Martín-Valdivia, Eugenio Martínez-Cámara, and L. Alfonso Ureña-López. 2019. 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.

Jiménez-Zafra, Salud María, Eugenio Martínez-Cámara, M. Teresa Martín-Valdivia, and M. Dolores Molina-González. 2015. Tratamiento de la negación en el análisis de opiniones en español. *Procesamiento del Lenguaje Natural*, 54:37–44.

Jiménez-Zafra, Salud María, Roser Morante, M. Teresa Martín-Valdivia, and L. Alfonso Ureña-López. 2018a. A review of Spanish corpora annotated with negation. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 915–924, Santa Fe, NM.

Jiménez-Zafra, Salud María , Mariona Taulé, M. Teresa Martín-Valdivia, L. Alfonso Ureña-López, and M. Antónia Martí.

2018b. 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.

Jimenez-Zafra, Salud Maria, M. Teresa Martin Valdivia, Eugenio Martinez Camara, and Luis Alfonso Urena-Lopez. 2017. Studying the scope of negation for Spanish sentiment analysis on Twitter. *IEEE Transactions on Affective Computing*, 10(1):129–141.

Kang, Tian, Shaodian Zhang, Nanfang Xu, Dong Wen, Xingting Zhang, and Jianbo Lei. 2017. Detecting negation and scope in Chinese clinical notes using character and word embedding. *Computer Methods and Programs in Biomedicine*, 140(C):53–59.

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

Kim, J.-D., Tomoko Ohta, Yuka Tateisi, and Jun'ichi Tsujii. 2003. Genia corpus—A semantically annotated corpus for bio-textmining. *Bioinformatics*, 19(suppl_1):i180–i182.

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

Kim, Jin-Dong, Tomoko Ohta, and Jun'ichi Tsujii. 2008. Corpus annotation for mining biomedical events from literature. *BMC Bioinformatics*, 9(1):1–25.

Kolhatkar, V., H. Wu, L. Cavasso, E. Francis, K. Shukla, and M. Taboada. 2019. The SFU opinion and comments corpus: A corpus for the analysis of online news comments. *Corpus Pragmatics*, pages 1–36.

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

Konstantinova, Natalia, Sheila C. M. De Sousa, and J. A. Sheila. 2011. Annotating negation and speculation: The case of the review domain. In *RANLP Student Research Workshop*, pages 139–144.

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

Lazib, Lydia, Bing Qin, Yanyan Zhao, Weinan Zhang, and Ting Liu. 2018. A syntactic path-based hybrid neural network for negation scope detection. *Frontiers of Computer Science*, pages 1–11.

Li, Fangtao, Sinno Jialin Pan, Ou Jin, Qiang Yang, and Xiaoyan Zhu. 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*, pages 410–419, Jeju Island.

Li, Hao and Wei Lu. 2018. Learning with structured representations for negation scope extraction. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 533–539, Melbourne.

Li, Junhui, Guodong Zhou, Hongling Wang, and Qiaoming Zhu. 2010. Learning the scope of negation via shallow semantic parsing. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 671–679, Beijing.

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

Liu, Bing. 2015. *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. Cambridge University Press.

Loharja, Henry, Lluìs Padró, and Jordi Turmo. 2018. Negation cues detection using CRF on Spanish product review text at NEGES 2018. In *Proceedings of NEGES 2018: Workshop on Negation in Spanish, CEUR Workshop Proceedings*, volume 2174, pages 49–54, Seville.

Marimon, Montserrat, Jorge Vivaldi, Núria Bel, and Roc Boronat. 2017. Annotation of negation in the IULA Spanish Clinical Record Corpus. *SemBEaR 2017*, 5(36.41):43–52.

Martí, M. Antónia, M. Teresa Martín Valdivia, Mariona Taulé, Salud María Jiménez Zafra, Montserrat Nofre, and Laia Marsó. 2016. La negación en español: análisis y tipología de patrones de negación. *Procesamiento del Lenguaje Natural*, 57:41–48.

Matsuyoshi, Suguru, Ryo Otsuki, and Fumiyo Fukumoto. 2014. Annotating the focus of negation in Japanese text. In *LREC*, pages 1743–1750.

Mehrabi, Saeed, Anand Krishnan, Sunghwan Sohn, Alexandra M. Roch, Heidi Schmidt, Joe Kesterson, Chris Beesley, Paul Dexter, C. Max Schmidt, Hongfang Liu, et al. 2015. DEEPEN: A negation detection system for clinical text incorporating dependency relation into NeGex. *Journal of Biomedical Informatics*, 54:213–219.

Minard, Anne-Lyse, Alessandro Marchetti, and Manuela Speranza. 2014. Event factuality in Italian: Annotation of news stories from the ITA-timebank. In *First Italian Conference on Computational Linguistics*, pages 260–264.

Minard, Anne-Lyse, Manuela Speranza, and Tommaso Caselli. 2016. The Evalita 2016 event factuality annotation task (FACTA). In *CLiC-it/EVALITA*, pages 32–39.

Miranda, Carlos Henriquez, Jaime Guzmán, and Dixon Salcedo. 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, Kevin J., Michael J. Becich, Jules J. Berman, Wendy W. Chapman, John R. Gilbertson, Dilip Gupta, James Harrison, Elizabeth Legowski, and Rebecca S. Crowley. 2004. Implementation and evaluation of a negation tagger in a pipeline-based system for information extraction from pathology reports. In *Medinfo*, pages 663–667.

Morante, Roser. 2010. Descriptive analysis of negation cues in biomedical texts. In *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC'10)*, pages 1429–1436, Valletta.

Morante, Roser and Eduardo Blanco. 2012. *SEM 2012 Shared Task: Resolving the scope and focus of negation. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics (*SEM)*, pages 265–274, Montréal.

Morante, Roser and Walter Daelemans. 2012. ConanDoyle-neg: Annotation of negation in Conan Doyle stories. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation*, pages 1563–1568, Istanbul.

Morante, Roser, Anthony Liekens, and Walter Daelemans. 2008. Learning the scope of negation in biomedical texts. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 715–724, Honolulu.

Morante, Roser, Sarah Schrauwen, and Walter Daelemans. 2011. Annotation of negation cues and their scope: Guidelines v1. *Computational Linguistics and Psycholinguistics Technical Report Series, CTRS-003*, pages 1–42.

Morante, Roser and Caroline Sporleder. 2012. Modality and negation: An introduction to the special issue. *Computational Linguistics*, 38(2):223–260.

Moreno, Antonio, Susana López, Fernando Sánchez, and Ralph Grishman. 2003. Developing a syntactic annotation scheme and tools for a Spanish treebank. In *Treebanks*. Springer, pages 149–163.

Mowery, Danielle L., Sumithra Velupillai, Brett R. South, Lee Christensen, David Martinez, Liadh Kelly, Lorraine Goeuriot, Noemie Elhadad, Sameer Pradhan, Guergana Savova, et al. 2014. Task 2: Share/clef ehealth evaluation lab 2014. In *Proceedings of CLEF 2014*, pages 31–42, Sheffield.

Mutalik, Pradeep G., Aniruddha Deshpande, and Prakash M. Nadkarni. 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.

Mykowiecka, Agnieszka, Małgorzata Marciniak, and Anna Kupść. 2009. Rule-based information extraction from patients' clinical data. *Journal of Biomedical Informatics*, 42(5):923–936.

Ohta, Tomoko, Yuka Tateisi, and Jin-Dong Kim. 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*, pages 82–86, San Diego, CA.

Oronoz, Maite, Koldo Gojenola, Alicia Pérez, Arantza D'ıaz de Ilarraza, and Arantza Casillas. 2015. On the creation of a clinical gold standard corpus in Spanish: Mining adverse drug reactions. *Journal of Biomedical Informatics*, 56:318–332.

Padró, Lluís and Evgeny Stanilovsky. 2012. Freeling 3.0: Towards wider multilinguality. In *Proceedings of the Language Resources and Evaluation Conference (LREC 2012)*, pages 2473–2479, Istanbul.

Palmer, Martha, Daniel Gildea, and Paul Kingsbury. 2005. The proposition bank: An

annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71–106.
Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up?: Sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing*, volume 10, of *EMNLP '02*, pages 79–86, Stroudsburg, PA.
Polanyi, Livia and Annie Zaenen. 2004. Contextual lexical valence shifters. In *Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications*, pages 1–6, Stanford, CA.
Polanyi, Livia and Annie Zaenen. 2006. Contextual valence shifters. In James G. Shanahan, Yan Qu, and Janyce Wiebe, editors, *Computing attitude and affect in text: Theory and applications*. Springer, Dordrecht, pages 1–10.
Pyysalo, Sampo, Filip Ginter, Juho Heimonen, Jari Björne, Jorma Boberg, Jouni Järvinen, and Tapio Salakoski. 2007. BioInfer: A corpus for information extraction in the biomedical domain. *BMC Bioinformatics*, 8(1):1–24.
Qian, Zhong, Peifeng Li, Qiaoming Zhu, Guodong Zhou, Zhunchen Luo, and Wei Luo. 2016. Speculation and negation scope detection via convolutional neural networks. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 815–825, Austin, TX.
Ren, Yafeng, Hao Fei, and Qiong Peng. 2018. Detecting the scope of negation and speculation in biomedical texts by using recursive neural network. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 739–742.
Rosenberg, Sabine and Sabine Bergler. 2012. UConcordia: CLaC negation focus detection at *SEM 2012. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics*, pages 294–300, Montreal.
Sandoval, Antonio Moreno and Marta Garrote Salazar. 2013. La anotacióin de la negación en un corpus escrito etiquetado sintácticamente. Annotation of negation in a written treebank. *Revista Iberoamericana de Lingüística: RIL*, 8:45–60.
Saurí, Roser and James Pustejovsky. 2009. FactBank: A corpus annotated with event factuality. *Language Resources and Evaluation*, 43(3):227–268.
Savova, Guergana K., James J. Masanz, Philip V. Ogren, Jiaping Zheng, Sunghwan Sohn, Karin C. Kipper-Schuler, and Christopher G. Chute. 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.
Segura Bedmar, Isabel, Paloma Martinez, and Cesar de Pablo Sánchez. 2011. Using a shallow linguistic kernel for drug–drug interaction extraction. *Journal of Biomedical Informatics*, 44(5):789–804.
Skeppstedt, Maria. 2011. Negation detection in Swedish clinical text: An adaption of NeGex to Swedish. *Journal of Biomedical Semantics*, 2 Suppl 3:1–12.
Socher, Richard, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D. Manning, Andrew Ng, and Christopher Potts. 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1631–1642, Seattle, WA.
Sohn, Sunghwan, Stephen Wu, and Christopher G. Chute. 2012. Dependency parser-based negation detection in clinical narratives. *AMIA Summits on Translational Science Proceedings*, 2012:1–8, San Francisco, CA.
Stricker, Vanesa, Ignacio Iacobacci, and Viviana Cotik. 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*, pages 1–7, Buenos Aires.
Styler IV, William F., Steven Bethard, Sean Finan, Martha Palmer, Sameer Pradhan, Piet C. de Groen, Brad Erickson, Timothy Miller, Chen Lin, and Guergana Savova, et al. 2014. Temporal annotation in the clinical domain. *Transactions of the Association for Computational Linguistics*, 2:143–154.
Szarvas, György, Veronika Vincze, Richárd Farkas, and János Csirik. 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*, pages 38–45, Columbus, OH.
Szarvas, György, Veronika Vincze, Richárd Farkas, György Móra, and Iryna Gurevych. 2012. Cross-genre and cross-domain detection of semantic uncertainty. *Computational Linguistics*, 38(2):335–367.

Taboada, Maite, Caroline Anthony, and Kimberly Voll. 2006. Methods for creating semantic orientation dictionaries. In *Proceedings of the 5th Conference on Language Resources and Evaluation (LREC'06)*, pages 427–432, Genoa.

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

Taylor, Ann, Mitchell Marcus, and Beatrice Santorini. 2003. The Penn treebank: An overview. In *Treebanks*. Springer, pages 5–22.

Uzuner, Özlem, Brett R. South, Shuying Shen, and Scott L. DuVall. 2011. 2010 I2B2/VA challenge on concepts, assertions, and relations in clinical text. *Journal of the American Medical Informatics Association*, 18(5):552–556.

Uzuner, Özlem, Xiaoran Zhang, and Tawanda Sibanda. 2009. Machine learning and rule-based approaches to assertion classification. *Journal of the American Medical Informatics Association*, 16(1):109–115.

Velldal, Erik, Lilja Øvrelid, Jonathon Read, and Stephan Oepen. 2012. Speculation and negation: Rules, rankers, and the role of syntax. *Computational Linguistics*, 38:369–410.

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

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

Vincze, Veronika. 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*, pages 28–31, Uppsala.

Vincze, Veronika, György Szarvas, Richárd Farkas, György Móra, and János Csirik. 2008. The BioScope corpus: Biomedical texts annotated for uncertainty, negation and their scopes. *BMC Bioinformatics*, 9(11):1–9.

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

Wilson, Theresa, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 347–354, Vancouver.

Wu, Stephen, Tim Miller, James Masanz, Matt Coarr, Scott Halgrim, and Cheryl Clark. 2014. Negation's not solved: Generalizability versus optimizability in clinical natural language processing. *PLoS ONE*, 9:1–11.

Zou, Bowei, Guodong Zhou, and Qiaoming Zhu. 2016. Research on Chinese negation and speculation: Corpus annotation and identification. *Frontiers of Computer Science*, 10(6):1039–1051.

Zou, Bowei, Qiaoming Zhu, and Zhou Guodong. 2015. Unsupervised negation focus identification with word-topic graph model. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1632–1636, Lisbon.