

Review

Negation and Speculation in NLP: A Survey, Corpora, Methods, and Applications

Ahmed Mahany ^{1,*} , Heba Khaled ¹, Nouh Sabri Elmitwally ^{2,3} , Naif Aljohani ⁴ and Said Ghoniemy ¹ 

¹ Faculty of Computer and Information Sciences, Ain Shams University, Cairo 11566, Egypt; heba.khaled@cis.asu.edu.eg (H.K.); ghoniemy1@cis.asu.edu.eg (S.G.)

² School of Computing and Digital Technology, Birmingham City University, Birmingham B4 7XG, UK; nouh.elmitwally@bcu.ac.uk

³ Faculty of Computers and Artificial Intelligence, Cairo University, Giza 12613, Egypt

⁴ Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia; nraljohani@kau.edu.sa

* Correspondence: ahmedmahany@cis.asu.edu.eg

Abstract: Negation and speculation are universal linguistic phenomena that affect the performance of Natural Language Processing (NLP) applications, such as those for opinion mining and information retrieval, especially in biomedical data. In this article, we review the corpora annotated with negation and speculation in various natural languages and domains. Furthermore, we discuss the ongoing research into recent rule-based, supervised, and transfer learning techniques for the detection of negating and speculative content. Many English corpora for various domains are now annotated with negation and speculation; moreover, the availability of annotated corpora in other languages has started to increase. However, this growth is insufficient to address these important phenomena in languages with limited resources. The use of cross-lingual models and translation of the well-known languages are acceptable alternatives. We also highlight the lack of consistent annotation guidelines and the shortcomings of the existing techniques, and suggest alternatives that may speed up progress in this research direction. Adding more syntactic features may alleviate the limitations of the existing techniques, such as cue ambiguity and detecting the discontinuous scopes. In some NLP applications, inclusion of a system that is negation- and speculation-aware improves performance, yet this aspect is still not addressed or considered an essential step.

Keywords: negation; speculation; uncertainty; natural language processing; corpora annotation; annotation guidelines; rule-based algorithms; machine learning; deep learning



Citation: Mahany, A.; Khaled, H.; Elmitwally, N.S.; Aljohani, N.; Ghoniemy, S. Negation and Speculation in NLP: A Survey, Corpora, Methods, and Applications. *Appl. Sci.* **2022**, *12*, 5209. <https://doi.org/10.3390/app12105209>

Academic Editors: Andrea Prati, Luis Javier García Villalba and Vincent A. Cicirello

Received: 23 April 2022

Accepted: 18 May 2022

Published: 21 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the rapid increase in the use of textual data in various domains—social media, product reviews, and the biomedical domain—the need arises to make use of all this available data. Social media has erupted as a method of online communication in which individuals can express their opinions and engage with each other at any time and place [1]. Through social media platforms, a vast amount of user-generated data in various formats is being shared and used online to express their opinions on products, services, and events in addition to communicating with each other [2]. These data are advantageous to governments and businesses because they obtain indirect insight into individuals' opinions about different aspects of life [3,4]. However, the majority of the users make use of colloquial language to express their opinions [5–7]. This adds to the challenge of text mining research, as these data are not syntactically correct sentences and feature both negation and speculation, which may misclassify the opinionated phrases [8]. Similar to social media data and product reviews, the rapid growth of utilizing electronic health records (EHRs) worldwide has made it possible to extract important and valuable information from biomedical data [9,10]. It is difficult to extract information from biomedical notes

and reports, given that they are written in narrative form and are prone to negating or speculative words that lead to incorrectly identifying the medical terms or events [11–13].

Both negation and speculation are commonly used in language, giving information on factuality and the polarity of facts. Negation is used to deny the characteristics of products or to reject something. On the other hand, speculative language is used to convey uncertainty about an event or idea. To identify instances of speculation and negation efficiently, it is necessary to find their cues, that is, those words (or morphological units) expressing speculation and negation, and then their scope, such as the tokens within the sentence that are affected by these cues [14]. Since negation and speculation are language-dependent, these phenomena are necessary in all natural languages [15]. Although the detection of negation has attracted much interest in the past decade, the available open-access resources for low-resource languages, such as models or corpora, are limited and tiny in size compared to the English language [16,17]. Speculation corpora are even scarcer than those for negation, and most focus on the biomedical domain [18–21].

1.1. Negation

Negation is considered a universal property of every human language [22]. It is defined as “the exact opposite of something; the act of causing something not to exist or to become its opposite” [23]. Negation means the refusal, denial, or rejection of a definite idea or event, expressed in four forms: correlation, contrariety, contradiction, and privation. Negation detection systems involve many aspects of all natural languages, but mainly cues, scope, event, and focus. First, the negation cues must be identified; that is, finding the words that express negation. Then it is necessary to discover their scope or event, that is, the part of the text that is affected by these negation cues. The cues may consist of a prefix, a single word, or multiple words, as negation can be expressed by a single or a compound cue, which may be morphological, lexical, or syntactic [24]. Sometimes a negation cue may be disconnected, where the word sequence does not indicate the negation. The negation scope may span tokens located both before and after the negation cue, yet it does not encompass the entire phrase or sentence [25]. Furthermore, in some languages it may be either connected or disconnected, as a language-dependent phenomenon. Events annotation is mainly performed in the biomedical domain, which includes many biomedical events where a negation cue directly negates an event; in most cases, this is a verb, noun, or adjective [21]. Recently, such annotation has been studied in other domains, such as literature [20] and reviewing [26]. Moreover, detection of the focus, which is the part with the explicitly negated scope, has been investigated by various research studies [27,28].

1.2. Speculation

Speculation, hedging, or uncertainty means that the author is not sure whether a sentence or an event is true or not [29]. Similar to that of negation, the detection of speculation is based on first identifying the speculation cue, then its scope or event. Moreover, the speculation cues may consist of single or multiple words, and their scope may be either coupled or detached. Nonetheless, the expression of speculation can be more complex in cases where the presence of a negation cue may be accompanied by a speculative content not negating [14]. Most work on speculation detection has focused on the biomedical domain, where clinical or medical notes contain uncertain expressions less frequently than negation [24,30].

Negation and speculation have been studied theoretically in almost all natural languages, and the detection of their scope is crucial to our recognition of a sentence’s meaning. The automatic detection of negation and speculation in texts is crucial to a wide range of natural language processing (NLP) tasks such as sentiment analysis (SA) [31,32], machine translation (MT) [33], and information extraction (IE) [34]. Indeed, it plays an important role and provides vital clues for further NLP applications because the factuality of information plays a vital role in these tasks. So, the computational treatment of these phenomena

has been studied in multiple languages [28,35,36] and domains [37–39]. However, we did not find a comprehensive survey of negation- and speculation-based systems.

The rest of the paper is organized as follows: Section 2 highlights the annotated corpora and summarizes their annotated elements and statistics. Section 3 serves as a comprehensive review on several techniques that have been used, including rule-based algorithms and machine- and deep-learning methods. Section 4 presents negation- and speculation-based systems and their effect on NLP tasks. The analysis of this review article, challenges, and open research issues are discussed in Section 5. Finally, Section 6 concludes the paper.

2. Corpora and Annotation Process

Several research works have addressed the construction of NLP resources to meet the challenges concerning the detection of negation and speculation [14,40]. Most of the annotated corpora have been collected in the biomedical domain, such as from clinical texts, radiology reports, and sometimes medical-related scientific publications [18,19]. For reasons of confidentiality and ethics, it is complicated to obtain access to biomedical corpora generated from patients' records, even if anonymized and in itself the anonymization of such data types may remove important events. Therefore, biomedical corpora obtained from scientific publications and reporting are increasingly available, such as those from the PubMed Central (PMC) portal (<https://pubmed.ncbi.nlm.nih.gov/> (accessed on 20 February 2022)). Other corpora are available for domains such as short stories [15,41], reviews [42], social media posts [8], news [28], and financial articles [43]. By contrast, regardless of their language, these and most data generated by social media platforms and product review websites are informal and contain grammatical mistakes. Such data types have many uncertainty and factuality issues, making them challenging to annotate.

The following subsections discuss the annotation process and guidelines on the phenomena of negation and speculation. In addition to exploring various corpora in several natural languages on the basis of analysis criteria such as domain, size, language, annotated elements, and availability, this section discusses the annotation process and the measure of inter-annotator agreement (IAA).

2.1. Annotation Guidelines

Negation and speculation corpora are usually annotated in two steps: first annotated is the token that indicates the phenomenon and the cue, then the cue's scope—the sequence of words that are negated/speculated. Especially in the biomedical domain, further elements, called events, can be annotated, and in the negation corpora of various domains besides the biomedical [35,44,45], the element of focus has been annotated.

Vincze et al. [14] created annotation guidelines (<https://rgai.sed.hu/sites/rgai.sed.hu/files/Annotation%20guidelines2.1.pdf> (accessed on 17 September 2021)) for use during the process of annotating biomedical texts in order to ensure the quality of the annotated data. These are the most-used guidelines for negation and speculation, and have been adapted for various domains [27,40,41] and multiple natural languages [9,19,46]. The general rule is to consider only those sentences with negating or speculative particles, including complex cases that assure non-existence/uncertainty about something without including any negating or speculative cues. The guidelines described suggest annotating keywords and their scope by following the 'min-max' strategy [14]. The smallest unit that expresses negation or uncertainty is labeled a cue, yet, in special cases, multiple tokens may be considered as the cue since a single word is unable to express the phenomenon. On the other hand, scope involves the longest sequence of tokens to be affected by a cue. To assure the quality of the annotation process, the guidelines clearly describe complex cases, such as the implicit negation employed in the Arabic language [19].

It is recommended that several annotators with the same level of experience regularly carry out the annotation process, during which they are not allowed to communicate but can consult an annotation expert. The guidelines written by an expert linguist should be

adhered to; however, any problematic cases that arise may lead this expert to adjust the guidelines. Although in most cases more than one annotator annotates the entire corpus, in large corpora another approach may be applied whereby a single annotator annotates each sentence, with random checks by a second annotator. A linguist expert resolves any disagreements that arise between annotators and assures the quality of the process by measuring IAA.

IAA measures the consistency between multiple annotators making the same decision and indicates how precise the annotation guidelines are. Usually, the Cohen [47] or Fleiss's [48] kappa coefficient measure is used in the annotation of negation and speculation data. Cohen's kappa measures the agreement between only two annotators, but Fleiss's kappa allows for more.

2.2. Corpora

2.2.1. English Corpora

This subsection explores ten English corpora that have been annotated with negation or speculation. As shown in Table 1, they cover texts extracted from various domains (biomedical, clinical, reviews, and others).

Table 1. English corpora annotated for negation and/or speculation.

| Ref. | Year | Corpus | Domain | Size | Neg. | Spec. | Scope | Event | Focus | Avail. |
|------|------|------------------|-----------------|------------------|------|-------|-------|-------|-------|--------|
| [49] | 2007 | BioInfer | Biomedical | 1100 | ✓ | | ✓ | | | ✓ |
| [21] | 2008 | GENIA | Biomedical | 9372 | ✓ | ✓ | | ✓ | | ✓ |
| [14] | 2008 | BioScope | Biomedical | 20,924 | ✓ | ✓ | ✓ | | | ✓ |
| [50] | 2010 | CoNLL-2010 | Biological | 40,289 | | ✓ | ✓ | | | ✓ |
| | | Wikipedia | | | | | | | | |
| [44] | 2010 | Product Review | Review | 2111 | ✓ | | ✓ | | | |
| [51] | 2011 | PropBank FOC | Journal stories | 3779 | ✓ | | | | ✓ | ✓ |
| [40] | 2012 | SFU Review | Review | 17,263 | ✓ | ✓ | ✓ | | | ✓ |
| [41] | 2012 | ConanDoyle-neg | Short stories | 4423 | ✓ | | ✓ | ✓ | | ✓ |
| [8] | 2015 | Twitter Negation | Tweets | 4000 | ✓ | | ✓ | | | |
| [27] | 2016 | DT-Neg | Dialogues | 27,785 responses | ✓ | | ✓ | | ✓ | ✓ |
| [45] | 2018 | SFU SOCC | Opinion | 1043 comments | ✓ | | ✓ | | ✓ | ✓ |

BioInfer was the first biomedical corpus to have negation annotation [49]. It is collected from abstracts of biomedical scientific papers, where 1100 sentences were annotated with entities, their relationship, and dependencies. In addition, 6% of the annotated relationships include negated cases using the 'not' cue. This corpus is publicly available (<http://mars.cs.utu.fi/BioInfer/> (accessed on 1 October 2021)), and was built for information extraction systems to find relations between proteins, genes, and RNAs.

The Genome Information Acquisition (GENIA) corpus was originally annotated with part of speech (PoS), syntactic trees, and terms [52,53]. In 2008, half the GENIA corpus, 1000 Medline abstracts, was annotated with negated biological events and two levels of uncertainty [21]. It consists of 9372 sentences in which 36,114 events were identified, and it is considered the first event-annotated corpus. A chief biologist and three graduates undertook the process of annotating negated, uncertain, and other event types (http://www.nactem.ac.uk/meta-knowledge/Annotation_Guidelines.pdf (accessed on 17 September 2021)). The corpus and its annotation guidelines are publicly available under the terms of the Creative Commons Attribution 3.0 Public License (<http://geniaproject.org/genia-corpus/event-corpus> (accessed on 17 September 2021)).

BioScope is a well-known biomedical corpus of information on negating/speculative cues and their scope [14]. It consists of documents from various sources (clinical radiology reports, full biological papers from the FlyBase and BioMed Central (BMC) website, and biological paper abstracts from the GENIA corpus [52]) covering many types of texts

in the biomedical domain. It comprises 6383 clinical texts (radiology reports), 2670 sentences from full scientific papers, and 11,871 sentences from scientific papers' abstracts. Two independent annotators annotated it, following the annotation guidelines written by a linguist expert. This expert followed the 'min-max' strategy during the annotation process, and modified it several times due to ambiguous cases arising between annotators. As a result, the guidelines followed in annotating the BioScope corpus were adapted to multiple domains [7]. Around 13% of the entire corpus contains negating expressions and more than 16% has speculative sentences. The reliability level of the annotation process was evaluated using the IAA rate, defined as the $F\beta=1$ measure of one annotation, considering the second to be the 'gold standard'. The average IAA of annotation was 0.85 for negation scope and 0.81 for speculation scope. The corpus is freely available for the purposes of academic and research (<https://rgai.sed.hu/sites/rgai.sed.hu/files/bioscope.zip> (accessed on 17 September 2021)).

The Computational Natural Language Learning (CoNLL) 2010 shared task was dedicated to identifying the speculation cues and their scope in two sub-corpora: biological publications and Wikipedia articles [50]. The first consists of 14,541 sentences, the biological part of the BioScope corpus representing the training set with an evaluation dataset formed of 790 of 5003 sentences from the PMC database. Wikipedia documents, the second sub-corpus, include 2484 of 11,111 sentences as a training set, with an evaluation set of 2346 of 9634 sentences. The corpus was annotated by two independent annotators who followed the 'min-max' strategy used on the BioScope corpus. The chief linguist who wrote the annotation guidelines resolved any disagreement between the annotators. The corpus is publicly available for research purposes (<http://www.inf.u-szeged.hu/rgai/conll2010st> (accessed on 2 February 2022)).

The research community has investigated other domains, and Morante et al. highlight the need for corpora covering domains other than the biomedical [54]. The annotation guidelines of the biomedical domain should be adapted to new domains, such as product reviews [44] and short stories by Conan Doyle [55].

The Product Review corpus is the first corpus in the review domain to have been annotated for negation [44]. It consists of 2111 sentences written as reviews for products, as extracted from Google Product Search. The number of negated sentences in this corpus is 679, and each sentence was annotated manually to define its cues and scope. Unlike the BioScope guidelines [14], the authors of this corpus included no negation cues in its scope. The IAA between two annotators on a dataset sample was 0.91, which is high. As a result, to complete the annotation of the entire corpus, each annotator applied the guidelines to a separate part of the reviews.

Blanco and Moldovan introduced the negation focus using the PropBank Corpus [51], the focus of negation in a sentence is the part of the scope that is most directly or explicitly negated. This corpus was selected for its semantic annotation, as it is not restricted to the biomedical domain [56]. The authors worked on 3779 sentences marked with MNEG, annotating the negation focus. Half of the PropBank FOC sentences were annotated by two annotators, achieving an IAA of 0.72. Any disagreements were then carefully examined and resolved, and the annotators were given a revised version of the guidelines to annotate the remaining half.

The Simon Fraser University (SFU) Review Corpus is a large annotated English corpus comprising 400 documents extracted from the Epinions.com website, belonging to various domains such as books, hotels, movies, and consumer product reviews [57]. Each sentence is labelled according to whether it is a positive or negative review, where a different person submits every review. Konstantinova et al. annotated the corpus at the token level with negating and speculative cues, and at the sentence level with the linguistic scope [40]. Their annotation guidelines were adapted from the BioScope corpus guidelines [14] and tailored to the review domain as an example: no cues were included in the scope. Of the sentences, 22% contained speculative instances, yet only 18% contained negating instances. Due to the large number of sentences, 17,263, the entire corpus was annotated by one linguist and

a second linguist annotated 10% of the documents at random to measure the IAA. The IAA was 0.92 in negation cues and 0.890 in speculation cues and 0.87 and 0.86 in their scope, respectively. The original corpus, its annotated form, and the annotation guidelines are published on their website (http://www.sfu.ca/~mtaboada/research/SFU_Review_Corpus.html) (accessed on 23 August 2021)).

The ConanDoyle-neg corpus consists of two of the 56 Sherlock Holmes short stories by Arthur Conan Doyle: *The Hound of the Baskervilles* (HB) and *The Adventure of Wisteria Lodge* (WL) [41]. Morante and Daelemans annotated them with negating cues, their scope, and event information. The annotation guidelines were adapted from the BioScope [14], yet have several differences. The authors focused on narrative texts, and in addition to defining the annotation of negating cues and their scope, they defined their events [55]. In this corpus, negation cues and their scope may be discontinuous. They annotated 850 and 145 negating sentences from the total of 3640 from the HB story and 783 from the WL. The corpus was annotated by two annotators, an MSc student and a researcher, both with a background in linguistics. The IAA was based on the F1 measure of 0.85 and 0.77 for scope in the HB and WL stories, respectively. The corpus and annotation guidelines are publicly available (<https://www.clips.ua.ac.be/BiographTA/corpora.html>) (accessed on 1 November 2021)). The corpus was used alongside the PropBank corpus in the SEM 2012 Shared Task*, (<https://www.clips.ua.ac.be/sem2012-st-neg/>) (accessed on 1 November 2021)) dedicated to resolving the scope and focus of negation [58].

The Twitter Negation Corpus contains tweets downloaded using Twitter API [8]. Two authors manually annotated the tweets with negation cues and their scope. The number of tweets involving negation was 539 of 4000 tweets, including 615 of negation scope. The IAA was measured at both the token and scope level, with values of 0.98 for the token scope and 0.73 for the full scope.

The DeepTutor Negation (DT-Neg) is the first corpus to focus on negation phenomena for dialogue-based systems [27]. It consists of texts extracted from tutorial interactions between high-school students to solve conceptual physics problems, as logged on an intelligent tutoring system. The authors automatically detected 2603 explicit negation cues in 27,785 student responses, using a compiled list of cue words [14,54]. Their annotation guidelines are based on Morante's work [54] to validate negation cues and annotate their scope and focus. The annotation process was performed by five graduate and research students to report 1088 negation cues and 458 instances of scope/focus. The IAA was based on 500 randomly selected instances divided into five parts, two annotators annotating each. The average sentence-level agreement was 0.66 for both scope and focus. The corpus is freely available for academic and research purposes (<http://deeptutor.memphis.edu/resources.htm>) (accessed on 23 September 2021)).

The SFU Opinion and Comments Corpus (SOCC): SOCC contains 10,339 opinion articles, with 663,173 comments, from *The Globe* and *Mail Canadian* newspapers [45]. The corpus has three categories: the articles, the comments, and the comment threads that are publicly available (<https://github.com/sfu-discourse-lab/SOCC>) (accessed on 5 February 2022)). The authors selected a subset of the corpus, 1043 comments, to annotate to four layers: constructiveness, toxicity, negation, and appraisal. The main target of this corpus was to study the relationship between negation and appraisal. A research student and a linguist performed the annotation using guidelines developed to annotate the negation cue, scope, and focus. The guidelines included a new annotation, 'xscope', to label the implied content of an inexplicit scope. The annotation process led to 1397 negation cues, and 1349 instances of scope, 1480 of focus, and 34 of 'xscope'. The IAA was based on 50 comments from the beginning of the annotation process and 50 from the conclusion. Agreement was measured using percentage agreement for nominal data with annotation and another for label and span, then combined to produce the average agreement. The agreement for the first 50 comments was 0.96, 0.88, and 0.47 for the cue, scope, and focus, respectively; for the last 50 comments, the agreement was 0.70, 0.63, and 0.43. The annotated corpus is publicly available under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International

License (<https://www.kaggle.com/mtaboada/sfu-opinion-and-comments-corpus-socc> (accessed on 7 March 2022)).

2.2.2. Corpora in Other Languages

This section explores non-English corpora annotated with negation and speculation, as shown in Table 2. Moreover, it highlights the first parallel negation corpus (English-Chinese). As seen in this survey, Spanish is the second language in terms of numbers of annotated corpora for these two phenomena, where it is the nearest language to Arabic.

Table 2. Non-English corpora annotated with negation and/or speculation.

| Ref. | Year | Corpus | Lang. | Domain | Size | Neg | Spec | Scope | Event | Focus | Avail |
|------|------|---------------------------------|----------------------|---|----------------------|-----|------|-------|-------|-------|-------|
| [20] | 2010 | Stockholm EPR | Swedish | Clinical | 6740 | ✓ | ✓ | | | | |
| [29] | 2014 | hUnCertainty | Hungarian | Misc. | 15,203 | | ✓ | | | | |
| [59] | 2014 | Review and Newspaper | Japanese | Review, Newspaper | 2147 | ✓ | | | | ✓ | ✓ |
| [60] | 2014 | EMC | Dutch | Clinical | 12,888 medical terms | ✓ | | | ✓ | | |
| [43] | 2015 | CNeSp | Chinese | literature, reviews, financial articles | 16,841 | ✓ | ✓ | ✓ | | | ✓ |
| [46] | 2016 | EMR | Chinese | Biomedical | 36,828 | ✓ | | ✓ | | | |
| [61] | 2016 | GNSC | German | Biomedical | 2234 | ✓ | ✓ | ✓ | ✓ | | |
| [19] | 2016 | BioArabic | Arabic | Biomedical | 10,165 | ✓ | ✓ | ✓ | | | |
| [9] | 2017 | IULA | Spanish | Biomedical | 3194 | ✓ | | ✓ | | | ✓ |
| [62] | 2017 | UHU-HUVR | Spanish | Clinical | 8412 | ✓ | | ✓ | ✓ | | |
| [26] | 2017 | SFU Review _{SP} NEG | Spanish | Review | 9455 | ✓ | | ✓ | ✓ | | ✓ |
| [28] | 2017 | News (Fact-Ita Bank) and Tweets | Italian | News stories, Tweets | 1591 | ✓ | | ✓ | ✓ | ✓ | |
| [15] | 2018 | NegPar | English-Chinese | Short stories | 5520 E5005 C | ✓ | | ✓ | ✓ | | ✓ |
| [24] | 2019 | ESSAI | French | Medical | 6547 | ✓ | ✓ | ✓ | | | |
| [24] | 2019 | CAS | French | Medical | 3811 | ✓ | ✓ | ✓ | | | |
| [13] | 2020 | REBEC | Brazilian Portuguese | Clinical | 3228 | ✓ | | ✓ | | ✓ | |
| [13] | 2020 | Clinical narratives | Brazilian Portuguese | Clinical | 9808 | ✓ | | ✓ | | ✓ | |
| [18] | 2020 | NUBES | Spanish | Biomedical | 29,682 | ✓ | ✓ | ✓ | ✓ | | ✓ |
| [35] | 2020 | NewsComm | Spanish | Comments | 4980 | ✓ | | ✓ | ✓ | ✓ | ✓ |
| [17] | 2021 | T-MexNeg | Mexican Spanish | Tweets | 13,704 | ✓ | | ✓ | ✓ | | ✓ |
| [36] | 2021 | ArNeg | Arabic | Wikipedia Biography Religion | 6000 | ✓ | | ✓ | | | |

Swedish: A subset of the Stockholm Electronic Patient Record (EPR) corpus [63] was randomly selected, giving 6740 sentences annotated with certain/uncertain expressions, as well as negating and speculative cues [20]. The Stockholm ERP corpus is a clinical corpus

with free texts in the category of Assessment, comprising patient records from the city of Stockholm. The annotation guidelines are similar to those for the BioScope corpus [14], but certain expressions, such as those containing question marks, were not annotated. Three annotators worked together on the entire corpus, consisting of 6996 expressions. The average number of annotated cues is 1624 speculative and 1008 negation. The IAA was measured by a pairwise F-measure, with approximate values of 0.8, 0.5, 0.9, and 0.6 for negation cues, speculation cues, certain cases, and uncertain cases. The authors plan to make the corpus available for research purposes.

Hungarian: Vincze presented the first Hungarian corpus to annotate uncertainty cues, where the hUnCertainty corpus consists of 15,203 sentences from several domains [29]. The author randomly selected 1081 paragraphs from the Hungarian Wikipedia, including 9722 sentences. The second sub-corpus consists of 300 parts (5481 sentences) of criminal news from the HVG Hungarian news portal. The annotation guidelines were adapted from seven categories from earlier works [64] and [38], with slight modifications. The frequency of uncertainty cues in Wikipedia and the news was 5980 and 2361, respectively.

Japanese: Matsuyoshi et al. annotated the first Japanese corpus with negation cues, scope, and focus, covering texts from two domains [59]. It consists of 5178 sentences of user review data randomly selected from Rakuten Travel data, and 5582 sentences from newspaper articles (Group A and B) from the Balanced Corpus of Contemporary Written Japanese (BCCWJ). After filtering the data, the authors proposed annotation guidelines for negation in Japanese, for which the total candidate sentences for reviews and newspaper articles were 1246 and 901, respectively. The total negation cues were 1023 and 762, respectively, where 300 and 190 sentences in each domain include a negated scope. Two annotators marked the focus for newspaper documents in Group A, where 66% of segments include focus particles. The next step was to resolve disagreements to annotate the remaining part of the entire corpus using a single annotator, and 165 out of 490 of instances of negation scope were found to include focus particles. The authors plan to publish their annotated corpus for the research community (<http://cl.cs.yamanashi.ac.jp/> (accessed on 1 January 2022)).

Dutch: The Erasmus Medical Center (EMC) clinical corpus includes several types of anonymized clinical documents, such as those by general practitioners (GP), specialists' letters (SP) from the IPCI database, radiology reports (RD), and discharge letters (DL) from the EMC in Netherlands [60]. The authors extracted the Dutch medical terms from the Unified Medical Language System (UMLS) for use in annotating the corpus with negation. A medical term was marked as 'Negated' if the evidence did not exist; otherwise, it was annotated as 'Not Negated'. Two independent annotators followed the compiled guidelines, with support from a linguist expert to resolve disagreements. As a result, 1804 negated terms were reported of the 12,888 instances of medical terms. The IAA for the entire corpus was 0.92, on average. The authors mention that their corpus is free for research, upon request.

Chinese: The Chinese Negation and Speculation (CNeSp) corpus consists of 16,841 sentences from various domains, annotated with negating and speculative cues and their scope [43]. It includes 19 scientific articles, 311 financial articles, and 821 product reviews. The authors adjusted the BioScope corpus guidelines [14] to make it suitable for the Chinese language. These modifications include that the subject should be within the scope, and that the scope should be a continuous set of tokens. Two annotators carried out the annotation process and a linguist expert resolved disagreements. As a result, the percentages of negating sentences were 13.2, 17.5, and 52.9 for scientific articles, financial articles, and product reviews, respectively. These percentages indicate that the review domain has high percentages of negation, independent of language. Moreover, the percentages of speculation were 21.6, 30.5, and 22.6, respectively, similar to the percentages of negation. IAA kappa-based average values of 0.90 for negation and 0.89 speculation show that the annotation guidelines were well formulated. The authors have made this corpus publicly

available for research purposes (<http://nlp.suda.edu.cn/corpus/CNeSp/> (accessed on 10 January 2022)).

Chinese: Kang et al. collected 36,828 sentences from 400 admission notes and 400 discharge summaries of one month's data, March 2011, from the EMR database of Peking Union Medical College Hospital [46]. This corpus includes data from four biomedical categories: diseases, symptoms, treatments, and laboratory tests. The BioScope corpus guidelines [14] and other Chinese guidelines [43] were adapted to annotate the corpus. Three domain experts formulated the initial draft and then applied the guidelines to a corpus sample to refine them. The outcome of the process is 21,767 negating sentences from the entire corpus. The IAA measure was based on only 80 notes, annotated by two annotators, reaching 0.79 measured by Cohen kappa.

German: The German Negation and Speculation Corpus (GNSC) is the first German annotated corpus in the biomedical domain [61]. It consists of eight anonymized discharge summaries containing medical histories and 175 clinical notes from the nephrology domain, which are shorter than discharge summaries. First, the medical terms were automatically pre-annotated using an annotation tool [65], following the UMLS unified coding standards for EHRs, as per the predefined types. Secondly, a human annotator rectified incorrect annotations and classified whether a given finding occurred in a positive, negative, or speculative context. Finally, a linguist revised the annotated sentences to assure the data quality. As a result, for discharge summaries the negative and speculated sentences were 106 and 22 of 1076 sentences, and for clinical notes 337 and 4 of 1158 sentences. The results show that in this corpus speculation rarely arises.

Arabic: The BioArabic is the first Arabic corpus annotated with negation and speculation for biomedical Arabic texts [19]. The corpus consists of 10,165 sentences extracted from 70 medical and biological articles collected from the *Iraqi Journal of Biotechnology*, the *Journal of Damascus University for Health Sciences*, and *Biotechnology News*. An expert linguist adapted the annotation guidelines of the BioScope corpus [14] to Arabic biomedical texts. Five linguist annotators performed the annotation according to the guidelines developed, as described in detail in the paper. The expert linguist resolved disagreements arising between the annotators to report 1297 negated sentences and 1376 speculative ones. Unfortunately, the measure of agreement between the annotators was not reported, and the corpus is not readily available online.

Spanish: The IULA Clinical Record corpus consists of 3194 sentences extracted from 300 anonymized clinical records from the main hospitals in Barcelona, Spain [9]. These sentences were manually annotated with negation cues and their scope by three linguists, advised by a clinician. Marimon et al. used the English annotation guidelines from several domains: BioScope from the biomedical domain, and ConanDoyle-neg from the short stories domain and general concepts [14,54,66]. As in the BioScope guidelines, the annotation guidelines include neither the negation cue nor the subject of the record in the scope. After applying these annotation rules to the entire corpus, the team reported 1093 negated sentences. The percentage of sentences that include negation scope, roughly 35%, is relatively high compared with English corpora. The IAA Cohen kappa rates were 0.85 between annotators 1 and 2 or 3 and 0.88 between annotators 2 and 3. The corpus is publicly available with a CC-BY-SA 3.0 license (http://eines.iula.upf.edu/brat/#/NegationOnCR_IULA/ (accessed on 5 January 2022)).

Spanish: UHU-HUVR, a corpus with 276 radiology reports and 328 patients' personal histories (written in free text) obtained from the Virgen del Rocío Hospital in Seville, Spain [62]. It is considered the first Spanish corpus to include affixal negation. Cruz et al. adapted the Thyme corpus guidelines to annotate the negation cues and scope, using two domain experts [67]. The results of the annotation are 1079 negated sentences of the 3065 in the patient histories and 1219 negated sentences of the 5347 in the radiology documents. As observed in the report, the percentage of negated sentences in the histories is relatively high since they are patients' descriptions of clinical conditions, which are not 100% factual,

whereas the radiology report gives the radiologists' observations. The IAA measure for the negation scope was 0.72.

Spanish: The SFU Review_{SP} Negation (NEG) extends the SFU Review corpus [57]; it consists of 400 review documents from domains such as movies, music, and various product reviews, extracted from the Ciao.es website [26]. Each domain contains 25 positive and 25 negative review documents, based on the stars awarded by the reviewer. It is the first Spanish corpus to include events in the annotation and reflect any discontinuous negation cues. Like its English equivalent, the entire corpus was manually annotated with negation cues and their consequent scope and events [40]. Although the Bioscope corpus guidelines had been written for the biomedical domain, for the review domain the authors adapted these guidelines alongside a typology of negation patterns in Spanish [14]. The annotation process was supervised by experts and performed by two trained annotators, who came up with 3022 negated sentences in the 9455 sentences. The IAA kappa coefficient is 0.95 for negated events and 0.94 for scope, both of which are relatively high compared to other Spanish negation annotated corpora. This corpus is publicly available under a Creative Commons Attribution Noncommercial-ShareAlike 4.0 International License (<http://clic.ub.edu/corpus/es/node/171> (accessed on 13 November 2021)).

Italian: Altuna et al. annotated two corpora from contrasting domains [28]. The first consists of 71 documents for stories adapted from Fact-Ita Bank [68]. The second consists of 301 tweets used in the Factuality Annotation (FactA) task at Evaluation of NLP and Speech Tools for Italian (EVALITA) 2016 [69]. The authors based their annotation guidelines on earlier guidelines [51,55], including negation cues, scope, and focus. In general, every negation cue is associated with its scope and focus, and the first corpus of 1290 sentences contains 282 negating cues and 278 negated sentences. The second corpus comprises 59 negated sentences and 71 negation cues. The authors are to make the annotated data available on the Human Language Technology-Natural Language Processing (HLT-NLP) website at the FBK organization to implement a system for negation detection in Italian. The agreement on the annotation of the scope achieved roughly 0.7 and 0.6 for focus, where IAA was based on the average pairwise F-measure.

English–Chinese: Many corpora in various natural languages have been annotated with monolingual negation to detect negation phenomena; however, they have not been studied in parallel across languages. The NegPar corpus is the first English–Chinese parallel corpus to be annotated with negation for narrative texts [15]. The corpus used four short stories from Conan Doyle's Sherlock Holmes short stories and mapped them to their Chinese translation by Mengyuan Lin. The annotation guidelines of the ConanDoyle-neg corpus [41] were adapted for the English part of the corpus. Although the English side of the corpus had already been annotated, most was reannotated to capture semantic phenomena. Qianchu Liu, a native Mandarin speaker, created the annotation guidelines for the Chinese part with support from the other two authors. Chinese translation often converts a positive English statement into a negative one; therefore, the number of negated sentences in the Chinese corpus is slightly higher than in the original English corpus. The annotation process was based on projections from the English corpus, where 1304 and 1762 negated sentences were annotated for the English and Chinese parts, respectively. The annotation projection offers imperfect help in the annotation process, where the word-level F1 measure was 0.39, 0.45, and 0.24 for negation cue, scope, and event, respectively. The corpus, with its annotation guidelines, has been published for public use (<https://github.com/qianchu/NegPar> (accessed on 25 January 2022)).

French: ESSAI is a corpus of French clinical trials obtained mainly from the National Cancer Institute registry website [13,24]. The French protocol for such clinical trials has two parts: a summary of the trial, which presents its purpose and the methods applied, and its detailed description. ESSAI comprises 6547 sentences; 1025 sentences are negating, while 630 are speculative. The IAA measure of the negation annotation is 0.80, which is not high compared to English corpora [13].

French: Grabar et al. introduced another French corpus called CAS. This consists of 3811 sentences for clinical cases adapted from scientific literature and training materials [70]. These clinical cases were published in various journals and websites in French-speaking countries such as Belgium, Switzerland, and Canada, and relate to medical specialties [53]. The purpose of clinical cases is to describe clinical situations for real (de-identified) or fake patients. This corpus is automatically annotated with negation using different supervised learning techniques, trained on the ESSAI corpus. Uncertainty cues and their scope were annotated for CAS on the basis of heuristic rules. Of the 3811 sentences, 804 are negating and 226 include speculation. Later in 2021, two annotators manually verified CAS for the negation annotation, achieving a Kappa Cohen's IAA of 0.84 [13].

Brazilian Portuguese: Dalloux et al. presented the Brazilian Portuguese clinical trial protocols provided by Brazilian registration of clinical trials (REBEC) [13]. Each protocol includes a sample title, scientific title, description, and inclusion and exclusion criteria. Negation in this corpus provides valuable information on the specification of the target and the patient's recruitment. Three students annotated different parts of these protocols and came up with 643 negating sentences out of 3228, which is considered relatively low in this corpus. The authors compiled another Brazilian Portuguese corpus, collected from three hospitals in Brazil, covering medical specialties such as cardiology and nephrology [13]. This corpus contains 9808 sentences on clinical narratives. The negation cues and their scopes were manually annotated, and include 1751 negating cases with a Cohen's kappa coefficient of 0.74.

Spanish: Negation and Uncertainty annotations in biomedical texts in the Spanish (NUBES) corpus make it the largest publicly available Spanish corpus for negation. It is the first to have annotations of speculative cues and scope and speculative events [18]. NUBES consists of 29,682 sentences from a Spanish private hospital's anonymized health records. The corpus was extracted from several sections, namely Surgical History, Physical Examination, and Diagnostic Tests, and divided into 10 batches. Together, two linguists drafted the initial annotation guidelines sourced from IULA [9] then extended them to include uncertainty. One batch was annotated using the initial guidelines, then improved after a medical expert resolved the disagreements between the annotators. The Cohen kappa IAA was calculated at 0.8 based on the sample batch. Later, the same batch was annotated by a third annotator to resolve all disagreements and create a 'gold standard'. Consequently, the other nine batches could be annotated by a single annotator using the final version of the annotation guidelines. This corpus of 29,682 sentences contains 7567 negated and 2219 uncertain instances. The authors enriched the IULA corpus by incorporating uncertainty and have made the guidelines publicly available (<https://github.com/Vicomtech/NUBes-negation-uncertainty-biomedical-corpus> (accessed on 20 January 2022)).

Spanish: NewsComm was the first Spanish corpus to have been collected from newspapers and annotated for negation phenomena. Also, it is considered to be the first Spanish corpus to be manually annotated with negation alongside negation cues and scope. The corpus consists of 2955 comments posted in response to 18 news articles on nine topics (immigration, politics, technology, terrorism, economy, society, religion, refugees, and real estate) in an online Spanish newspaper, two articles per topic [35]. Linguistic analysis of the negation focus arrived at 10 conditions for the various forms of negation in this corpus. The criteria for annotation are described in detail in the guidelines (<http://clic.ub.edu/publications> (accessed on 25 February 2022)). The two trained annotators selected to annotate this corpus manually had earlier annotated the SFU Review_{SP} NEG corpus [26]. They found 2965 negating structures in the corpus, with corresponding negation cues, scope, and focus. The result of the annotation process shows that 45% (2247 of 4980 sentences) are negating. Furthermore, the kappa IAA measure is 0.83, a high value for this first annotated Spanish corpus to have a negation focus. The authors made the corpus freely available for research purposes.

Mexican Spanish: T-MexNeg is the first Mexican Spanish corpus annotated for negating phenomena [17]. It consists of 13,704 tweets collected from September 2017 to April 2019 using the standard streaming APIs from Twitter (<https://developer.twitter.com/en/docs/tutorials/consuming-streaming-data> (accessed on 1 March 2022)). To limit the collection to Mexico, tweets were filtered by the language tag ‘es’ and the user location ‘mx’. Although the corpus is very large, it was annotated manually for negation cue, scope, and event. In addition, the authors adapted the SFU Review_{SP} NEG [26] guidelines to Mexican Spanish and the nature of the collected corpus. The annotation process was carried out in two stages: binary classification for the presence of negation, and manual annotation by three teams of two annotators, who were linguistics students, together with a linguist. As a result, the T-MexNeg corpus of 13,704 sentences includes 4895 tweets with negation. The IAA was measured with Cohen’s kappa coefficient and has a value of 0.89, which is relatively high. The authors made this corpus publicly available for research purposes (https://gitlab.com/gil.iingen/negation_twitter_mexican_spanish (accessed on 1 March 2022)).

Arabic: The ArNeg corpus is another Arabic corpus that has been annotated with negation for formal Arabic texts [36]. It consists of 6000 sentences collected from Wikipedia and the King Saud University-Corpus for Classical Arabic (KSUCCA). The corpus covers sentences from topics such as biography, media, science, and technology. Mahany et al. wrote clear annotation guidelines, to which the two independent Arabic native speakers adhered in the annotation process. The percentage of the negated sentences was found to be 18% for the Wikipedia sub-corpus and 29% for the KSUCCA sub-corpus. One of the annotators applied the guidelines to 20% of the entire corpus, and between all five the IAA was recorded at 0.98.

This section has explored corpora annotated for negation and speculation in 13 languages from six language families: Germanic languages (English, German, Swedish), Romance (French, Spanish), Uralic (Hungarian), Semitic (Arabic), Sinitic (Chinese), and the Japonic language family (Japanese).

3. Methods and Techniques

The numerous systems developed for negation and speculation detection range from algorithmic (rule-based) [10,37,71], machine [72,73], and deep learning approaches [13,36,74], or hybrids [75,76]. Most rule-based algorithms depend on simple rules made up of regular expressions and detect the scope of negation using dependency and parse trees. By contrast, supervised learning techniques use various classifiers to detect these phenomena at different levels: cue, scope, event, and focus.

3.1. Rule-Based Methods

Initially, the processing of negation and speculation in natural language processing (NLP) was investigated using a handcrafted list of semantic and syntactic rules. The most famous rule-based methods are shown in Figure 1, and are presented in this section.

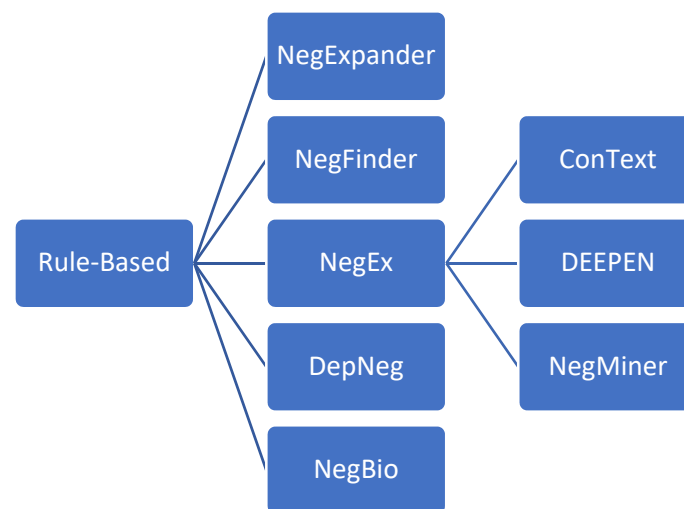


Figure 1. Taxonomy of Rule-based Algorithms.

In the late 1990s, Aronow, Fangfang, and Croft developed the NegExpander algorithm to solve the problem of finding negated UMLS, coding standard for EHRs, by constructing conjunctive phrases [77]. The algorithm's input is a sentence with definite UMLS phrases and part of speech (POS) tags to detect the negation scope. It initially looks for conjunctive words within a sentence, then searches for negation words within the conjunctive phrases. If a single UMLS term is negative, all UMLS phrases within the conjunctive sentence are considered to be negative.

In 2001, NegFinder was developed. This is a negation-detecting algorithm using a context-free grammar parser to detect it in medical documents [66]. A set of 40 medical documents from several specialties were manually checked to develop a set of rules to detect the negated patterns. The input to the algorithm consists of coded representations for UMLS concepts extracted from a medical narrative. As a result, the algorithm generates an index with one entry for each concept that shows negated concepts.

NegEx is a simple, regular, expression-based algorithm developed to detect negation cues in the English language and identify the affected medical terms [71]. The algorithm takes a sentence with identified UMLS terms and negation words called triggers, which occur before or after a finding. The algorithm relies on three negation phrases: pre-UMLS, post-UMLS, and pseudo-negation. A pseudo-negation expression comprises negative grammatical words, yet this expression is an unreliable negation indicator. All UMLS terms within a window size of 0 to 5 are assigned a negation status. A UMLS term is considered negated, within the scope of negation, if a negation word is located within five tokens of a finding.

ConText, derived from the NegEx [71], covers negation, temporality, and the subject concerned for conditions in the clinical texts [78]. Although it depends on regular expressions, it defines the scope of the trigger terms differently and identifies three more contextual properties (hypothetical, historical, and experienter). The algorithm searches for the negated trigger terms preceding or following a clinical condition. It changes the default value, affirmed for negation, of the trigger term if the condition is within the scope of this trigger term. The negation scope extends from the right of a trigger term to the end of a sentence, unlike NegEx, or to a specific window size.

Elazhary extended the NegEx algorithm to propose the NegMiner tool to handle connecting and multiple negation [79]. The pre-negation and post-negation terms were updated by adding more tokens, and new pre-contiguous and post-contiguous negation terms were added. This tool consists of three modules: the knowledge base and update module, which allows the user to modify negation, and termination terms, in addition to the mining module. Furthermore, the tool takes input as a PDF file that includes multiple sentences, one per line, and generates the results as an output PDF file.

Some other researchers have proposed dependency-based parsers to detect the negation scope in complex sentences. In the simplistic form, regular expressions, the researchers assumed that in their corpora, mainly in clinical reports, negation does not include complexly structured sentences [71]. The dependency parser-based negation (DepNeg) used to dig deeply into the syntactic structure of the sentences to detect the complex negation cases for the target-named entities in clinical reports [80]. Using a dependency parser allowed the authors to reduce the problem encountered by rule-based algorithms in detecting the scope of negation. It expresses the syntactic structure as directed dependencies between the tokens. The authors built their syntactic rules for negation using 41 sentences from the I2B2/VA corpus [81] based on the dependency paths between the named entity and the negation terms. If a named entity is on the same dependency path as the negation clue, this means that it is negated. DepNeg was compared with NegEx by adding to the Clinical Text Analysis and Knowledge Extraction System (cTAKES) (<http://www.ohnlp.org> (accessed on 10 March 2022)), an information extraction tool for medical records, and the results showed that DepNeg reduced the number of false positives.

DEpEndency ParsEr Negation (DEEPEN) (<https://code.google.com/p/deepen> (accessed on 12 March 2022)) is an open-source algorithm for negation detection in clinical texts [10]. It uses dependency patterns after NegEx as a post-processing step to remove false positives of negative findings. DEEPEN uses the transition-based dependency parser to find the dependency relationship between the negation words and medical terms, unlike DepNeg, which used the graph-based dependency parser. The algorithm rules were built using the Stanford Dependency Parser (SDP) [82] based on 1929 clinical reports from the Indiana University dataset, where only medical terms negated by the NegEx algorithm are considered. After that, a production chain of three levels of tokens is generated for each sentence, marking the medical term as negated if found in the production chain. Unfortunately, the basic rule failed in sentences with complex structures; further rules were added to address these structures. Moreover, two domain experts reviewed the output after adding these rules, and the IAA between them was 0.95.

NegBio is an open-source tool for detecting negation and uncertainty in radiology reports [37]. This algorithm uses the universal dependency graph to find the grammatical relationships in a sentence, in addition to subgraph matching for graph traversal search. Consequently, the scope of negation and uncertainty is not limited to a specific window size. As a result, NegBio should outperform other rule-based algorithms by identifying the scope of triggers that include explicit negation cases as complex sentences. The input is a list of findings with corresponding UMLS, which are processed to detect negated and uncertain cases.

3.2. Supervised Techniques

In the past decade, interest has increased in the automatic detection of negation and speculation using supervised machine learning approaches, especially in English. Nevertheless, the lack of annotated corpora for the other natural languages has impeded the development of such approaches. This subsection presents the commonly used supervised learning approaches for negation and speculation detection, and Table 3 summarizes these approaches.

Table 3. Supervised techniques for negation and speculation detection.

| Ref. | Year | Language | Negation | | Speculation | |
|------|------|------------------|--|-----------------|-----------------------|----------------------------|
| | | | Features | Methods | Features | Methods |
| [83] | 2007 | English | | | BoW | SVM |
| [84] | 2007 | English | | | Bigrams, Trigrams | Weakly Supervised Learning |
| [72] | 2009 | English | PoS, Lemma, Syntactic | SVM, CRF, TiMBL | | |
| [73] | 2009 | English | | | PoS, Lemma, Syntactic | SVM-CRF-TiMBL |
| [75] | 2009 | English | | | BoW, PoS, Syntactic | SVM |
| [44] | 2010 | English | PoS | CRF | | |
| [85] | 2012 | English | | | Syntactic | SVM |
| [86] | 2013 | English | Syntactic | SVM Tree Kernel | Syntactic | SVM Tree Kernel |
| [76] | 2014 | English | | | Handcrafted Rules | MRS Crawler |
| [87] | 2016 | English | Syntactic, Cue | CNN | Syntactic, Cue | CNN |
| [88] | 2016 | English | WE | BiLSTM | | |
| [89] | 2016 | English | WE, PoS | BiLSTM | | |
| [90] | 2017 | English | | | WE | CNN with attention |
| [91] | 2017 | English | Lexical, Syntactic Embedding, BoW, BoC | SVM+CRF | | |
| [46] | 2017 | Chinese | | CRF | | |
| [92] | 2018 | English, Chinese | Latent Structural, Cue | Semi-CRF | | |
| [93] | 2018 | Spanish | WE, PoS | LSTM | | |
| [94] | 2018 | English, Chinese | PoS | BiLSTM | | |
| [74] | 2019 | English | Cue | NegBERT | | |

Table 3. Cont.

| Ref. | Year | Language | Negation | | Speculation | |
|------|------|--|-----------|--------------|----------------|--------------------|
| | | | Features | Methods | Features | Methods |
| [11] | 2019 | English, Chinese | BERT, PoS | BiLSTM | BERT, PoS | BiLSTM |
| [24] | 2019 | French | WE, PoS | BiLSTM-CRF | WE, PoS | BiLSTM-CRF |
| [95] | 2019 | Arabic | | | Syntactic, PoS | GAN with Attention |
| [96] | 2020 | English, Chinese | Syntactic | RNN-CRF | Syntactic | RNN-CRF |
| [39] | 2020 | English | Syntactic | BiLSTM & CNN | | |
| [97] | 2020 | English, French, Brazilian, Portuguese | Syntactic | STRNN | | |
| [13] | 2020 | Brazilian, Portuguese | WE, PoS | BiLSTM-CRF | | |
| [36] | 2021 | Arabic | WE, Cue | BiLSTM | | |
| [17] | 2021 | Mexican Spanish | PoS, Cue | CRF | | |

Medlock and Briscoe treated speculation detection as a binary classification problem and proposed a weakly supervised machine learning model [83]. Their support vector machine (SVM)-based model depends on a representation approach using a manually labeled corpus' bag-of-words (BoW). In 2008, Szarvas adapted this model [83] by manually limiting the feature set and including bigrams and trigrams into the feature representation [84]. The weakly supervised models considered the entire sentence as either speculative or not speculative.

Since 2008, with the construction of the BioScope corpus [14], the scope detection problem began to be extensively solved using supervised machine learning approaches [14]. Morante and Daelemans modeled the negation scope detection problem as two consecutive classification tasks: task 1 detects the negation signal, and task 2 finds the scope of the detected negation signal [72]. A memory-based learning algorithm, Tilburg Memory-Based Learner (TiMBL) (<https://languagemachines.github.io/timbl/> (accessed on 20 March 2022)), was used with features such as PoS and lemma of the token to identify the negation signals. In the second phase, the TiMBL, SVM, and conditional random field (CRF) classifiers were combined with shallow syntactic features to pass their predictions to a metalearner CRF classifier to predict the negation scope. In another study [73], the authors implemented the same approach to find the scope of speculation cues in biomedical texts. Although this system is portable for similar negating and speculative detection, it depends on a highly engineered features set. Zou et al. proposed a tree kernel classifier based on the constituent and dependency structured syntactic features, proving their efficiency in capturing the cues' relation and scope [86]. This system considered the problem a binary classification task, where the SVM classifier used a kernel function between two trees to classify whether the token lay inside or outside the scope.

The first joint conference on lexical and computational semantics (SEM 2012*) published an annotated corpus with negation to prompt for systems to detect negation cues, scope, and negated events [58]. Of the 12 participating teams, the first to submit a system was from the University of Oslo (UiO1), and it was selected as the best-performing system for this task [85]. As in previous studies, an SVM classifier was used, but it depends on syntactic constituents generated by following the path from the negation cue to the root of the syntactic tree. In this approach, Read et al. tried to handle the discontinuous scope of negation by removing unnecessary words from the negation scope. Even though their system achieved the top-ranked system with an F1-score of 78, it is a highly engineered system and in the case of a single constituent, fails to capture discontinuous scope. Enger et al. developed a simple open-source Python-based tool (<https://github.com/marenger/negtool> (accessed on 18 March 2022)) to identify negation scope using an approach called maximum-margin [91]. This approach divides the problem into two sub-tasks; the cue detection task uses lexical features with a binary SVM classifier. The second sub-task, scope resolution, is a sequence-labeling task that uses syntactic-based features with a linear-chain CRF model; however, like (UiO1), it failed to handle discontinuous scope [85].

Other studies tackled the problem of scope detection as a sequential labeling task, using CRF-based models. For example, Li and Lu proposed a semi-CRF-based model (<https://github.com/leodotnet/negationscope> (accessed on 3 March 2022)) to identify negation scope in the SEM 2012* corpus [92]. The features set includes long-distance dependencies, latent structural information, and negation cue-related features such as the relative positions of each token and cue, as in another study [98]. Their model was designed to identify the negation scope at the span level to capture discontinuous and partial scope, and another latent-based CRF model was used to represent the implicit patterns. As a result, this system outperformed the best performing system (UiO1) at SEM 2012* [85].

CRF-based models were used to detect negation scope in other natural languages. Kang et al. used a CRF model with character and word embedding for Chinese clinical notes to identify negation scope [46]. Besides the character and word embedding, a feature known as 'In Previous Scopes' was added to handle the nested negation scope within multi-cue sentences. The sequence labeling task was divided into cue and scope detection, where the output of the cue detection task was an input to the CRF-based scope detection. Bel-Enguix et al. used two methodologies to compare their T-MexNeg (TMN) corpus to the SFU Review corpus in detecting negation phenomena in Spanish social media data [17]. The CRF-based system depends on several attributes, such as PoS and cue features.

The previously explored studies prove machine learning methods' effectiveness in the scope detection task. In treating the scope detection as a sequence labeling problem, these CRF-based models outperform the other machine learning method, SVM. Nevertheless,

their success depends heavily on a highly engineered feature set that needs more effort. Therefore, deep learning-based methods are the best candidates that learn latent features automatically on different NLP sequence labeling tasks.

Qian et al. proposed a convolutional neural network (CNN) model to capture various syntactic information between a cue and its tokens within a sentence to detect the negation and speculative scope [87]. The feature set consists of the widely used constituency, dependency parse trees, and relative position of cues and candidate tokens. These are combined to form a single vector that finally is fed into a SoftMax layer to detect the negation or speculative scope. By modeling the relationship between cues and their scope accurately rather than considering the problem as a binary classification task, this work performed better than previous studies; however, it did not capture long-distance syntactic dependency and performed poorly on the BioScope sub-corpus of full papers. McKenna and Steedman proposed the Structural Tree Recursive Neural Network (STRNN) to detect the negation scope on the SEM 2012* test corpus [97]. First, this network takes a binarized syntax tree, where a local vector represents each word, as input for the sentence's syntactic structure. After that, it recursively combines the local vectors for constituent states to form a global state vector in an upward phase. Finally, it recursively divides the global vector through the parse tree during the downward phase to obtain the output using a SoftMax layer.

Other studies addressed speculation detection with attention-based deep learning architectures, proving their efficiency in many NLP tasks. These architectures mainly help the neural network to focus on the most relevant information. Adel and Schütze proposed CNNs with external and sequence-preserving attention mechanisms to capture the most relevant features that represent the long input sequences [90]. Their CNN model includes a convolutional layer, three max-pooling layers, one hidden layer, and a logistic output (https://github.com/heikeadel/attention_methods (accessed on 9 March 2022)). Al-Khawaldeh proposed an attention-based generative adversarial networks (GAN) model to identify the speculation scope in Arabic biomedical data [95]. This model takes PoS, constituency, and dependency features to detect the speculative phrase. The experimental results show that the performance of speculation detection systems was improved by using attention-based models.

As in our survey, the bidirectional long short-term memory network (BiLSTM) is the usual deep learning method for negation and speculation scope tasks. Lazib et al. tackled negation detection in the review domain, using various recurrent neural networks that are recurrent neural networks (RNNs)-based (LSTM, BiLSTM, and GRU) with word embedding representation [88]. The word embedding representation of tokens was automatically extracted as the only feature used in their model. The system's performance proves that RNNs with WE models are candidates to replace the rich features-dependent classifiers for negation scope detection. Fencellu et al. used the BiLSTM model, producing in- and out-of-scope tokens within a sentence to identify the negation scope in the SEM 2012* corpus [89]. The word embedding and PoS cue feature fed into the BiLSTM network to detect the negation scope of this cue. This approach modeled the negation scope detection as independent predictions for each token.

In English, negation scope boundaries often depend on punctuation, especially in old-style writing as in Conan Doyle's short stories. Fencellu et al. tried to alleviate the effect of this issue with a dependency-based BiLSTM model that processes dependency trees using encodings for words and dependency relations [94]. The authors developed a set of models based on English, with a cross-lingual word representation and universal PoS features, yet then applied them to Chinese. These models were evaluated on the NegPar corpus [15] using several performance measures, but some cross-lingual features related to negation scope are still missing.

BiLSTM-based models were also used for negation and speculation scope detection in languages other than English. Fabregat et al. adapted the proposed model by Fencellu [89] to detect negation cues in the Spanish reviews domain [93]. This model used the encoded

tokens and lemmas generated by a pre-trained Spanish word embedding model in addition to the PoS feature. Mahany et al. proposed a BiLSTM model with three dense layers and a single output layer to detect negation scope in Arabic texts [36]. This network takes an input vector of the FastText word embeddings representation and another vector of the cues features to identify the scope of a negation cue within a sentence. The results show that BiLSTM-based models with word embedding representation achieve a remarkable performance in English, Arabic, and Spanish.

Without any handcrafted or highly engineered features, the experimental results of RNNs and CNNs-based models outperform feature-rich SVM and CRF-based models. Nonetheless, the researchers combined RNN/CNN networks and CRF-based models in some studies. Fei et al. used RNN and CRFs, RecurCRFs, as a negation and speculation scope detection framework [96]. At first, the RNN model learns a high-level representation of each sentence and detects negating or speculative words with their scope through the dependency tree structure, where an embedding vector represents each word in the input sentence. After that, a linear chain of CRFs layer takes this representation from RNN as input to decode each label. The authors assumed in this work that the negation and speculation scopes impact each other. This framework was tested on two corpora from separate languages: BioScope [14] and CNeSp [43]. Each covers annotated data from different domains. This framework had a generalization across languages and domains. Therefore, Dalloux et al. designed a cross-domain approach to identify the negation in French and Brazilian Portuguese [13]. In their study, they built word-embedding models for the French language that were composed of French Wikipedia articles and biomedical data. For the Brazilian Portuguese, they used pre-trained models from the NILC website. In addition, various supervision-based models were adapted from Fancellu [89], BiLSTM, BiGRU, and BiLSTM-CRF to detect the negation cue and its scope automatically. The BiLSTM-CRF-based model achieved the best performance in the French and Brazilian Portuguese cross-domain experiments. By contrast, the French cross-domain result for the ESSAI-CAS corpora had a low value for exact matching scope detection, because the CAS corpus was automatically annotated.

In another study, syntactic path-based hybrid architecture combines the BiLSTM and CNN networks [39]. The BiLSTM learns a sentence's context representation by encoding the semantics of each word, while the CNN network captures the syntactic features between the negation cue and its surroundings using dependency parse trees. Finally, the features produced by these two networks are concatenated into a single global feature vector and fed into a binary classifier to decide whether or not this token is part of the negation scope.

Recently, to address the tasks with a lack of resources, transfer of deep learning systems trained on a massive corpus, such as Google's Bidirectional Encoder Representation for Transformers (BERT) [99], has become extensively used to transfer the pretrained architectures. Therefore, transfer learning systems can potentially address problems with various languages and domains' negation and speculation. For example, Khandelwal and Sawant proposed NegBERT [74] to detect negation cue and scope in two stages using the BERT model. Their BERT model depends on multiple corpora from different domains: BioScope, ConanDoyle, and the SFU review corpus. The input to this model is a sequence of tokenized and encoded words with their scores. Then NegBERT generates an output vector for each token, to be fed to a classification layer for the cue and scope detection. This model achieved good performance on the negation detection for unseen corpora from various domains.

Finally, Sergeeva and Zhu [11] proposed a framework based on context-aware representations: global vectors (Glove) [100], embeddings from language models (ELMo) [101], or BERT. Variations of the word embedding vectors and the other features, such as PoS, were fed into two BiLSTM networks to predict the scope of the phenomena. BioScope and the English part of NegPar's corpora, contrasting domains, were used to evaluate the performance of the proposed framework. The results show that the BERT model out-

performs the other word embedding models, proving that it has a superiority effect on syntactic-based tasks.

3.3. Hybrid Approaches

The hybrid approaches integrate empirical models with handcrafted. As an example, Özgür and Radev suggested a hybrid approach to detecting the scope of speculation [75]. First, the speculative cue was identified by linguistic features such as POS, BoW of a cue, and the cue's positional features within a sentence. Then, these feature sets fed an SVM network to detect whether or not the cue is actually speculative. Second, to determine the speculation scope, the authors developed rules according to the syntactic trees.

ScopeFinder is a linguistically motivated system for the detection of negation and speculation scope [102]. It focuses simultaneously on scope detection and cue detection, because the presence of a cue does not suggest the scope in every instance. The proposed rules rely on lexica-syntactic patterns automatically extracted from the BioScope corpus. The authors used the Stanford parser [82] to generate the rules, based on the negating and speculative sentences. As a result, 439 and 1000 generated rules were extracted for the scope of negation and speculation, respectively. This performance evaluation suggests that the same approach will motivate other researchers to apply it to other domains.

Packard et al. revisited Task 1 from SEM 2012* and proposed an approach that relies on the formal representation of propositional semantics [76]. First, the negative sentences are parsed using handcrafted rules to find their minimal recursion semantics (MRS) and identify the negated tokens. Then, if the semantic parser fails to represent a sentence, the system switches to a hybrid model that instead uses syntactic information [85]. The authors in this work were attempting to resolve shortcomings in handling discontinuous scope; however, the issue persists, given that MRS representation is not suitable for all NLP languages.

The problem of negation detection has been addressed by many studies, for many languages, using various techniques, but speculation detection has not been covered to a similar extent. However, the phenomena are semantically and syntactically similar, so most negation detection approaches also promise to address the issue of speculation detection.

4. Applications

Negation and speculation systems have been deployed in various domains, including those of biomedical, product reviews, news, and social media. In this section, we explore the experimental results of these systems and also discuss the impact of negation and speculation on NLP tasks such as SA [103] and IE [104], named entity recognition [105] and machine translation [106]. Besides the F1-score evaluation metric, the percentage of correct scopes (PCS) in some studies is reported to show their systems' performance. The PCS metric is calculated as the number of exactly matched scopes, divided by the number of actual scopes.

4.1. Review Domain

Table 4 summarizes the performance of negation scope detection systems in the review domain. The results show that the BERT-based model [74] outperforms the other proposed systems when evaluated on the SFU review corpus [40]. However, BiLSTM networks are still proving their efficiency in such tasks. By contrast the negation systems in other languages, such as Spanish, still do not perform well in detecting negation scope. Bel-Enguix et al. tested the detection of cross-domain negation on two forms of the Spanish language [17], but doing so reduced the system's performance. Therefore, it is inapplicable to handle negation across domains or language variations.

Table 4. Negation detection systems in the review domain.

| Ref. | Algorithm/Method | Phenomenon | Language | Corpus | F1-Score | PCS |
|------|------------------|--------------|-----------------|--------------------------|----------|-----|
| [44] | CRF | Negation | English | Review | 80 | 39 |
| [88] | BiLSTM | Negation | English | SFU Review | 89 | |
| [74] | NegBERT | Negation | English | SFU Review | 90 | |
| [93] | LSTM | Negation | Spanish | SFU Review _{SP} | 77 | |
| [17] | CRF | Cross-domain | Mexican Spanish | T-MexNeg | 75 | |
| | | Negation | Spanish | SFU Review _{SP} | 68 | |

Negation is of great significance since it directly changes the polarity of a text, especially one expressing an opinion on a product, trend, or service [107,108]. Negation cues affect the contextual polarity of words, yet it does not mean that every sentiment word is inverted. It is considered one of the dominant linguistic methods to change a text's polarity in the SA task [31]. For example, Dadvar et al. [31] investigated the problem of determining the polarity of sentiment in movie reviews when negation words occur. Their approach defined a list of negation cues and studied their effect at several window sizes. As a result, a significant improvement in the classification of the documents was observed. Some research works have addressed the negation problem in SA or opinion retrieval tasks, and most focused on rule-based methods. For example, Polanyi and Zaenen introduced the idea of valence shifters, where a token may strengthen, weaken, or invert the polarity of sentiment as in 'so', 'rarely', and 'not' [109]. Moreover, Jia et al. introduced the identification of the scope of a negation term to determine the affected sentence's clause [110]. However, these proposed rules did not report their impact on detecting the negation cues or scopes. Mukherjee et al. designed a pre-processing algorithm to identify the words with negations and to replace them with negator tags plus their lemmatized forms [2]. By including this negation detection algorithm, the results classification system's performance improved by 1% on average in terms of F1 score. The improvement in performance is limited, as it handles only the explicit form of negation.

Other research works have used a lexicon of common negation cues with supervised learning techniques to reveal the negation scope. Councill et al. [44] described a system that can identify the scope of negation in free English texts. A CRF model was built and then evaluated on an SA task. In contrast to this work [44], which addressed the scope of a negation term within various window sizes, Misra et al. [111] used a binary SVM classifier to identify the true negation cues in the sentence's syntactic structure to detect the scope. Although these systems dramatically improved the F1 score by 20% and 5% on average, again they handle only explicit negation.

Negation in English has been detected by both simple sets of rules and sophisticated approaches; however, it is a language-dependent phenomenon, and the negation models for English cannot be used on text in the Arabic language, for example, as the syntactic structure of negation in Arabic differs from that in English. Few Arabic studies have addressed negation's effect on the SA task, but simple rule-based approaches are clearly insufficient to handle every possible instance of negation in the various Arabic language categories and dialects [112]. The frequency of negation terms in the Arabic Sentiment Analysis (ASA) task has been considered a classification feature, but its effect on the classification process has gone unreported [113,114]. Hamouda and Akaichi assumed that negating terms always invert the polarity of a word preceded by an inverter [115]. By contrast, El-Beltagy concluded in her research that, in addition, the presence of a negation term might change the polarity of the following opinionated word; however, there are still odd cases that may confirm the polarity of the following lexeme [116].

In 2015, Duwairi and Alshboul defined six handcrafted rules to handle negation in the Modern Standard Arabic (MSA) texts in the review domain to enhance the performance of the SA [117]. Even though they addressed the MSA, which follows well-defined rules,

the simplistic approach proved insufficient for a syntactically and morphologically rich language like Arabic. Dialectal Arabic has obstacles such as no syntactic rules and multiple forms of the same word, making ANLP tasks challenging. El-Naggar et al. considered several valences to build a negation-aware classifier for SA in MSA and the Egyptian dialect [118]. Later, Assiri et al. formulated four rules to handle negation in the Saudi dialect [3]. Recently, Kaddoura et al. proposed a system that inverts the polarity of a sentence's clause if a negation term precedes a positive or negative pattern [5]. Regardless of the improvement in performance in these systems' experimental results [3,5,118], none handled the implicit form of negation frequently used in Arabic.

As discussed in Section 3.1, many research studies have addressed the negation problem using rule-based algorithms in various languages [60,78]. For example, Ljajić et al. proposed a lexicon-based approach to handle negation rules in the Serbian language to improve the SA for Twitter data [119]. Jimenez-Zafra et al. studied the effect of negation detection in the Spanish language in the SA task [6]. As in our survey, most negation-aware SA models improved their system's performance in the different language yet could not handle implicit negation.

Other research works improve performance in the SA task by automatically detecting negation and speculation. Cruz et al. proposed a machine learning approach to detect these problems in the review domain in Spanish [32]. The results showed an improvement of 13% and 10% in terms of the F1 score in negation and speculation scope detection, respectively.

4.2. Biomedical Domain

Section 3.1 discussed the rule-based methods used to address the biomedical domain's negation and speculation phenomena. In 2001, the NegFinder algorithm used context-free grammar [66] and the regular expression-based NegEx [71] algorithm to address the negation phenomenon. As seen in Table 5, both proved that negation in medical documents is tractable with a simple algorithm. Due to the simplicity and promising results of the NegEx algorithm, it has been adapted for languages like Spanish [120], Swedish [121], German [61], and French [122]. Although Elazhary has proved that incorporating these cases would improve performance by 4%, NegEx did not handle connecting and multiple negations. Furthermore, the NegEx adaptation for detecting the scope of speculation in German biomedical data achieved poor performance.

ConText extended the NegEx algorithm, addressed the negation scope differently, and added further contextual properties [78]. As a result, it obtained a F1 score value of 93 in identifying the negation scope in six types of clinical reports. The evaluation results prove that the regular expression approach is robust in identifying further contextual properties. ConText has been extended to the Dutch language, contextD, by translating the English trigger terms into Dutch [60].

In other research, DepNeg [71], DEEPEN [10], and NegBio [37] used dependency-based parsers to improve the handling of negation and speculation in biomedical data. For example, on 139 clinical notes from the Mayo clinical dataset DEEPEN decreased the number of false positives cases compared to the NegEx algorithm [123]. In addition, NegBio improved the F1 score in the information extraction task on two datasets, OpenI [124] and ChestX-ray [37], by 64% and 12%, respectively.

Table 5. Negation and speculation detection systems for the biomedical domain.

| Ref. | Algorithm/Method | Phenomenon | Language | Corpus | F1-Score | PCS |
|------|--------------------|---------------------------|-----------------------------------|--|---------------------|-----|
| [66] | NegFinder | Negation | English | Surgical notes and discharge summaries | 96 | |
| [71] | NegEx | Negation | English | Discharge summaries | 90 | |
| [78] | ConText | Negation | English | Clinical reports | 93 | |
| [79] | NegMiner | Negation | English | Narrative Medical Documents | 95 | |
| [10] | DEEPEN | Negation | English | Clinical notes | 81 | |
| [37] | NegBio | Negation Speculation | English | OpenI [124] ChestX-ray [37] | 87 * 94 * | |
| [83] | SVM | Speculation | English | Papers from FlyBase | 76 PEP ¹ | |
| [84] | SVM | Speculation | English | Papers from BMC | 85 PEP | |
| [90] | CNN with attention | Speculation | English | CoNLL 2010 | 85 | |
| [60] | contextD | Negation | Dutch | Clinical reports | 87–93 | |
| [46] | CRF | Negation | Chinese | EMR | 95 | |
| [95] | GAN with attention | Negation Speculation | Arabic | BioArabic | 79 79 | |
| [61] | NegEx | Negation Speculation | German | GNSC | 94 42 | |
| [24] | BiLSTM-CRF | Negation Speculation | French | ESSAI-CAS | 90 86 | |
| [13] | BiLSTM-CRF | Cross-lingual Negation | French Brazilian Portuguese | ESSAI-CASREBEC | 53 73 | |

¹ Precision/recall break-even point (PEP). * Information extraction system's performance.

Table 5 shows the evaluation results of the other supervised learning techniques. As in the review domain, BiLSTM and BiLSTM-CRF models obtained remarkable results. The results in this work [13] also further prove that attempting negation across languages leads to poor performance.

The BioScope corpus has been used to evaluate various rule- and supervised-based models. Table 6 presents the F1 score for these models, where RecurCRFs is the best performing model [96]. This model has a generalization ability across languages and domains, because it performed well on the Chinese CNeSp corpus [43].

Table 6. Detection of negation and speculation scope systems for the BioScope corpus.

| Ref. | Algorithm/Method | Negation | | Speculation | |
|-------|------------------|----------|-----|-------------|-----|
| | | F1-Score | PCS | F1-Score | PCS |
| [72] | SVM-CRF-TiMBL | 80 | 60 | | |
| [73] | SVM-CRF-TiMBL | | | 59 | 31 |
| [75] | SVM | | | 71 | |
| [44] | CRF | 75 | 53 | | |
| [86] | SVM Tree Kernel | 84 | 53 | 94 | 71 |
| [87] | CNN | 89 | 74 | 91 | 74 |
| [92] | Semi-CRF | 90 | 78 | | |
| [74] | NegBERT | 93 | | | |
| [11] | BiLSTM | | 90 | | 86 |
| [96] | RNN-CRF | 93 | 92 | 91 | 88 |
| [39] | BiLSTM & CNN | 90 | 75 | | |
| [71] | NegEx | 82 | | | |
| [37] | NegBio | 95 | | | |
| [102] | ScopeFinder | 75 | | 76 | |

Negation- and speculation-aware models have been included in the pipeline of biomedical NLP tasks, where the biomedical data commonly include negating and speculative instances. For example, cancer diagnosis is a sensitive task requiring highly accurate systems. Gkotsis et al. assessed the classification task of identifying patients expected to be at high risk of suicide with negation detection [125]. In their study, if a negation cue was found in the syntactic tree of an input text, a rule-based algorithm detected its scope. As a result, their approach outperformed the NegEx algorithm and achieved a recall metric of 95%. Another study used a hybrid approach that combines deep learning techniques and rule-based methods to diagnose lung cancer in Spanish clinical narratives [12]. First, the BiLST-CRF model extracts the named entities; next, handcrafted rules detect the negating and speculative scope, and finally, it reports the lung cancer diagnosis. The proposed negation and speculation rule-based approach improved the task of relating cancer diagnosis to a correct date by 18% and achieved an 89 F1 score. These evaluation results prove that clinical notes are sensitive to negation and speculation phenomena.

4.3. Others

The SEM 2012 Shared Task* was announced at the first joint conference on lexical and computational semantics to motivate the research community to solve the scope and focus of negation [58]. This competition relied on the corpus of journal stories, PropBank FOC [51], and the corpus of short stories Conan Doyle [41]. In 2012, the best performing system was the SVM-based UiO1 system. Later, other research groups resolved this task by other approaches. As listed in Table 7, NegBERT obtained the highest F1 score, but it did not mention the PCS metric. For the other languages, Arabic and Chinese, the RNN-based models showed good performance.

Further NLP projects have studied the effect of negation and speculation, as in Hossain et al. [106], which looked at tackling phenomena of negation in MT systems. Negation-related translation issues include the omission and reversal of negation that contradicts the translated sentence, the incorrect scope of negation that negates the wrong token, and the mistranslation of the negation object. These issues vary in severity, but each may completely change the meaning of the translated sentence. The authors targeted the translation tasks in WMT18 and WMT19 competitions that translate news both from and to English and other languages. Using the BiLSTM-CRF (<https://github.com/mosharafhossain/negation-mt> (accessed on 28 March 2022)) model based on the SEM 2012* corpus, an F1 score of 92 was achieved when detecting the negation cue in English. The evaluation results showed that the presence of negation reduces the system's performance, in some cases by 60%.

In recent years, news articles have become rapidly shared and consumed; thus, there is a need for systems to extract valuable and accurate information. In one work [126], uncertainty detection in the newswire domain was addressed as a binary classification task. The authors proposed a hybrid machine learning approach with semantic, syntactic features to classify an event as certain or uncertain. For the uncertainty identification task, this approach achieved a 71 F1-score.

Table 7. Negation and speculation detection systems for other domains.

| Ref. | Algorithm/Method | Phenomenon | Language | Corpus | F1-Score | PCS |
|------|------------------|---------------------------|-----------------|-------------|----------|-----|
| [76] | MRS Crawler | Negation | English | Conan Doyle | | 75 |
| [85] | SVM | Negation | English | SEM 2012* | 78 | |
| [89] | BiLSTM | Negation | English | SEM 2012* | 89 | |
| [91] | SVM+CRF | Negation | English | SEM 2012* | 71 | |
| [97] | STRNN | Negation | English | SEM 2012* | 89 | |
| [74] | NegBERT | Negation | English | SEM 2012* | 92 | |
| [92] | Semi-CRF | Negation | English | SEM 2012* | 88 | 82 |
| | | | Chinese | ChNeSp | 90 | 72 |
| [96] | RNN-CRF | Negation | Chinese | CNeSp | 93 | 92 |
| | | Speculation | | | 91 | 88 |
| [11] | BiLSTM | Negation | English-Chinese | NegPar | | 79 |
| [94] | BiLSTM | Cross-Lingual Negation | English-Chinese | NegPar | 72 | 24 |
| [36] | BiLSTM | Negation | Arabic | ArNe | 89 | |

5. Discussion

Corpora annotated with negation and speculation vary in language, domain, size, and annotated span level (cue, scope, event, and focus). Cue is the most important element, as it influences the other annotated elements, and without it, negation or speculation cannot be handled. Negation has four tasks: cue detection, scope identification, event extraction, and focus identification. Almost all corpora annotated with negation include an annotation for a negation cue and its scope; however, some corpora do not include any negation cue, as in English [21,51], Dutch [60], and Hungarian [29]. Event element is annotated mainly for the biomedical domain, and recently has been included in further domains, such as reviews [26] and news [35] in the Spanish language. Corpora annotated with focus are scarce, because annotating a corpus appropriately involves great attention. However, it is necessary to review the annotation guidelines to eliminate inconsistent annotations, even in the same language and in the same domain.

Corpora annotated with negation were mainly built to develop negation-aware systems. At first, the research community focused on the treatment of negation in clinical texts using the rule-based algorithms. The NegEx algorithm was widely used to detect negation especially in clinical texts, since the nature of any text in this domain is syntactically restricted. The BioScope corpus is an example [14]. Despite its simplistic form, the algorithm achieved remarkable results on a variety of clinical notes. Dependency-based parser tools did not achieve the same desired performance compared to their complexities. Nonetheless, NegBio was the first rule-based algorithm to handle the speculative instances in the biomedical domain. Most of these algorithms were built to handle negation phenomena for clinical reports; however, other research studies have handcrafted the rules to the task of ASA [3,117]. Despite NegEx being language-dependent and seemingly a relatively basic tool with which to identify negating events/text, it is considered the better rule-based algorithm compared to more complex ones [10,37].

In the biomedical domain, rule-based algorithms succeeded in identifying the scope of negation, yet they were not used in other domains, such as review data or general text data. This is for several reasons: their text may include spelling mistakes and implicit negation or speculation, plus it is freestyle writing that does not follow grammatical rules. As a result, in various domains supervised-based models were proposed to address the problem of negation. These models typically either use external resources to extract complex syntax and grammar features or are based on neural architectures, such as LSTM and CNNs, to extract the semantic representations and capture the hidden features automatically. Although RNNs, CNNs, and CRF-based systems with the data's syntactic structure achieved better performance than traditional machine learning techniques, their lack of annotated data limited the progress in treating negation. Possible solutions to redress the imbalance for low-resource languages include merging multiple corpora of the same language, translating the well-known English corpora such as BioScope [14] and SFU Review [40], into other languages, or transferring the models cross-lingually. When transfers between Spanish and Mexican Spanish [17] and between French and Brazilian Portuguese [13] were evaluated, this approach seemed to have resolved the problem somewhat.

As with corpora for negation, speculation corpora include cue, scope, and events elements, but they are less available. Most were annotated for the clinical and biomedical domain, and it is rare to find a corpus annotated for speculation in other domains, such as for English reviews [40] and Chinese literature or financial articles [43]. A trial to handle the detection of speculative scope uses a modified version of NegEx on German biomedical data, but it records low performance [61]. As seen in our survey, there are no rule-based algorithms to have been developed to handle the problem of speculation. However, the supervised learning techniques that have been used for the detection of negation also achieved good results on the problem of speculation [86,96].

The previous studies addressed the negation and speculation phenomena utilizing a variety of methodologies. However, there are still certain challenges that have not been addressed. Among the challenges are the ambiguity of cues and instances in which a simple cue is part of a discontinuous cue or scope [35,36]. In addition, poor handling of discontinuous scopes and implicit negations and speculations have not been addressed [5]. In order to determine the patterns of the nontrivial cases of negation and speculation, additional research is required. Taking into account more sophisticated and syntactic features may be useful for detecting the discontinuous and implicit negation and speculation. Additionally, recent embedding techniques, like BERT [99] and ELMo [101], may provide a more precise representation of the negating and speculative sentences. Some research teams adopted transfer learning-based systems with pre-trained models, achieving better results than with other supervised learning techniques [74].

Section 4 discusses the importance in multiple NLP tasks of determining both negating and speculative clauses within a sentence. For example, it is crucial to detect speculation in the task of SA to differentiate between the objective and subjective content, yet insufficient studies have addressed this issue. Furthermore, any inappropriate handling of negation leads to the SA task generating false results. In the biomedical domain, spotting negation and speculation can save people's lives by identifying those patients at high risk of suicide [125]. Negation and speculation detection systems should be incorporated into the pipeline of NLP tasks, including SA and IE in the biomedical domain.

6. Conclusions

In this review we present the corpora annotated with negation and speculation for use in handling these linguistic phenomena in NLP tasks. In addition to that, this survey reviewed the different approaches that have been applied in a variety of languages and domains. Negation- and speculation-aware systems for many languages have limitations, ranging from the availability of corpora to inconsistent annotation guidelines. Some take a sophisticated approach, such as transfer learning techniques. The complex syntactic structures of negation and speculation make it hard to build a general system to handle

them across multiple domains, and, although they need further enhancement, cross-lingual systems are an excellent way to start addressing the problem. Therefore, experimentation with languages close to each other, such as Arabic and Spanish, is an eligible approach. Moreover, we suggest making use of more syntactic features to correctly detect discontinuous scopes, implicit negations, and speculations, as well as to find a solution to the problem of ambiguity in cues. The performance of NLP tasks was shown to be improved as a result of incorporating such phenomena into them.

Author Contributions: Conceptualization, A.M., N.A. and S.G.; methodology, A.M., H.K. and N.S.E.; validation, A.M., H.K., N.S.E. and N.A.; investigation, A.M. and H.K.; resources, A.M., N.S.E. and N.A.; writing—original draft preparation, A.M.; writing—review and editing, A.M., H.K., N.S.E., N.A. and S.G.; supervision, H.K. and S.G.; project administration, H.K. and S.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Belkahla Driss, O.; Mellouli, S.; Trabelsi, Z. From Citizens to Government Policy-Makers: Social Media Data Analysis. *Gov. Inform. Q.* **2019**, *36*, 560–570. [\[CrossRef\]](#)
2. Mukherjee, P.; Badr, Y.; Doppalapudi, S.; Srinivasan, S.M.; Sangwan, R.S.; Sharma, R. Effect of Negation in Sentences on Sentiment Analysis and Polarity Detection. *Procedia Comput. Sci.* **2021**, *185*, 370–379. [\[CrossRef\]](#)
3. Assiri, A.; Emam, A.; Al-Dossari, H. Towards Enhancement of a Lexicon-Based Approach for Saudi Dialect Sentiment Analysis. *J. Inform. Sci.* **2018**, *44*, 184–202. [\[CrossRef\]](#)
4. Alomari, E.; Katib, I.; Albeshri, A.; Yigitcanlar, T.; Mehmood, R. Iktishaf+: A Big Data Tool with Automatic Labeling for Road Traffic Social Sensing and Event Detection Using Distributed Machine Learning. *Sensors* **2021**, *21*, 2993. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Kaddoura, S.; Itani, M.; Roast, C. Analyzing the Effect of Negation in Sentiment Polarity of Facebook Dialectal Arabic Text. *Appl. Sci.* **2021**, *11*, 4768. [\[CrossRef\]](#)
6. Jimenez Zafra, S.M.; Martin Valdivia, M.T.; Martinez Camara, E.; Urena Lopez, L.A. Studying the Scope of Negation for Spanish Sentiment Analysis on Twitter. *IEEE Trans. Affect. Comput.* **2019**, *10*, 129–141. [\[CrossRef\]](#)
7. Konstantinova, N.; De Sousa, S.C.M. Annotating Negation and Speculation: The Case of the Review Domain. In Proceedings of the Second Student Research Workshop Associated with the International Conference on Recent Advances in Natural Language Processing (RANLP), Hissar, Bulgaria, 12–14 September 2011; pp. 139–144.
8. Reitan, J.; Faret, J.; Gambäck, B.; Bungum, L. Negation Scope Detection for Twitter Sentiment Analysis. In Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, Stroudsburg, PA, USA, 17 September 2015; pp. 99–108.
9. Marimon, M.; Vivaldi, J.; Bel, N. Annotation of Negation in the IULA Spanish Clinical Record Corpus. In *Proceedings of the Workshop Computational Semantics Beyond Events and Roles*; Association for Computational Linguistics (ACL): Stroudsburg, PA, USA, 2017; pp. 43–52.
10. Mehrabi, S.; Krishnan, A.; Sohn, S.; Roch, A.M.; Schmidt, H.; Kesterson, J.; Beesley, C.; Dexter, P.; Max Schmidt, C.; Liu, H.; et al. DEEPEN: A Negation Detection System for Clinical Text Incorporating Dependency Relation into NegEx. *J. Biomed. Inform.* **2015**, *54*, 213–219. [\[CrossRef\]](#)
11. Sergeeva, E.; Zhu, H.; Tahmasebi, A.; Szolovits, P. Neural Token Representations and Negation and Speculation Scope Detection in Biomedical and General Domain Text. In *Proceedings of the Tenth International Workshop on Health Text Mining and Information Analysis (LOUHI 2019)*; Association for Computational Linguistics (ACL): Stroudsburg, PA, USA, 2019; pp. 178–187.
12. Solarte Pabón, O.; Torrente, M.; Provencio, M.; Rodríguez-Gonzalez, A.; Menasalvas, E. Integrating Speculation Detection and Deep Learning to Extract Lung Cancer Diagnosis from Clinical Notes. *Appl. Sci.* **2021**, *11*, 865. [\[CrossRef\]](#)
13. Dalloux, C.; Claveau, V.; Grabar, N.; Oliveira, L.E.S.; Moro, C.M.C.; Gumiel, Y.B.; Carvalho, D.R. Supervised Learning for the Detection of Negation and of Its Scope in French and Brazilian Portuguese Biomedical Corpora. *Nat. Lang. Eng.* **2021**, *27*, 181–201. [\[CrossRef\]](#)
14. Vincze, V.; Szarvas, G.; Farkas, R.; Móra, G.; Csirik, J. The BioScope Corpus: Biomedical Texts Annotated for Uncertainty, Negation and Their Scopes. *BMC Bioinform.* **2008**, *9* (Suppl. S11), S9. [\[CrossRef\]](#)

15. Liu, Q.; Fancellu, F.; Webber, B. NegPar: A Parallel Corpus Annotated for Negation. In *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018)*; European Language Resources Association (ELRA): Miyazaki, Japan, 2018; pp. 3464–3472.
16. Jiménez-Zafra, S.M.; Morante, R.; Martín, M.; Ureña-López, L.A. A Review of Spanish Corpora Annotated with Negation. In *Proceedings of the 27th International Conference on Computational Linguistics*, Santa Fe, NM, USA, 20–26 August 2018; pp. 915–924.
17. Bel-Enguix, G.; Gómez-Adorno, H.; Pimentel, A.; Ojeda-Trueba, S.-L.; Aguilar-Vizuet, B. Negation Detection on Mexican Spanish Tweets: The T-MexNeg Corpus. *Appl. Sci.* **2021**, *11*, 3880. [\[CrossRef\]](#)
18. Lima, S.; Perez, N.; Cuadros, M.; Rigau, G. NUBES: A Corpus of Negation and Uncertainty in Spanish Clinical Texts. In *Proceedings of the LREC 2020—12th International Conference on Language Resources and Evaluation*; European Language Resources Association (ELRA): Marseille, France, 2020; pp. 5772–5781.
19. Al-Khawaldeh, F.T. Speculation and Negation Annotation for Arabic Biomedical Texts: BioArabic Corpus. *World Comput. Sci. Inform. Technol. J. (WCSIT)* **2016**, *6*, 8–11.
20. Dalianis, H.; Velupillai, S. How Certain Are Clinical Assessments? Annotating Swedish Clinical Text for (Un) Certainties, Speculations and Negations. In *Proceedings of the 7th International Conference on Language Resources and Evaluation, LREC 2010*; European Language Resources Association (ELRA): Valletta, Malta, 2010; pp. 3071–3075.
21. Kim, J.-D.; Ohta, T.; Tsujii, J. Corpus Annotation for Mining Biomedical Events from Literature. *BMC Bioinform.* **2008**, *9*, 10. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Greenberg, J.H. *Universals of Human Language*; Stanford University Press: Stanford, UK, 1978; p. 4. [\[CrossRef\]](#)
23. Oxford Learner's Dictionaries. Available online: <https://www.oxfordlearnersdictionaries.com/> (accessed on 2 October 2021).
24. Dalloux, C.; Claveau, V.; Grabar, N. Speculation and Negation Detection in French Biomedical Corpora. In *Proceedings of the Recent Advances in Natural Language Processing*, Varna, Bulgaria, 2–4 September 2019; pp. 223–232.
25. Jiménez-Zafra, S.M.; Morante, R.; Blanco, E.; Martín-Valdivia, M.T.; Alfonso Ureña-López, L. Detecting Negation Cues and Scopes in Spanish. In *Proceedings of the LREC 2020—12th International Conference on Language Resources and Evaluation*; European Language Resources Association (ELRA): Marseille, France, 2020; pp. 6902–6911.
26. Jiménez-Zafra, S.M.; Taulé, M.; Martín-Valdivia, M.T.; Ureña-López, L.A.; Martí, M.A. SFU ReviewSP-NEG: A Spanish Corpus Annotated with Negation for Sentiment Analysis. A Typology of Negation Patterns. *Lang. Resour. Eval.* **2018**, *52*, 533–569. [\[CrossRef\]](#)
27. Banjade, R.; Rus, V. DT-Neg: Tutorial Dialogues Annotated for Negation Scope and Focus in Context. In *Proceedings of the Proceedings of the 10th International Conference on Language Resources and Evaluation, LREC 2016*; European Language Resources Association (ELRA): Portorož, Slovenia, 2016; pp. 3768–3771.
28. Altuna, B.; Minard, A.-L.; Speranza, M. The Scope and Focus of Negation: A Complete Annotation Framework for Italian. In *Proceedings of the Workshop Computational Semantics Beyond Events and Roles*; Association for Computational Linguistics (ACL): Valencia, Spain, 2017; pp. 34–42.
29. Vincze, V. Uncertainty Detection in Hungarian Texts. In *Proceedings of COLING 2014: Technical Papers, Proceedings of the COLING 2014–25th International Conference on Computational Linguistics*, Dublin, Ireland, 23–29 August 2014; University and Association for Computational Linguistics: Dublin, Ireland, 2014; pp. 844–1853.
30. Ren, Y.; Fei, H.; Peng, Q. Detecting the Scope of Negation and Speculation in Biomedical Texts by Using Recursive Neural Network. In *Proceedings of the 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Madrid, Spain, 3–6 December 2018; pp. 739–742.
31. Dadvar, M.; Hauff, C.; De Jong, F. Scope of Negation Detection in Sentiment Analysis. In *Proceedings of the Dutch-Belgian Information Retrieval Workshop (DIR 2011)*, Amsterdam, The Netherlands, 4 March 2011; pp. 16–20.
32. Cruz, N.P.; Taboada, M.; Mitkov, R. A Machine-Learning Approach to Negation and Speculation Detection for Sentiment Analysis. *J. Assoc. Inform. Sci. Technol.* **2016**, *67*, 2118–2136. [\[CrossRef\]](#)
33. Baker, K.; Bloodgood, M.; Dorr, B.J.; Callison-Burch, C.; Filardo, N.W.; Piatko, C.; Levin, L.; Miller, S. Modality and Negation in SIMT Use of Modality and Negation in Semantically-Informed Syntactic MT. *Comput. Linguist.* **2012**, *38*, 411–438. [\[CrossRef\]](#)
34. Savova, G.K.; Masanz, J.J.; Ogren, P.V.; Zheng, J.; Sohn, S.; Kipper-Schuler, K.C.; Chute, C.G. Mayo Clinical Text Analysis and Knowledge Extraction System (CTAKES): Architecture, Component Evaluation and Applications. *J. Am. Med. Inform. Assoc.* **2010**, *17*, 507–513. [\[CrossRef\]](#)
35. Taulé, M.; Nofre, M.; González, M.; Martí, M.A. Focus of Negation: Its Identification in Spanish. *Nat. Lang. Eng.* **2021**, *27*, 131–152. [\[CrossRef\]](#)
36. Mahany, A.; Fouad, M.M.; Aloraini, A.; Khaled, H.; Nawaz, R.; Aljohani, N.R.; Ghoniemy, S. Supervised Learning for Negation Scope Detection in Arabic Texts. In *Proceedings of the Tenth International Conference on Intelligent Computing and Information Systems (ICICIS)*, Cairo, Egypt, 5 December 2021; pp. 177–182.
37. Peng, Y.; Wang, X.; Lu, L.; Bagheri, M.; Summers, R.; Lu, Z. NegBio: A High-Performance Tool for Negation and Uncertainty Detection in Radiology Reports. *AMIA Joint Summits on Translational Science proceedings. AMIA Jt. Summits Transl. Sci.* **2017**, *2017*, 188–196. [\[CrossRef\]](#)

38. Vincze, V. Weasels, Hedges and Peacocks: Discourse-Level Uncertainty in Wikipedia Articles. In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*; Asian Federation of Natural Language Processing: Nagoya, Japan, 2013; pp. 383–391.
39. Lazib, L.; Qin, B.; Zhao, Y.; Zhang, W.; Liu, T. A Syntactic Path-Based Hybrid Neural Network for Negation Scope Detection. *Front. Comput. Sci.* **2020**, *14*, 84–94. [\[CrossRef\]](#)
40. Konstantinova, N.; De Sousa, S.C.M.; Cruz, N.P.; Maña, M.J.; Taboada, M.; Mitkov, R. A Review Corpus Annotated for Negation, Speculation and Their Scope. In *Proceedings of the 8th International Conference on Language Resources and Evaluation, LREC 2012*; European Language Resources Association (ELRA): Istanbul, Turkey, 2012; pp. 3190–3195.
41. Morante, R.; Daelemans, W. ConanDoyle-Neg: Annotation of Negation Cues and Their Scope in Conan Doyle Stories. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*; European Language Resources Association (ELRA): Istanbul, Turkey, May, 2012; pp. 1563–1568.
42. Jiménez-Zafra, S.M.; Díaz, N.P.C.; Morante, R.; Martín-Valdivia, M.T. NEGes 2019 Task: Negation in Spanish. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2019)*, CEUR Workshop, Bilbao, Spain, 24 September 2019; pp. 329–341.
43. Zou, B.; Zhu, Q.; Zhou, G. Negation and Speculation Identification in Chinese Language. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*; Association for Computational Linguistics (ACL): Stroudsburg, PA, USA, 2015; pp. 656–665.
44. Councill, I.G.; McDonald, R.; Velikovich, L. What's Great and What's Not: Learning to Classify the Scope of Negation for Improved Sentiment Analysis. In *Proceedings of the ACL Workshop on Negation and Speculation in Natural Language Processing*, Uppsala, Sweden, 10 July 2010; pp. 51–59.
45. Kolhatkar, V.; Wu, H.; Cavasso, L.; Francis, E.; Shukla, K.; Taboada, M. The SFU Opinion and Comments Corpus: A Corpus for the Analysis of Online News Comments. *Corpus Pragmat.* **2020**, *4*, 155–190. [\[CrossRef\]](#) [\[PubMed\]](#)
46. Kang, T.; Zhang, S.; Xu, N.; Wen, D.; Zhang, X.; Lei, J. Detecting Negation and Scope in Chinese Clinical Notes Using Character and Word Embedding. *Comput. Methods Prog. Biomed.* **2017**, *140*, 53–59. [\[CrossRef\]](#) [\[PubMed\]](#)
47. Cohen, J. A Coefficient of Agreement for Nominal Scales. *Educ. Psychol. Measur.* **1960**, *20*, 37–46. [\[CrossRef\]](#)
48. Fleiss, J.L. Measuring Nominal Scale Agreement among Many Raters. *Psychol. Bull.* **1971**, *76*, 378–382. [\[CrossRef\]](#)
49. Pyysalo, S.; Ginter, F.; Heimonen, J.; Björne, J.; Boberg, J.; Järvinen, J.; Salakoski, T. BioInfer: A Corpus for Information Extraction in the Biomedical Domain. *BMC Bioinform.* **2007**, *8*, 50. [\[CrossRef\]](#)
50. Farkas, R.; Vincze, V.; Móra, G.; Csirik, J.; Szarvas, G. The CoNLL-2010 Shared Task: Learning to Detect Hedges and Their Scope in Natural Language Text. In *Proceedings of the CoNLL 2010—14th Conference on Computational Natural Language Learning: Shared Task*, Uppsala, Sweden, 15–16 July 2010; ACL Anthology: New York, NY, USA, 2010; pp. 1–12.
51. Blanco, E.; Moldovan, D. Semantic Representation of Negation Using Focus Detection. In *Proceedings of the ACL-HLT 2011—49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*; Association for Computational Linguistics (ACL): Portland, OR, USA, 2011; pp. 581–589.
52. Ohta, T.; Tateisi, Y.; Kim, J.D. The GENIA Corpus: An Annotated Research Abstract Corpus in Molecular Biology Domain. In *Proceedings of the Second International Conference on Human Language Technology Research*, San Diego, CA, USA, 24–27 March 2002; pp. 82–86.
53. Kim, J.-D.; Ohta, T.; Tateisi, Y.; Tsujii, J. GENIA Corpus-A Semantically Annotated Corpus for Bio-Textmining. *Bioinformatics* **2003**, *19*, i180–i182. [\[CrossRef\]](#)
54. Morante, R.; Schrauwen, S.; Daelemans, W. Corpus-Based Approaches to Processing the Scope of Negation Cues: An Evaluation of the State of the Art. In *Proceedings of the 9th International Conference on Computational Semantics, IWCS 2011*; Association for Computational Linguistics (ACL): Stroudsburg, PA, USA, 2011; pp. 350–354.
55. Morante, R.; Schrauwen, S.; Daelemans, W. *Annotation of Negation Cues and Their Scope Guidelines v1.0*; University of Antwerp: Antwerp, Belgium, 2011.
56. Palmer, M.; Gildea, D.; Kingsbury, P. The Proposition Bank: An Annotated Corpus of Semantic Roles. *Comput. Linguist.* **2005**, *31*, 71–106. [\[CrossRef\]](#)
57. Taboada, M.; Anthony, C.; Voll, K. Methods for Creating Semantic Orientation Dictionaries. In *Proceedings of the 5th International Conference on Language Resources and Evaluation, LREC 2006*; European Language Resources Association (ELRA): Genoa, Italy, 2006; pp. 427–432.
58. Morante, R.; Blanco, E. *SEM 2012 Shared Task: Resolving the Scope and Focus of Negation. In *Proceedings of the SEM 2012—1st Joint Conference on Lexical and Computational Semantics*; Association for Computational Linguistics (ACL): Montréal, ON, Canada, 2012; pp. 265–274.
59. Matsuyoshi, S.; Otsuki, R.; Fukumoto, F. Annotating the Focus of Negation in Japanese Text. In *Proceedings of the 9th International Conference on Language Resources and Evaluation, LREC 2014*; European Language Resources Association (ELRA): Reykjavik, Iceland, 2014; pp. 1743–1750.
60. Afzal, Z.; Pons, E.; Kang, N.; Sturkenboom, M.C.J.M.; Schuemie, M.J.; Kors, J.A. ContextD: An Algorithm to Identify Contextual Properties of Medical Terms in a Dutch Clinical Corpus. *BMC Bioinform.* **2014**, *15*, 373. [\[CrossRef\]](#)
61. Cotik, V.; Roller, R.; Xu, F.; Uszkoreit, H.; Budde, K.; Schmidt, D. Negation Detection in Clinical Reports Written in German. In *Proceedings of the 5th Workshop on Building and Evaluating Resources for Biomedical Text Mining (BioTxtM 2016)*; The COLING 2016 Organizing Committee: Osaka, Japan, 2016; pp. 115–124.

62. Cruz, N.; Morante, R.; Maña López, M.J.; Mata Vázquez, J.; Parra Calderón, C.L. Annotating Negation in Spanish Clinical Texts. In *Proceedings of the Workshop Computational Semantics Beyond Events and Roles*; Association for Computational Linguistics (ACL): Stroudsburg, PA, USA, 2017; pp. 53–58.
63. Dalianis, H.; Hassel, M.; Velupillai, S. The Stockholm EPR Corpus: Characteristics and Some Initial Findings. In *Proceedings of the 14th International Symposium for Health Information Management Research (ISHIMR 2009)*, Kalmar, Sweden, 14–16 October 2009; pp. 243–249.
64. Szarvas, G.; Vincze, V.; Farkas, R.; Móra, G.; Gurevych, I. Cross-Genre and Cross-Domain Detection of Semantic Uncertainty. *Comput. Linguist.* **2012**, *38*, 335–367. [[CrossRef](#)]
65. Roller, R.; Uszkoreit, H.; Xu, F.; Seiffe, L.; Mikhailov, M.; Staeck, O.; Budde, K.; Halleck, F.; Schmidt, D. A Fine-Grained Corpus Annotation Schema of German Nephrology Records. In *Proceedings of the Clinical Natural Language Processing Workshop*; Clinical {NLP}: Osaka, Japan, 2016; pp. 69–77.
66. Mutalik, P.G.; Deshpande, A.; Nadkarni, P.M. Use of General-Purpose Negation Detection to Augment Concept Indexing of Medical Documents. *J. Am. Med. Inform. Assoc.* **2001**, *8*, 598–609. [[CrossRef](#)]
67. Styler, W.F.; Bethard, S.; Finan, S.; Palmer, M.; Pradhan, S.; de Groen, P.C.; Erickson, B.; Miller, T.; Lin, C.; Savova, G.; et al. Temporal Annotation in the Clinical Domain. *Trans. Assoc. Comput. Linguist.* **2014**, *2*, 143–154. [[CrossRef](#)] [[PubMed](#)]
68. Minard, A.; Marchetti, A.; Speranza, M. Event Factuality in Italian: Annotation of News Stories from the Ita-TimeBank. In *Proceedings of CLiC-it 2014, First Italian Conference on Computational Linguistic*, Pisa, Italy, 9–11 December 2014; pp. 260–264.
69. Minard, A.-L.; Speranza, M.; Caselli, T. The EVALITA 2016 Event Factuality Annotation Task (FactA). In *EVALITA. Evaluation of NLP and Speech Tools for Italian*; Accademia University Press: Torino, Italy, 2016; pp. 32–39.
70. Grabar, N.; Claveau, V.; Dalloux, C. CAS: French Corpus with Clinical Cases. In *Proceedings of the Ninth International Workshop on Health Text Mining and Information Analysis*; Association for Computational Linguistics (ACL): Stroudsburg, PA, USA, 2018; pp. 122–128.
71. Chapman, W.W.; Bridewell, W.; Hanbury, P.; Cooper, G.F.; Buchanan, B.G. A Simple Algorithm for Identifying Negated Findings and Diseases in Discharge Summaries. *J. Biomed. Inform.* **2001**, *34*, 301–310. [[CrossRef](#)] [[PubMed](#)]
72. Morante, R.; Daelemans, W. A Metalearning Approach to Processing the Scope of Negation. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL '09)*; Association for Computational Linguistics (ACL): Morristown, NJ, USA, 2009; pp. 21–29.
73. Morante, R.; Daelemans, W. Learning the Scope of Hedge Cues in Biomedical Texts. In *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing*, Boulder, CO, USA, 5 June 2009; pp. 28–36.
74. Khandelwal, A.; Sawant, S. NegBERT: A Transfer Learning Approach for Negation Detection and Scope Resolution. In *Proceedings of the LREC 2020—12th International Conference on Language Resources and Evaluation*, Marseille, France, 11 November 2019; pp. 5739–5748.
75. Özgür, A.; Radev, D.R. Detecting Speculations and Their Scopes in Scientific Text. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing Volume 3—EMNLP '09*; Association for Computational Linguistics: Morristown, NJ, USA, 2009; pp. 1398–1407.
76. Packard, W.; Bender, E.M.; Read, J.; Oepen, S.; Drīdan, R. Simple Negation Scope Resolution through Deep Parsing: A Semantic Solution to a Semantic Problem. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*; Association for Computational Linguistics (ACL): Stroudsburg, PA, USA, 2014; pp. 69–78.
77. Aronow, D.B.; Fangfang, F.; Croft, W.B. Ad Hoc Classification of Radiology Reports. *J. Am. Med. Inform. Assoc.* **1999**, *6*, 393–411. [[CrossRef](#)] [[PubMed](#)]
78. Harkema, H.; Dowling, J.N.; Thornblade, T.; Chapman, W.W. ConText: An Algorithm for Determining Negation, Experiencer, and Temporal Status from Clinical Reports. *J. Biomed. Inform.* **2009**, *42*, 839–851. [[CrossRef](#)]
79. Elazhary, H. NegMiner: An Automated Tool for Mining Negations from Electronic Narrative Medical Documents. *Int. J. Intell. Syst. Appl.* **2017**, *9*, 14–22. [[CrossRef](#)]
80. Sohn, S.; Wu, S.; Chute, C.G. Dependency Parser-Based Negation Detection in Clinical Narratives. In *Proceedings of the AMIA Joint Summits on Translational Science*; American Medical Informatics Association: Bethesda, MD, USA, 2012; pp. 1–8.
81. Uzuner, Ö.; South, B.R.; Shen, S.; DuVall, S.L. 2010 I2b2/VA Challenge on Concepts, Assertions, and Relations in Clinical Text. *J. Am. Med. Inform. Assoc. JAMIA* **2011**, *18*, 552–556. [[CrossRef](#)]
82. De Marneffe, M.C.; MacCartney, B.; Manning, C.D. Generating Typed Dependency Parses from Phrase Structure Parses. In *Proceedings of the 5th International Conference on Language Resources and Evaluation, LREC 2006*; European Language Resources Association (ELRA): Genoa, Italy, 2006; pp. 449–454.
83. Medlock, B.; Briscoe, T. Weakly Supervised Learning for Hedge Classification in Scientific Literature. In *Proceedings of the ACL 2007—45th Annual Meeting of the Association for Computational Linguistics*; Association for Computational Linguistics (ACL): Prague, Czech Republic, 2007; pp. 992–999.
84. Szarvas, G. Hedge Classification in Biomedical Texts with a Weakly Supervised Selection of Keywords. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-08: HLT)*, Columbus, OH, USA, 20 June 2008; pp. 281–289.

85. Read, J.; Velldal, E.; Øvrelid, L.; Oepen, S. UiO1: Constituent-Based Discriminative Ranking for Negation Resolution. In *Proceedings of the SEM 2012—1st Joint Conference on Lexical and Computational Semantics*, Montreal, QC, Canada, 7–8 June 2012; ACL Anthology: New York, NY, USA, 2012; pp. 310–318.
86. Zou, B.; Zhou, G.; Zhu, Q. Tree Kernel-Based Negation and Speculation Scope Detection with Structured Syntactic Parse Features. In *Proceedings of the EMNLP 2013—2013 Conference on Empirical Methods in Natural Language Processing*, Seattle, WA, USA, 18–21 October 2013; pp. 968–976.
87. Qian, Z.; Li, P.; Zhu, Q.; Zhou, G.; Luo, Z.; Luo, W. Speculation and Negation Scope Detection via Convolutional Neural Networks. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*; Association for Computational Linguistics (ACL): Stroudsburg, PA, USA, 2016; pp. 815–825.
88. Lazib, L.; Zhao, Y.; Qin, B.; Liu, T. Negation Scope Detection with Recurrent Neural Networks Models in Review Texts. In *Proceedings of the International Conference of Young Computer Scientists, Engineers and Educators*, Harbin, China, 20–22 July 2016; pp. 494–508.
89. Fancellu, F.; Lopez, A.; Webber, B. Neural Networks for Negation Scope Detection. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*; Association for Computational Linguistics (ACL): Stroudsburg, PA, USA, 2016; pp. 495–504.
90. Adel, H.; Schütze, H. Exploring Different Dimensions of Attention for Uncertainty Detection. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*; Association for Computational Linguistics (ACL): Stroudsburg, PA, USA, 2017; pp. 22–34.
91. Enger, M.; Velldal, E.; Øvrelid, L. An Open-Source Tool for Negation Detection: A Maximum-Margin Approach. In *Proceedings of the Workshop Computational Semantics Beyond Events and Roles*; Association for Computational Linguistics (ACL): Stroudsburg, PA, USA, 2017; pp. 64–69.
92. Li, H.; Lu, W. 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)*; Association for Computational Linguistics (ACL): Stroudsburg, PA, USA, 2018; pp. 533–539.
93. Fabregat, H.; Martinez-Romo, J.; Araujo, L. Deep Learning Approach for Negation Cues Detection in Spanish. In *Proceedings of NEGES 2018: Workshop on Negation in Spanish*, Seville, Spain, 18 September 2018; pp. 43–48.
94. Fancellu, F.; Lopez, A.; Webber, B. Neural Networks for Cross-Lingual Negation Scope Detection. *arXiv* **2018**, arXiv:1810.02156.
95. Al-Khawaldeh, F.T. Hierarchical Attention Generative Adversarial Networks for Biomedical Texts Uncertainty Detection. *Int. J. Adv. Stud. Comput. Sci. Eng.* **2019**, *8*, 1–12.
96. Fei, H.; Ren, Y.; Ji, D. Negation and Speculation Scope Detection Using Recursive Neural Conditional Random Fields. *Neurocomputing* **2020**, *374*, 22–29. [[CrossRef](#)]
97. McKenna, N.; Steedman, M. Learning Negation Scope from Syntactic Structure. In *Proceedings of the Ninth Joint Conference on Lexical and Computational Semantics*, Barcelona, Spain, 12–13 December 2020; pp. 137–142.
98. Lafferty, J.; McCallum, A.; Pereira, F. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In *Proceedings of the Eighteenth International Conference on Machine Learning*, San Francisco, CA, USA, 28 June–1 July 2001; pp. 282–289.
99. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of NAACL-HLT 2019 NAACL HLT 2019-North American Chapter of the Association for Computational Linguistics: Human Language Technologies*; Association for Computational Linguistics (ACL): Minneapolis, MN, USA, 2019; pp. 4171–4186.
100. Pennington, J.; Socher, R.; Manning, C. Glove: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, 25–29 October 2014; ACL Anthology: New York, NY, USA, 2014; pp. 1532–1543. [[CrossRef](#)]
101. Peters, M.; Neumann, M.; Iyyer, M.; Gardner, M.; Clark, C.; Lee, K.; Zettlemoyer, L. Deep Contextualized Word Representations. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2018; pp. 2227–2237.
102. Apostolova, E.; Tomuro, N.; Demner-Fushman, D. Automatic Extraction of Lexico-Syntactic Patterns for Detection of Negation and Speculation Scopes. In *Proceedings of the ACL-HLT 2011—49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, Portland, OR, USA, 19–24 June 2011; pp. 283–287.
103. Anis, S.; Saad, S.; Aref, M. A Survey on Sentiment Analysis in Tourism. *Int. J. Intell. Comput. Inform. Sci.* **2020**, *163*, 1–15. [[CrossRef](#)]
104. Chowdhury, M.F.M.; Lavelli, A. Exploiting the Scope of Negations and Heterogeneous Features for Relation Extraction: A Case Study for Drug-Drug Interaction Extraction. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*; Association for Computational Linguistics (ACL): Atlanta, GA, USA, 2013; pp. 765–771.
105. Mady, L.; Afify, Y.; Badr, N. Nested Biomedical Named Entity Recognition. *Int. J. Intell. Comput. Inform. Sci.* **2022**, *22*, 98–107. [[CrossRef](#)]
106. Hossain, M.M.; Anastasopoulos, A.; Blanco, E.; Palmer, A. It's Not a Non-Issue: Negation as a Source of Error in Machine Translation. In *Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2020*; Association for Computational Linguistics (ACL): Stroudsburg, PA, USA, 2020; pp. 3869–3885.

107. Wiegand, M.; Balahur, A.; Roth, B.; Klakow, D.; Montoyo, A. A Survey on the Role of Negation in Sentiment Analysis. In *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*; University of Antwerp: Antwerp, Belgium, 2010; pp. 60–68.
108. Pang, B.; Lee, L.; Vaithyanathan, S. Thumbs up? Sentiment Classification Using Machine Learning Techniques. In *Proceedings of the ACL—02 Conference on Empirical Methods in Natural Language Processing-EMNLP '02*; Association for Computational Linguistics (ACL): Morristown, NJ, USA, 2002; pp. 79–86.
109. Polanyi, L.; Zaenen, A. Contextual Valence Shifters. In *Computing Attitude and Affect in Text: Theory and Applications*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 1–10.
110. Jia, L.; Yu, C.; Meng, W. The Effect of Negation on Sentiment Analysis and Retrieval Effectiveness. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management—CIKM '09*; ACM Press: New York, NY, USA, 2009; pp. 1827–1830.
111. Misra, A.; Bhuiyan, M.; Mahmud, J.; Tripathy, S. Using Structured Representation and Data: A Hybrid Model for Negation and Sentiment in Customer Service Conversations. In *Proceedings of the Tenth Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*; Association for Computational Linguistics (ACL): Stroudsburg, PA, USA, 2019; pp. 46–56.
112. Habash, N.Y. Introduction to Arabic Natural Language Processing. In *Synthesis Lectures on Human Language Technologies (Book 10)*, 1st ed.; Hirst, G., Ed.; Morgan and Claypool Publishers: San Rafael, CA, USA, 2010; Volume 3, ISBN 9781598297959.
113. Abdul-Mageed, M.; Diab, M.T. Subjectivity and Sentiment Annotation of Modern Standard Arabic Newswire. In *Proceedings of the ACL HLT 2011-LAW 2011: 5th Linguistic Annotation Workshop*, Portland, OR, USA, 23–24 June 2011; pp. 110–118.
114. El-Dine, A.; El-zahraa, F. Sentiment Analyzer for Arabic Comments System. *Int. J. Adv. Comput. Sci. Appl.* **2013**, *4*, 99–103. [\[CrossRef\]](#)
115. Ben Hamouda, S.; Akaichi, J. Social Networks' Text Mining for Sentiment Classification: The Case of Facebook' Statuses Updates in the "Arabic Spring" Era. *Int. J. Appl. Innov. Eng. Manag.* **2013**, *2*, 470–478.
116. El-Beltagy, S.R. NileULex: A Phrase and Word Level Sentiment Lexicon for Egyptian and Modern Standard Arabic. In *Proceedings of the 10th International Conference on Language Resources and Evaluation, LREC 2016*, Portorož, Slovenia, 23–28 May 2016; pp. 2900–2905.
117. Duwairi, R.M.; Alshboul, M.A. Negation-Aware Framework for Sentiment Analysis in Arabic Reviews. In *Proceedings of the 2015 3rd International Conference on Future Internet of Things and Cloud, Rome, Italy, 24 August 2015*; pp. 731–735.
118. El-Naggar, N.; El-Sonbaty, Y.; El-Nasr, M.A. Sentiment Analysis of Modern Standard Arabic and Egyptian Dialectal Arabic Tweets. In *Proceedings of the IEEE Computing Conference 2017*, London, UK, 18–20 July 2017; pp. 880–887.
119. Ljajic, A.; Marovac, U. Improving Sentiment Analysis for Twitter Data by Handling Negation Rules in the Serbian Language. *Comput. Sci. Inform. Syst.* **2019**, *16*, 289–311. [\[CrossRef\]](#)
120. Costumero, R.; Lopez, F.; Gonzalo-Martín, C.; Millan, M.; Menasalvas, E. An Approach to Detect Negation on Medical Documents in Spanish. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 366–375, ISBN 9783319098906.
121. Skeppstedt, M. Negation Detection in Swedish Clinical Text: An Adaption of NegEx to Swedish. *J. Biomed. Semant.* **2011**, *2*, S3. [\[CrossRef\]](#) [\[PubMed\]](#)
122. Deléger, L.; Grouin, C. Detecting Negation of Medical Problems in French Clinical Notes. In *Proceedings of the 2nd ACM SIGHIT Symposium on International Health Informatics-IHI'12*, Miami, FL, USA, 28–30 January 2012; pp. 697–701.
123. Ogren, P.V.; Savova, G.K.; Chute, C.G. Constructing Evaluation Corpora for Automated Clinical Named Entity Recognition. In *Proceedings of the 6th International Conference on Language Resources and Evaluation, LREC 2008*; European Language Resources Association (ELRA): Marrakech, Morocco, 2008; pp. 3143–3150.
124. Demner-Fushman, D.; Kohli, M.D.; Rosenman, M.B.; Shooshan, S.E.; Rodriguez, L.; Antani, S.; Thoma, G.R.; McDonald, C.J. Preparing a Collection of Radiology Examinations for Distribution and Retrieval. *J. Am. Med. Inform. Assoc.* **2016**, *23*, 304–310. [\[CrossRef\]](#) [\[PubMed\]](#)
125. Gkotsis, G.; Velupillai, S.; Oellrich, A.; Dean, H.; Liakata, M.; Dutta, R. Don't Let Notes Be Misunderstood: A Negation Detection Method for Assessing Risk of Suicide in Mental Health Records. In *Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology*; Association for Computational Linguistics (ACL): Stroudsburg, PA, USA, 2016; pp. 95–105.
126. Zerva, C.; Ananiadou, S. Paths for Uncertainty: Exploring the Intricacies of Uncertainty Identification for News. In *Proceedings of the Workshop on Computational Semantics beyond Events and Roles*; Association for Computational Linguistics (ACL): Stroudsburg, PA, USA, 2018; pp. 6–20.