Natural Language Processing 13

Negation and Speculation Detection

Noa P. Cruz Díaz
and Manuel J. Maña López

JOHN BENJAMINS PUBLISHING COMPANY

Negation and Speculation Detection

Natural Language Processing (NLP)

ISSN 1567-8202

The scope of NLP ranges from theoretical Computational Linguistics topics to highly practical Language Technology topics. The focus of the series is on new results in NLP and modern alternative theories and methodologies.

For an overview of all books published in this series, please see
http://benjamins.com/catalog/nlp

Volume 13

Negation and Speculation Detection
by Noa P. Cruz Díaz and Manuel J. Maña López

Negation and Speculation Detection

Noa P. Cruz Díaz
Manuel J. Maña López
University of Huelva

John Benjamins Publishing Company
Amsterdam / Philadelphia

The paper used in this publication meets the minimum requirements of the American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

DOI 10.1075/nlp.13

Cataloging-in-Publication Data available from Library of Congress:
LCCN 2018047742 (PRINT) / 2018059189 (E-BOOK)

ISBN 978 90 272 0217 8 (HB)
ISBN 978 90 272 0216 1 (PB)
ISBN 978 90 272 6295 0 (E-BOOK)

John Benjamins Publishing Company · https://benjamins.com

Table of contents

Acknowledgements

Writing a book that collects all the knowledge about the subject in which you have invested so many hours of your life is one of the most rewarding experiences I have ever had. This experience will always be part of the beautiful and inexplicable path that I chose: that of science.

The result that you have today in your hands would be impossible without all those masters who preceded us, all those who were also interested in science, technology and natural language processing. Thank you to all of them for showing us the way.

The things that scare us are those that we end up remembering. Of course, my most sincere gratitude goes also to all the people who trusted me, assuaged my fears and believed that I would be up to what is required to write a book such as this.

In the difficult and arduous task of writing a book while carrying out many other activities, there are countless moments when your strength and resilience waver. These are the moments when beautiful people provide you with that toughness of spirit. It is impossible to name everyone but you know that regardless of the moments we have shared in this time, a part of this book is also yours. Thank you!

And finally, I would like to thank you, reader, for taking an interest in this task that has gained so much relevance in recent years and that can benefit so many other tasks in natural language processing. Science and technology give us the opportunity to create a better world; they provide us with the necessary tools that make the future ours to create. Therefore, thank you to science and technology and thank you to everyone who is a part of them.

The time has come, this book is already yours.

Seville (Spain), January 2, 2018.
Noa P. Cruz Díaz

In these lines I would like to thank the people who, during the past 20 years, have accompanied me in the exciting task of research in the field of natural language processing. My first thanks I would like to extend to Manuel de Buenaga, who taught me a course in a doctoral programme and introduced me to this exciting field of research. Manuel de Buenaga was also my doctoral advisor and since then I have been lucky enough to collaborate with him on many projects and research papers, to continue learning from him and to count on his friendship. Manuel, thank you for your brilliance.

In this very long and geographically varied period of my professional career there are many colleagues with whom I have had the opportunity to share hours of work and research, most of them from the Universities of Vigo and Huelva, where I have worked during these years. From of all of them I have learnt important lessons, so I would like to thank them for the fact that I have grown as a researcher and as a person.

Working at the University has given me the opportunity to meet hundreds of students during this time. Among other subjects, I have taught courses in natural language processing, information retrieval, search engines and text mining to a large number of them. I have tried to instil in all my students the passion that I feel about these subjects and many of them have returned it to me in equal measure. I would like to extend my most sincere gratitude to all of them.

In September 2008, I began a new line of research: the detection of negation and speculation in biomedical documents. I proposed to Noa Cruz that I advise her on a master's thesis in this subject. Ten years later, we wrote this book. Thank you, Noa, for the last ten years of work and collaboration.

Now it is the reader's turn. I trust that this book will help other students and researchers to develop their own interests in the problems that, in recent years, researchers from all over the world have devoted so much effort to solving. If this book succeeds in awakening readers' curiosity – or even their passion – for the work associated with the detection of negation and speculation, then we shall declare ourselves well and truly satisfied.

Huelva (Spain), January 2, 2018.
Manuel J. Maña López.

List of Abbreviations

BEP	Break-Even Point
BIO	Begin/Inside/Outside
CNN	Convolutional Neural Network
CPG	Clinical Practice Guideline
CRC	Colorectal Cancer
CRF	Conditional Random Fields
DCBS	Dependency-based Candidate Boundary Selection
ECG	Electrocardiogram
EHR	Electronic Health Record
EM	Expectation-Maximisation parameter estimation
FN	False Negative
FP	False Positive
HMM	Hidden Markov Model
HPSG	Head-driven Phrase Structure Grammar
KD	Knowledge Discovery
LSTM	Long Short-Term Memory
MLE	Maximum Likelihood parameter Estimation
NLP	Natural Language Processing
PoS	Part-of-Speech
PPV	Positive Predictive Value
RNN	Recurrent Neural Network
SDP	Stanford Dependency Parser
SFU	Simon Fraser University
SVM	Support Vector Machine
TP	True Positive
UMLS	Unified Medical Language System

CHAPTER 1

Introduction

1.1 Motivation

Negation and speculation are complex expressive linguistic phenomena that have been extensively studied from a theoretical perspective (Morante & Sporleder, 2012). They modify the meaning of the phrases in their scope. Negation denies or rejects statements, transforming a positive sentence into a negative one, e.g., *Mildly hyperinflated lungs without focal opacity*. Speculation, also known as hedging, is used to express a statement without attributing certainty to it, e.g., *Atelectasis in the right mid-zone is, however, possible*. These two phenomena are interrelated (De Haan, 1997) and have similar characteristics in text: they both have scope, so both affect the part of the text that is denoted by the presence of negation or speculation cue words (Konstantinova & de Sousa, 2011).

The amount of negative and speculative information in texts cannot be underestimated. (Szarvas, Vincze, Farkas, & Csirik, 2008) report that 13.45% of sentences in the abstracts section of the BioScope corpus and 13.76% of sentences in the full papers section contain negations. In addition, they show that the percentage of sentences with hedge cues in the abstracts and full papers section of the BioScope corpus are 17.70% and 19.44% respectively. In the review domain, this proportion is slightly higher. (Konstantinova et al., 2012) show that 18% of sentences in the Simon Fraser University (SFU) review corpus[1] (Taboada, Gillies, & McFetridge, 2006) contain negation cues and 22.7% of the sentences include speculation keywords. Therefore, the information that is inside the scope of any negation or speculation cue cannot be treated as factual. It should be discarded or presented separately with less confidence.

Negation and speculation detection is becoming an important task in natural language processing (NLP). In recent years, several challenges and shared tasks have included the extraction of these language forms, such as the BioNLP'09 Shared Task 3 (J. Kim, Ohta, Pyysalo, Kano, & Tsujii, 2009), the CoNLL-2010 Shared Task (Farkas, Vincze, Móra, Csirik, & Szarvas, 2010), the i2b2 NLP Challenge (Uzuner, South, Shen, & DuVall, 2011), the SEM 2012 Shared Task (Morante & Blanco, 2012) and the ShARe/CLEF eHealth Evaluation Lab 2014 Task 2 (Mowery et al., 2014).

1. <https://www.sfu.ca/~mtaboada/research/SFU_Review_Corpus.html>

In addition, a special issue of *Computational Linguistics* has been published on negation and speculation (Morante & Sporleder, 2012). This shows how computational linguistics has started to take into account the subjective aspects of language.

Detecting uncertain and negative assertions is relevant in a wide range of applications such as information extraction (Savova et al., 2010), machine translation (Baker et al., 2012), sentiment analysis (Cruz, Taboada, & Mitkov, 2015; Reitan, Faret, Gambäck, & Bungum, 2015), paraphrasing and recognising textual entailment (AL-Khawaldeh, 2015; Sharma, Sharma, & Biswas, 2015). For all of these tasks it is crucial to know when a part of the text should have the opposite meaning (in the case of negation) or should be treated as subjective and non-factual (in the case of speculation). This part of the text is what is known as the scope.

At first glance, negation and speculation might seem easy to deal with. The problem could be broken down into finding negative and speculative cues and determining their scope. However, dealing with them is in fact much more problematic. Negation and speculation pose considerable challenges since they interact with many other phenomena and are used for so many different purposes. This means that a deep analysis of the text is needed.

This book is motivated by the fact that this is an emerging topic that has attracted the attention of many researchers, and there is clearly a lack of relevant textbooks and survey texts. It aims to define negation and speculation from an NLP perspective, to explain the need for processing these phenomena, to summarise existing research on processing negation and speculation, to provide a list of resources and tools, and to speculate about future developments in this research area. An advantage of this book is that it will not only provide an overview of the state of the art in negation and speculation detection, but will also introduce newly developed data sets and scripts.

This manuscript will be useful for students of NLP subjects who are interested in understanding this task in more depth, as well as for researchers in this field. It is also aimed at developers and researchers with an interest in the phenomena of negation and speculation in order to improve performance in other NLP tasks, such as text mining or sentiment analysis.

1.2 Negation and speculation in natural language

Negation and speculation in natural language present a real challenge to researchers. They are also a recurring theme in grammar. Negation and speculation are not limited to the linguistic field but are also related to many disciplines and domains, including philosophy, logic, mathematics and sociology.

Speculation should be studied within the framework of modality, i.e., what allows the expression of the attitude of the speaker (Morante & Sporleder, 2012). As described in (De Haan, 1997), the study of modality has a long history. However, it was not until relatively recently that it was recognised as an area of linguistic research in its own right. Evidently, there is by no means uniformity or agreement on what exactly is meant by the term modality.

For its part, negation, unlike modality, can be considered as a universal feature of natural language, in the sense that all languages have a system for denying statements in one way or another. In addition, not only is its existence apparently universal, but also it appears that the way in which it is manifested in different languages tends to move in a general direction (Horn & Kato, 2000).

The large number of publications and conferences held on this subject indicate the complexity of these linguistic phenomena and their importance.

1.3 Basic notions

This section presents a brief overview of basic concepts in negation and speculation detection that are essential for a better understanding of the following chapters of the book.

NLP, also known as computational linguistics, was born in the 1950s at the intersection of artificial intelligence and linguistics (Nadkarni, Ohno-Machado, & Chapman, 2011). NLP is the subfield of computer science concerned with using computational techniques to learn, understand, and produce human language content (Hirschberg & Manning, 2015).

Examples of common low-level sub-problems in NLP are sentence boundary detection, tokenisation, part-of-speech, tagging and shallow parsing (chunking). Higher-level sub-problems build on low-level tasks and are usually problem-specific. Examples are named entity recognition, word sense disambiguation and relationship extraction (Jurafsky & Martin, 2008). Negation and speculation detection can also be included in this type of sub-problem.

Negation and speculation extraction is typically broken down into two subtasks: cue identification and scope recognition. The cues are the words that have negative or speculative meaning (e.g., *without*, *not*, *suggest*, *may*). They can be single words, multiwords, prefixes or suffixes.

The scope is a text fragment governed by the corresponding cue in a sentence (Qian et al., 2016). Therefore, detecting the scope consists of marking the sequence of words in the sentence that is affected by each cue. Both cues and scopes may be discontinuous.

In Example (1), the scope is marked with square brackets and the cue in bold.

(1) [I] **fail** to [see how you could have done more]

Recently, some works have also included the recognition of the event(s) affected by the cue. In a very general sense, the term "event" can denote a process, an action or a state. Negated events are frequently reported in both biological literature and clinical notes. In Example (1), the event (underlined) is the verb "see".

The other task, introduced in the SEM 2012 Shared Task (Morante & Blanco, 2012) and related to negation detection, is detecting the focus.

Focus is the part of the scope that is most prominently or explicitly negated. Both concepts are interconnected (Huddleston & Pullum, 2002). Scope corresponds to all elements any of whose individual falsity would make the negated statement true. Focus is the element of the scope that is intended to be interpreted as false to make the overall negative true. The focus is more difficult to identify, especially without knowing stress or intonation (Blanco & Moldovan, 2011b).

Example (2) shows an annotated sentence where focus is marked with curly brackets and the cue in bold.

(2) His new job **doesn't** {require} anything.

Finally, there are two basic approaches to solving this kind of problem in NLP. The first uses ruled-based systems and makes use of linguistic information directly integrated into the workflow. The second approach is called statistical machine learning and relies on textual data from which the algorithm learns generalisations on its own.

1.4 Application domains

Medical practitioners are increasingly incorporating results and findings from clinical studies into their work. The availability of vast databases of scientific articles allows access to this material, although the huge volume also makes it difficult to locate relevant material. Furthermore, many hospitals have electronic records of their patients' medical backgrounds and several others are proceeding to digitise records. This enables physicians to carry out clinical studies that allow progress in evidence-based medicine. However, as in the case of access to scientific information, physicians need to have efficient tools for accessing this information and then analysing the text in greater depth. This analysis should include negation and speculation detection because these linguistic phenomena are used extensively in this domain with the aim to express impressions, hypothesised explanations of experimental results, or negative findings. A superficial analysis of medical text could result in an automated indexing system that suffers in terms of precision.

Chapman's work, for example (W. W. Chapman, Bridewell, Hanbury, Cooper, & Buchanan, 2001a), shows that the proportion of clinical findings that are negated in clinical reports ranged between 39% and 83%, depending on the type of report analysed. This research studied ten kinds of clinical reports and a total of 42,160 documents. The lowest percentage was found for surgical pathology reports and the highest for mammograms. Clearly, when querying large medical free-text databases, the presence of negations can yield numerous false-positive matches and therefore it is necessary to acknowledge whether words have been negated or not.

Also in the biomedical domain, other tasks such as interaction extraction could benefit from this type of text analysis. In interaction extraction, the aim is to mine text evidence for biological entities with certain relations between them. Here, an uncertain relation or the nonexistence of a relation might be of some interest for an end-user, so this information must not be confused with real textual evidence (Szarvas et al., 2008).

Sentiment analysis deals with the automatic detection and treatment of opinion in natural language applications. It is important for reasons in areas such as recommendation systems, affective computing and market research (Lapponi, Read, & Ovrelid, 2012). In this domain, hedges are linguistic tools that allow authors to indicate that they cannot back their opinions with facts. Thus, speculations include certain modal constructions, along with other markers such as indirect speech (e.g., *according to certain researchers*). On the other hand, there are modal constructions that are not hedges, i.e., when expressing a factual possibility, without uncertainty on the part of the speaker (e.g., "*these insects may play a part in the reproduction of plants as well*") (Benamara, Chardon, Mathieu, Popescu, & Asher, 2012).

Negation is one of the most common linguistic means for changing polarity (e.g., the polarity of the statement *Just a V–5 engine, spectacular* should be the opposite of its negation, as in *Just a V–5 engine, nothing spectacular*). There are different types of negation, such as negative operators (*not*, *no more*, *without*), negative quantifiers (*nobody*, *nothing*, *never*), and lexical negations (*lack*, *absence*, *deficiency*), each of which has different effects on both the polarity and the strength of the negation. As discussed in Benamara et al. (2012), negation always changes the polarity, but the strength of an opinion expression in the scope of negation is not greater than that of the opinion expression alone. Furthermore, opinions in the scope of multiple negatives have a higher strength than if in the scope of a single negative. Hence, dealing with negation requires going beyond polarity reversal, since simply reversing the polarity of sentiment upon the appearance of negation may result in inaccurate interpretation of sentiment expressions.

The literature on sentiment analysis and opinion mining (Councill, McDonald, & Velikovich, 2010; Dadvar, Hauff, & de Jong, 2011; Lapponi et al.,

2012) has emphasised the need for robust approaches to negation detection, and for rules and heuristics for assessing the impact of negation on evaluative words and phrases.

1.5 Structure of the book

A detailed outline of the book is described below.

Chapters 2 and 3 go into detail about the concepts of negation and speculation, including a classification of the different types of each concept. In addition, relevant literature is analysed and descriptions and comparisons are made of the most relevant negation and speculation detection systems found in the literature, with the different approaches followed by the authors in order to solve the problem highlighted.

Chapter 4 is an in-depth description of the applications for which information about negation and speculation has proven to be useful, e.g., text mining, sentiment analysis and opinion mining, recognising textual entailment, machine translation and information retrieval.

Chapter 5 presents a set of relevant resources for any researcher or developer interested in the problem. It also includes information about available scripts for evaluation.

Finally, Chapter 6 discusses the possibilities for future work and the challenges remaining in each domain.

CHAPTER 2

Negation

Chapter 2 is an overview of the concept of negation and the major topics concerning it, including a classification of the different types of negation. In addition, this chapter analyses the most relevant negation detection systems found in the literature, showing that this task has been an active research area during recent years in the NLP community.

2.1 Definition of negation

Negation is a feature in all languages. Its most obvious function is to turn a proposition into its opposite. In its more sophisticated forms, it can be strongly expressive and include euphemisms and irony. Unlike affirmative statements, negation is always marked by words (e.g., *not*, *without*), prefixes (e.g., *un*-, *in*-) or suffixes such as *-less* (Blanco & Moldovan, 2011b). As introduced in Section 1.3, in most cases, negation involves a cue and a negated syntagma which contains one or more words that are within the scope of negation (Ballesteros, Francisco, Díaz, Herrera, & Gervás, 2012). For instance, in (1), *not* is the negation cue used to denote that the following concept (in this example, *expensive*) is negated.

(1) The chair is *not* expensive but comfortable.

However, negation is much more than a grammatical phenomenon present in all languages. It is a linguistic, cognitive and intellectual phenomenon, as Lawler (2010) affirms. Authors like Horn and Kato (2000) add that negation is a central feature of language and cognition, which interacts with all areas of grammar and with the philosophy of language. In fact, negation in logic is well defined and syntactically simple (i.e., it is a unary operator which reverses the truth value) but in natural language it is complex.

The study of negation from a philosophical perspective dates back to Aristotle. He defines the law of contradiction (a statement cannot be true and false at the same time) and the law of excluded middle (a statement must be either true or false). Since then, many studies have been carried out in this regard, many of them collected in Seifert and Welte (1987). Additionally, the research conducted by Horn (1989) ought to be mentioned here since it is currently considered a masterpiece.

In this work, Horn outlines all the major questions concerning negation since Aristotle, and touches on negative polarity as well.

From a linguistic perspective, Tottie (1991) provides a quantitative analysis of negation, including a discussion about its linguistic variation. She finds, for example, that there are twice as many negation words in speech as in writing (2.67 vs. 1.28, per 100 words). Valencia (1991) and Dowty (1994) study how negation influences reasoning while Hintikka (2002) supports the argument that negation is a complex subject. At the same time, he explains that negation normally constitutes a barrier to anaphora and that it interacts with quantifiers. In addition, he makes a distinction between contradictory and contrary negation. Van der Wouden (2002) defines the concept of negative context and discusses collocation, polarity and multiple negation. He argues that these topics are closely related since collocation is the general phenomenon of lexical items having a restricted distribution whereas polarity items are a specific class of such lexical elements. He also adds that the same formal apparatus used to explain the behaviour of polarity items can be applied to other phenomena, such as some types of multiple negations. Polarity and multiple negations are also covered in *The Cambridge Grammar of the English Language* (Huddleston & Pullum, 2002), which includes a chapter tackling the question of negation. Morante and Sporleder (2012) summarise several aspects of negation and show that negative polarity and negation are different, though related, concepts. Essentially, they explain that negation and polarity are related in the sense that negation can reverse the polarity of an expression. In this context, negative polarity items can be seen as expressions with a limited distribution, part of which includes negative sentences like *any* in the sentence *I didn't read any book*. Other works such as those presented by Laka (2013) explore negation from a syntactic point of view.

Finally, it is worth noting that negation is common in language. Indeed, as mentioned in Section 1.1, Szarvas et al. (2008) report that the number of negative sentences in the BioScope corpus is about 13% depending on the type of documents. Also in the biomedical domain, Nawaz, Thompson and Ananiadou (2010) explain that more than 3% of the biomedical events in 70 abstracts from the GENIA corpus (J. D. Kim, Ohta, Tateisi, & Tsujii, 2003) are negated. For their part, Councill et al. (2010) annotate a corpus of product reviews with negation information, finding that 19% of the sentences contain negations. More recently, Konstantinova et al. (2012), show that 18% of the SFU Review corpus sentences include negative information. This proportion is higher in the ConanDoyle-neg corpus where 22.49% of sentences are negative (Morante & Daelemans, 2012).

2.2 Types of negation

The major distinction can be made between constituent (or local) negation and clausal (or sentential) negation (Klima, 1964). A clausal negation negates an entire preposition (e.g., *he does not have money*) while a constituent negation is associated with some constituent or clause (e.g., *he has no money*). Although their effects can be similar or identical, the latter is less common grammatically.

Tottie (1991) presents the following comprehensive taxonomy of English clausal negation:

a. Denials. They are the most common form and constitute an unambiguous negation of a particular clause (e.g., *The audio system on this television is not very good, but the picture is amazing.*)
b. Rejections. They appear in expository text where a writer explicitly rejects a previous supposition (e.g., *Given the poor reputation of the manufacturer, I expected to be disappointed with the device. This was not the case.*)
c. Imperatives. They instruct an audience not to take a particular action (e.g., *Do not neglect to order their delicious garlic bread.*)
d. Questions. For instance, *Why couldn't they include a decent speaker in this phone?*
e. Supports and Repetitions. They express agreement and add emphasis or clarity. They involve multiple expressions of negation.

Tottie includes rejections and supports in intersentential negation (i.e., the language used in one sentence may explicitly negate a proposition or implication found in another sentence) while denials, imperatives, and questions are examples of sentential negation.

For his part, Payne (1997) defines different types of both clausal and constituent negation in any language. Clausal negation can be divided into the following categories:

a. Lexical negation, which describes a situation in which the concept of negation is part and parcel of the lexical semantics of a particular verb.
b. Morphological negation, where the morphemes that express clausal negation are associated with the verb.
c. Analytic negation, in which the negative particles are normally associated with the main verb of the clause (e.g., *n't, not, never*).

The different types of constituent negation are described as follows:

a. Derivational negation. Languages allow a stem to convert into its opposite by the use of derivational morphology (i.e., suffixes and prefixes).

b. Negative quantifiers. Many languages employ quantifiers that are either inherently negative (e.g., *none*) or negated independently of clausal negation (e.g., *not many*).

Other authors identify different classes of negation in English. For example, Huddleston and Pullum (2002) define the four contrasting pairs of negation classes presented below:

a. Verbal vs. Non-Verbal. In verbal negation, the negative particle is associated with the verb whereas in non-verbal negation, the negation cue is related to a dependent of the verb.
b. Clausal vs. Subclausal. A negation is clausal if it yields a negative clause. Otherwise, the negation is subclausal.
c. Analytics vs. Synthetic. Negation that is analytic is denoted by words whose only syntactic function is to mark negation. In synthetic negation, the words that mark negation also have other functions in the sentence.
d. Ordinary vs. Metalinguistic. Ordinary negation indicates that something is not the case while metalinguistic negation does not dispute the truth but rather reformulates a statement.

Harabagiu, Hickl and Lacatusu (2006) distinguish two main classes of negation: overt (directly licensed) negations and indirectly licensed negations. The former includes overt negative markers such as *n't*, negative quantifiers (e.g., *no*) and strong negative adverbs like *never*. The latter consists of verbs or phrasal verbs (e.g., *fail, keep from*), prepositions such as *except*, weak quantifiers like *few* and traditional negative polarity items (e.g., *any more*).

2.3 Negation detection

Studies into the problem of negation detection evolve from rule-based approaches to machine learning techniques. In addition, recent studies are trying to explore how efficient the deep-learning algorithms are when applied to this task. Negation detection has been focused mainly on the biomedical domain because of the different challenges and shared tasks related to this area that have been carried out in the past and because of the set of resources developed (e.g., the BioScope corpus annotated for negation and speculation (Szarvas et al., 2008)). However, other areas such as literature or reviews, where some corpora have recently been published, have been investigated. The following sections provide an overview of the most relevant works on the recognition of negation cues and their scope.

2.3.1 Rule-based approaches

The study by W. W. Chapman, Bridewell, Hanbury, Cooper and Buchanan, (2001b) stands out above all others in the biomedical domain. Their algorithm, NegEx, which is based on regular expressions, determines whether a finding or disease mentioned in narrative medical reports is present or absent. Although the algorithm is described by the authors themselves as simple, it has proven to be powerful in negation detection in discharge summaries. The reported results of NegEx show a positive predictive value (PPV or precision) of 84.5%, sensitivity (or recall) of 77.8%, and specificity of 94.5%.[1] However, when NegEx is applied to a set of documents from a different domain than that for which it was conceived, the overall precision is about 20 percentage points lower (Mitchell, 2004). Trying to eliminate NegEx's false positives, Mehrabi et al. (2015) develop a negation algorithm called DEEPEN which takes into account the dependency relationship between negation words and concepts within a sentence using the Stanford dependency parser (De Marneffe, MacCartney, & Manning, 2006). The evaluation results demonstrate that DEEPEN can reduce the number of incorrect negation assignments for patients with positive results in Electronic Health Records (EHRs). Also based on the simple approach used by NegEx for finding negated conditions in text, Harkema, Dowling, Thornblade and Chapman (2009) present an algorithm called ConText. It relies on trigger terms, pseudo-trigger terms, and termination terms for identifying the values of three contextual features (Negation, Temporality and Experiencer). In spite of its simplicity, this approach performs well at identifying negation and hypothetical statuses. However, it performs only moderately well at determining whether a condition was experienced by someone other than the patient and whether the condition has occurred historically. Elazhary (2017) proposes an algorithm (NegMiner) to address some of the shortcomings of the NegEx system. It exploits some basic syntactic and semantic information to deal with more negations compared to the NegEx algorithm, including the ability to deal with multiple negations. Each term of the Unified Medical Language System (UMLS) in an output sentence is accompanied by explanation of the mining decision to help highlight any shortcomings that would necessitate future updates. This capability also helps in addressing one of the most prominent problems of the NegEx algorithm, which is its inability to deal with the existence of a UMLS term several times in a single sentence. Experimental results show a superior performance by the NegMiner algorithm compared to the simulated NegEx algorithm.

Mutalik, Deshpande and Nadkarni (2001) and Elkin et al. (2005) also conduct research into regular expressions. Gindl, Kaiser and Miksch (2008) propose a

1. The measures of effectiveness are explained in Section 5.3 of this book.

method called NegHunter, which classifies negations in clinical practice guidelines (CPGs) according to identified negation types. Results show that the use of syntactical methods can improve negation detection, not only in medical writings but also in arbitrary narrative texts. Apostolova, Tomuro and Demner-Fushman (2011) present a linguistically motivated rule-based system for the detection of negation scopes. The system rule set consists of lexico-syntactic patterns extracted automatically from the BioScope corpus and it outperforms the baseline in all cases and exhibits results comparable to machine learning systems. Ballesteros et al. (2012) incorporate syntactic parsing to improve negation detection and to infer the scope of negations. Their system consists of two algorithms: the first detects words affected by the negative operators' (cues) traversing dependency trees and the second uses rules to annotate sentences within the scope of negations. From the results, they conclude that dependency parsing is a valuable auxiliary technique for negation detection, at least in English.

In the review domain, many existing approaches have relatively straightforward conceptualisations of the scope of negation keywords. For instance, Pang and Lee (2004) assume that the scope of a negation cue consists of the words between the negation keyword and the first punctuation mark following it. Kennedy and Inkpen (2006) introduce the concept of contextual valence shifters (i.e., negation, intensifier and diminisher). They experiment with taking as the scope the remainder of the sentence as well as the first sentiment-carrying word following the negation cue. Other approaches only consider specific types of words. For example, Hu and Liu (2004) suggest that the scope of negation is bounded by the adjectives that appear closely around the negation cue. They remark that the word distance between the negation keyword and the words in the scope should not exceed a threshold of around five. However, these solutions are not accurate enough.

L. Jia, Yu, and Meng (2009) propose a rule-based system that uses information derived from a parse tree. This algorithm computes a candidate scope, which is then pruned by removing those words that do not belong to the scope. Heuristic rules are used to detect the boundaries of the candidate scope. These rules include the use of delimiters (i.e., unambiguous words such as *because*) and conditional word delimiters (i.e., ambiguous words like *for*). There are also defined situations in which a negation cue does not have an associated scope. The authors evaluate the effectiveness of their approach on polarity determination. The first set of experiments concerns the accuracy of computing the polarity of a sentence while the second involves the ranking of positively and negatively opinionated documents in the TREC blogosphere collection (Macdonald & Ounis, 2006). In both cases, their system outperforms the other approaches described in the literature.

Many systems have been developed in relation to the *SEM 2012 shared task, namely resolving the scope and focus of negation (Morante & Blanco, 2012), although most of them are based on machine learning techniques as described in the next section. Rule-based systems for detecting the presence of negations and delimitating their scope are described by de Albornoz, Plaza, Díaz, and Ballesteros (2012) and Ballesteros et al. (2012). The former was initially designed for processing opinionated texts. It applies a dictionary approach to cue detection, with the detection of affixal cues being performed using WordNet (Miller, 1995). Non-affixal cue detection is performed by consulting a predefined list of cues. It then uses information from the syntax tree in order to get a first approximation of the scope, which is later refined using a set of post-processing rules. In the case of the latter system, an algorithm detects negation cues and their scope by traversing Minipar dependency structures. Finally, the scope is refined with post-processing rules that take into account the information provided by the first algorithm and linguistic-clause boundaries.

2.3.2 Machine learning based systems

Much of the work in the field of negation detection is based on machine learning approaches. Examples of the detection of negated concepts in medical narrative using machine learning techniques are to be found in studies by Averbuch, Karson, Ben-Ami, Maimon, and Rokach (2004) and Goldin and Chapman (2003).

The research conducted by Morante, Liekens and Daelemans (2008) is worth highlighting. It shows a high performance in all sub-collections of the BioScope corpus (Szarvas et al., 2008). Their machine learning system consists of two classifiers. The first decides if the tokens in a sentence are negation cues. The second determines which words in the sentence are affected by the negation. They apply post-processing to increase the number of fully correct scopes. With this approach, the algorithm shows an F-score of 80.99%; and 50.05% of scopes are correctly identified. An improvement on this system is presented by the authors in Morante and Daelemans (2009b). They employ four classifiers instead of one to find the full scope of the negation cues. Three classifiers predict whether a token is the first token, the last, or neither in the scope sequence. A fourth classifier uses these predictions to determine the scope classes. To predict the cues, a list of 17 negation keywords extracted from the training data set is used. Instances with these negation cues are directly assigned to their class, so the classifiers only predict the class of the rest of the tokens. The set of documents employed for experimentation is wider (they use the whole BioScope corpus instead of just the abstracts as the previous system did). The third difference between these two approaches is that, in this case, a more refined set of attributes is used. For clinical documents, the F-score of negation detection is 84.2%; 70.75% of scopes are correctly identified. For full papers, the F-score is 70.94%; 41% of scopes are correctly predicted. In the case of abstracts, the F-score is 82.60% and the percentage of scopes correctly classified is 66.07%.

Another system worth mentioning is the one developed by Agarwal & Yu (2010). In this work, the authors detect negation cue phrases and their scope in clinical notes and biological literature from the BioScope corpus using conditional random fields (CRF) as a machine learning algorithm. The authors select all negation sentences from the three sub-corpora and an equal number of randomly chosen non-negation sentences. These new sub-corpora are divided into two groups; one is used for training and the other for testing. The best CRF-based model achieves F-scores of 98% and 95% on detecting negation cue phrases and their scope in clinical notes, and F-scores of 97% and 85% on determining negation cue phrases and their scope in biological literature. However, owing to the fact that the corpus partitions and the evaluation measures are different, this system is not comparable with the approaches previously described.

Zhu, Li, Wang and Zhou (2010) present an interesting approach to scope learning. They formulate it as a simplified shallow semantic parsing problem by regarding the cue as the predicate and mapping its scope onto several constituents as the arguments of the cue. Evaluation on the BioScope corpus shows an F-score of 78.50% for abstracts, 57.22% for papers and 81.41% for clinical documents (using as cues those previously detected for the classifier). With the gold-standard cues (those that appear annotated as such in the corpus), the results are notably higher. This means that this kind of system, together with an accurate cue classifier, could be used to tackle the task.

Cruz Díaz, Maña López, Vázquez and Álvarez (2012) propose a two-stage approach: first, a binary classifier determines whether each token in a sentence is a negation cue or not. A second classifier is trained to determine, at sentence level, which tokens are affected by the cues previously identified. This system is trained and evaluated on the clinical texts of the BioScope corpus. In the cue-detection task, the classifier obtains an F_1-score of 97.3%. In the scope detection task, a token is correctly classified if it has been properly identified as being inside or outside the scope of all the negation cues present in the sentence, achieving an F_1-score of 93.2%. Also using the BioScope corpus as a learning and evaluation source, Zou, Zhou and Zhu (2013) propose a novel approach for tree kernel-based scope detection by using the structured syntactic parse information. In addition, they explore methods of selecting compatible attributes for different parts-of-speech (PoS) since features have imbalanced efficiency for scope classification, which is normally affected by the PoS. Evaluation produces an F-score of 76.90% for the abstracts sub-collection, 61.19% for papers and 85.31% for clinical documents (using the gold-standard cues).

For their part, Wu et al. (2014) introduce a machine learning-based polarity module for detecting negation in clinical texts, and extensively compare its performance across domains. They train and test their system on four manually annotated corpora of clinical texts: the SHARPn natural language processing (NLP) Seed corpus (Rea et al., 2012); the 2010 i2b2/VA NLP Challenge corpus (Uzuner et al., 2011), the MiPACQ corpus (Cairns et al., 2011) and the NegEx test set. The authors conclude from this study that practical negation detection is not reliable without in-domain training data and/or development. Thus, it can be optimised for a domain, but is difficult to generalise across domains.

Recently, Attardi, Cozza and Sartiano (2015) describe a two-step approach to negation recognition in which the scope-detection step exploits the structure of sentences as represented by a sentence's dependency parse tree. The novelty of this approach is that the dependency tree is used as a guide in the choice of how to extend the current scope, which avoids producing spurious scopes (e.g., *discontinuous scopes*). The algorithm may also gather partial subtrees of the parse, providing more resilience and flexibility. Experiments on the BioScope corpus show that the algorithm achieved accuracy scores above the state-of-the-art. Finally, Shivade, de Marneffe, Fosler-Lussier and Lai (2015) apply kernel methods to extend the NegEx system. Using the source code of the rule-based system, the authors construct a binary feature corresponding to each cue and conjunction, and thus generate a feature vector for every sentence in the dataset, building a linear kernel, which was implemented in LibSVM (Chang & Lin, 2011). They also design a kernel that augments the decision by NegEx with the bag-of-words model. The approach is evaluated on the NegEx text set and four other datasets adapted from the 2010 i2b2 challenge (Uzuner et al., 2011). The results show that using a simple bag-of-words kernel with the NegEx output as an additional feature yields significantly improved results compared to the NegEx rule-based system, mainly owing to an increase in recall. This kernel generalises well and shows promising results when trained and tested on different datasets.

In contrast to the biomedical domain, the impact of negation detection on sentiment analysis using machine learning techniques has not been sufficiently investigated, perhaps because reasonably sized standard corpora annotated with this kind of information have only recently become available. Councill et al. (2010) define a system that can precisely identify the scope of negation in free text. The cues are detected using a lexicon (i.e., a dictionary of 35 negation keywords). A CRF is employed to predict the scope. This classifier incorporates, among others, features from dependency syntax. The approach is trained and evaluated on a product-review corpus. It yields an 80.0% F-score and correctly identifies 39.8% of scopes. The authors conclude that, as they had expected, performance is improved dramatically by introducing negation scope detection (29.5% for positive sentiment and 11.4% for negative sentiment, both in terms of F-score). Using the same corpus, Lapponi et al. (2012) present a state-of-the-art system for negation detection. At the heart of the system is the application of CRF models for sequence labelling, which makes use of a wealth of lexical and syntactic features, together with a fine-grained set of labels that capture the scopal behaviour of tokens. At the same time, they demonstrate that the choice of representation has a significant effect on performance. Also in the review domain, Cruz et al. (2015) present a machine learning system that automatically identifies negation cues and their scope in the SFU Review corpus, annotated for negation and speculation (Konstantinova et al., 2012). The results obtained by this system are in line with the results of other authors in the same task and domain such as Councill et al. (2010) and Lapponi et al. (2012). Skeppstedt, Paradis and Kerren (2016) train machine learning models to recognise markers for negation. Two English corpora are used in the experiments, the Bioscope corpus (Szarvas et al., 2008) and the SFU Review corpus (Konstantinova et al., 2012). Three setups are used: models trained on a subset of the BioScope corpus and evaluated on another subset of the same corpus, models trained on a subset of the SFU Review corpus and evaluated on another subset of this corpus, and finally models trained on the SFU Review corpus and evaluated on the BioScope corpus. In the BioScope corpus, when the algorithm is trained on text of the same genre, the method achieves results in line with the inter-annotator agreement. However, results for detecting negation in the SFU Review corpus are much lower than the measured agreement figures. To train the model on consumer reviews and apply it to clinical text also yields low results, showing that neither the trained models nor the choice of appropriate algorithms and features are transferable across the two text genres.

Regarding the machine learning systems developed in the *SEM 2012 shared task (Morante & Blanco, 2012); the approach presented by Read, Velldal, Øvrelid and Oepen (2012) deserves a mention. It combines support vector machines (SVM) cue classification with SVM-based ranking of syntactic constituents for scope resolution. The approach is extended to identify negated events by first classifying negations as factual or non-factual, and then applying an SVM ranker over candidate events. The original treatment of factuality in this system results in the highest score for both the negated event subtask and the global task. Lapponi, Velldal, Øvrelid and Read (2012) propose a system which combines SVM cue classification with CRF-based sequence labelling. An original aspect of this approach is the model representation for scopes and negated events, where tokens are assigned a set of labels that attempts to describe their behaviour within the mechanics of negation. After unseen sequences are labelled, in-scope and negated tokens are assigned to their respective cues using simple post-processing heuristics. M. Chowdhury and Mahbub (2012) present a system, consisting of three different CRF classifiers, that exploits phrasal and contextual clues separately from various token specific features. This system is ranked third among the participating teams and attains the highest F_1-score for negation cue detection. Abu-Jbara and Radev (2012) also uses three different CRF classifiers. A characteristic of the cue model of this system is that tokens are assigned five labels in order to represent the different types of negation.

2.3.3 Hybrid approaches

Goryachev, Sordo, Zeng and Ngo (2006) modify two existing regular expression-based algorithms, NegEx and NegExpander, in an attempt to improve their performance, and create two classification-based methods, both of which are machine learning algorithm-based classifiers, a Naïve Bayes and a SVM. The algorithms are trained on 1745 discharge reports from a Boston-based hospital and evaluated on 100 randomly chosen outpatient reports from two different Boston-based hospitals. The accuracy of regular expressions methods (91.9 for NegEx versus 92.3% for NegExpander) is higher than that of classification-based methods (83.5 for Naïve Bayes versus 89.9% for SVM). Therefore, NegEx is the algorithm which produces the best results. Huang and Lowe (2007) report that negated terms may be difficult to identify if negation cues are more than a few words away from them. To address this limitation in the automatic detection of negations in clinical radiology reports, the authors propose a novel hybrid approach, combining regular expression with grammatical parsing. The sensitivity of negation detection is 92.6%, the PPV is 98.6%, and the specificity is 99.8%. Drawing on the BioScope corpus, Velldal, Øvrelid, Read and Oepen (2012) combine manually crafted rules with machine learning techniques. Dependency rules are used for all cases where a head-driven phrase structure grammar (HPSG) parser is not available. Where a HPSG parser is available, the scope predicted by these rules is included as a feature in a constituent ranker model which automatically learns a discriminative ranking function by choosing subtrees from HPSG-based constituent structures. Although the results obtained by this system can be considered as the state-of-the-art, the combination of novel features together with the classification algorithm chosen in the system developed by Cruz Díaz et al. (2012) improves the results for the sub-collection of clinical documents to date. Fujikawa, Seki and Uehara (2013) propose a hybrid approach to negation identification combining statistical and heuristic approaches, named NegFinder. The system is composed of three phases: identification of negation cues, identification of negation scopes and adjustment of negation scope. The first two phases are based on supervised classifiers, IGTree (Daelemans, Van Den Bosch, & Weijters, 1997), and the last phase is based on a heuristic rule using grammatical parsing. To demonstrate the effectiveness of the approach, the authors conduct experiments on the BioScope corpus, reporting F-scores of 89.8% for abstracts, 78.7% for full papers and 94.2% for clinical texts.

A hybrid approach presented in the *SEM 2012 shared task (Morante & Blanco, 2012) is the system designed by White (2012). It has a CRF sequence tagger for scope and negated event detection, while cues are recognised by four different regular expression rule patterns: affixes (partial token), single (whole) token, continuous multiple token and discontinuous multiple token. The system developed by Gyawali and Solorio (2012) follows the same approach but employs SVM instead of CRF.

Reitan et al. (2015) describe an approach to negation scope detection for Twitter sentiment analysis. The system consists of two parts: a negation cue detector which uses a lexicon lookup and a CRF-based scope classifier. It is evaluated on the Twitter Negation corpus, a set of 4,000 tweets annotated for the task by two of the authors. The cue detector yields high recall, but modest precision. The negation scope classifier obtains an F_1-score of 85.3%, with 64.5% of scopes correctly classified. The authors also develop a sentiment classifier for Twitter data, confirming that taking negation into account tends to improve sentiment classification performance significantly.

Finally, Pröllochs, Feuerriegel and Neumann (2017) propose a novel learning strategy to detect negations, i.e., they apply reinforcement learning to develop a system that replicates the human perception of negations based on an exogenous response, such as a user rating for reviews. The evaluation reveals a superior performance in predicting negation scopes. In addition, reinforcement learning allows for hypothesis testing in order to pinpoint how humans process and act on negations.

2.3.4 Deep learning

Deep learning has recently been growing in popularity and some authors have investigated whether this kind of approach is a valid alternative when it comes to detecting negation in NLP.

For example, Fancellu, Lopez and Webber (2016) design two different neural networks architecture: a one hidden layer feed-forward neural network and a bidirectional long short-term memory (LSTM) model. Training, development and tests are carried out using the Conan Doyle corpus (Morante & Blanco, 2012). Both training and testing are done on negative sentences only. The results show that neural networks perform on par with previously developed classifiers, with a bi-directional LSTM outperforming them when tested on data from the same genre. In addition, the authors analyse in greater detail the difficulty of detecting negation scope by testing the model on data of different genre (a set of negative sentences extracted from Simple English Wikipedia and annotated according to the guidelines released during the *SEM2012 shared task) and find that the performance of word-embedding features is comparable to that of more fine-grained syntactic feature.

Qian et al. (2016) propose a convolutional neural network (CNN)-based model with probabilistic weighted average pooling to address negation scope detection. It first extracts path features from syntactic trees with a convolutional layer and concatenates them with their relative positions into one feature vector, which is then fed into a soft-max layer to compute the confidence scores of its location labels. The system is trained on the abstract sub-collection of the BioScope corpus. Experimental results show that the proposed model gets the second highest performances for negation scopes on abstracts. It achieves comparable results on clinical reports and performs worse on full papers.

Lazib, Zhao, Qin and Liu (2016) regard the task of negation scope detection as a token-level sequence-labelling problem. They propose different models based on recurrent neural networks (RNNs) and word embedding that can be successfully applied to such tasks without any task-specific feature engineering efforts. Experimental results show that RNNs, without using any hand-crafted features, outperform feature-rich CRF-based model.

2.3.5 Other works

The approaches aimed at identifying negation in other languages are few in number. Some of them are oriented to adapt NegEx (W. W. Chapman et al., 2001b) and ConText (Harkema et al., 2009) algorithms into these different languages.

W. W. Chapman et al. (2013) ported and evaluated NegEx on clinical texts in Swedish (Skeppstedt, 2011) and French (Deléger & Grouin, 2012). This adaptation performs well (recall 82%; precision 75%) for Swedish assessment sections of the Stockholm EPR corpus (Dalianis, Hassel, Henriksson, & Skeppstedt, 2012) and even better (recall 85%; precision 89%) for French cardiology notes. In both studies, they achieve comparable recall to the English NegEx but with observable differences in precision (differences of −9.3% and 4.4%, respectively). Error analyses from these studies suggest that increasing lexicon coverage improves scope detection. Also for Swedish, Skeppstedt (2011) adapts NegEx, resulting in a precision of 75.2% and a recall of 81.9% when evaluated on 558 manually classified sentences containing negation triggers, and a negative predictive value of 96.5% when evaluated on 342 sentences not containing negation triggers. The triggers used for the evaluation of the Swedish adaptation of NegEx are available at <http://people.dsv.su.se/~mariask/resources/triggers.txt> and can be used together with the original NegEx program for negation detection in Swedish clinical text. Velupillai et al. (2014) successfully port pyConTextNLP (B. E. Chapman, Lee, Kang, & Chapman, 2011), an extension of the ConText algorithm, to Swedish (pyConTextSwe) by creating an optimised assertion lexicon for clinical Swedish. They integrate cues from four external lexicons, along with generated inflections and combinations, using subsets of a clinical corpus in Swedish. Four assertion classes (definite existence, probable existence, probable negated existence and definite negated existence) and two binary classes (existence yes/no and uncertainty yes/no) are applied to pyConTextSwe. The system's final F-scores on an evaluation set are 81%. For the individual assertion classes, F-score results are 88% (definite existence), 81% (probable existence), 55% (probable negated existence), and 63% (definite negated existence). For the binary classifications existence yes/no and uncertainty yes/no, final system performance is 97%/87% and 78%/86% F-score respectively.

English ConText algorithm is also adapted into the Dutch language (Afzal et al., 2014). To do this, the authors create a Dutch clinical corpus containing four types of anonymised clinical documents: entries from general practitioners, specialists' letters, radiology reports, and discharge letters. Using a Dutch list of medical terms extracted from the UMLS (Lindberg, Humphreys, & McCray, 1993), the medical terms in the corpus with exact matches are identified and annotated for negation, temporality, and experiencer properties. In addition, the authors translate English trigger terms into Dutch and add several general and document-specific enhancements, such as negation rules for general practitioners' entries and a regular-expression-based temporality module. For the negation property, the algorithm obtains an F-score from 87% to 93% for the different document types.

In line with this research, Cotik et al. (2016) adapt NegEx with the aim of detecting negation and speculation information in German. Two approaches are introduced: a dictionary look-up algorithm, which is taken as a baseline, and an approach based on a revised version of an existing German NegEx trigger set. Tests are also performed with triggers that were previously translated into German. The system has been tested on two different data sets: German discharge summaries and German clinical notes in the nephrology domain. In both cases, the German NegEx system outperforms the baseline and achieves an F_1-Score above 90%. However, applying NegEx to other text types might turn out to be more challenging. Another interesting approach is the one developed by Gros and Stede (2013), called Negtopus, which identifies negations and their scope in medical diagnoses written in German and in English. However, Negtopus currently focuses only on negation terms.

For the Chinese language, we find a greater number of works. Z. Chen, Zou, Zhu, and Li (2013) develop a supervised machine learning method with CRFs to detect negative information in scientific literature. They also evaluate the effectiveness of each feature under the character-based and word-based framework, as well as of the combination of features. Experimental results show that the single-word feature and the PoS feature are effective, and the combined features improve performance most significantly. This Chinese negation-detection system achieves an accuracy of 94.70%. Y. Zhang et al. (2014) develop an algorithm to detect negative expressions in Chinese EHR, combining rules with word co-occurrences. In the experiments, 200 medical texts comprising 150,865 Chinese characters are used to test the method. The negative predictive value is 99.85%. The specificity of the system is not reported, however. Z. Jia et al. (2014) propose an algorithm called NegDetector for locating the concerned clinical terms mentioned in EHR and for detecting whether the particular terms appearing in different positions are negated or affirmed. The algorithm infers the status of a condition with regard to the property from simple lexical clues occurring in the context of the condition, sometimes more than a few words away from the term. When evaluating the system by testing case history, it shows a recall of 99.85%, a precision of 94.98% and fallout of 51.47%. Zou, Zhu, and Zhou (2015) construct a Chinese corpus for negation and speculation identification (CNeSp), which annotates cues and their linguistic scopes. This corpus is divided into three sub-corpora: product reviews, financial articles and computer-related articles. For cue detection, the authors present a feature-based sequence-labelling model, in which the morpheme feature is employed to catch the composition semantics in the Chinese words more reliably. Complementally, a cross-lingual cue-expansion strategy is proposed to increase the coverage in cue detection. For scope resolution, a syntactic structure-based framework to identify the linguistic scope of a cue is presented. Evaluation shows that this approach outperforms the state-of-the-art chunking ones on negation identification in the Chinese language. Kang et al. (2017) construct a Chinese clinical corpus consisting of admission and discharge summaries, and propose sequence-labelling-based systems for negation and scope detection. These systems rely on features from bag-of-characters models, bag-of-words models, character embedding and word embedding. For scopes, they introduce an additional feature to handle nested scopes with multiple negations. In cue detection, these approaches are able to achieve a performance as high as 99% when measured by F-score, significantly outperforming their rule-based counterparts (79%). The best system uses word embedding as a feature, which yields a precision of 99% and recall of 99.1%. In scope detection, it achieves a performance of 94.6% in terms of F-score. Experimental results demonstrate that word embedding is effective in identifying negations, and that nested scopes can be identified effectively. H. He, Fancellu and Webber (2017) also address the problem of automatically detecting the negation cues in Chinese. In particular, they investigate whether character-based neural networks are able to achieve a similar or better performance than previous highly engineered sequence classifiers. They use the CNeSp corpus in the experiments. Results confirm that these models can be a valid alternative to previous ones, although they still suffer from over-generating the negation cue.

For Japanese, Mizuno et al. (2015) propose a combination of a supervised classifier and clusters of n-grams derived from 115,125 tweets posted during a one-month period after the Great East Japan earthquake, with the aim of recognising the negation of predicates on Twitter to identify the tweets that rebut false rumours. The authors show that the n-gram clusters improve the F-score by about 22% for complex forms of negations.

Finally, for Spanish, Costumero, Lopez, Gonzalo-Martín, Millan, and Menasalvas (2014) adapt the NegEx algorithm to detect negation regarding diseases in medical documents. They translate the list of terms previously identified for English in NegEx, enriching it with synonyms and terms extracted from the manual annotation of medical texts in Spanish. This adaptation is evaluated on a corpus of 500 reports where 267 unique clinical conditions are identified, achieving an accuracy for negated terms of 83.37%. Stricker et al. (2015) also carry out an adaptation of NegEx into Spanish (SpRad-Neg), and compare their results with those of the approach taken by Costumero et al. (2014). This adaptation differs from the previous one in the clinical domain of the texts (multiple vs. radiology) and the length of the texts (5 lines vs. about 20 lines). They use two different datasets: a set of reports of ultrasonography studies performed in a public hospital, in which the findings are annotated as affirmed or negated, and the corpus used originally by Costumero et al. (2014), which the authors extract from SciELO (Packer, 1998). When they test SpRadNeg in radiology reports, it achieves precision of 87% and recall of 49%. The comparison of both NegEx adaptations with SciELO data shows that Costumero et al.'s adaptation achieves better results. This might be because their triggers are better adapted to their domain. In terms of the performance of SpRadNeg on the two corpora the authors employ, the experiment on SciELO data resulted in a higher accuracy, recall and F_1-score than the experiment on the radiology corpus.

Following this research, Cotik, Stricker, Vivaldi and Rodriguez (2016) present a system that detects negations in radiology reports written in Spanish, for which they develop a baseline lexical look-up algorithm that contains negation triggers identified by an expert radiologist. They also adapt NegEx. For this adaptation, they automatically translate the negation triggers into English and perform two different experiments. For the first, they use only the translated triggers, and for the second they combine the translated triggers, a set of bi- and tri-grams, and a list of triggers provided by an expert. They also build three sets of rules based on i) PoS patterns, ii) constituent tree patterns and iii) dependency tree patterns. The baseline system obtains a F_1-score of 75%, the one based on the dependency tree patterns achieves 81%, the constituent tree patterns system obtains a F_1-score of 90% and the NegEx adaptation and the PoS tagging pattern system achieves 92% for the F-measure.

2.4 Conclusions and chapter summary

This chapter is an overview of the concept of negation and the major topics concerning it. It is complex subject which has been studied for a long time. Negation dates back to Aristotle and in its most trivial level, it reverses the truth value of a preposition. However, in more subtle examples it is strongly expressive and includes irony and euphemisms.

In addition, this chapter has shown that negation detection has been an active research area during recent years in the NLP community, even inspiring some shared tasks in NLP-related conferences. In fact, negation detection constitutes a challenge in which many applications can benefit from identifying this kind of information (e.g., *recognising textual entailment, sentiment analysis, information extraction*). Main tasks have been focused on determining the negation cues and the resolution of their scope (i.e., identifying at sentence level which tokens are affected by the cues).

2.5 Further reading and relevant resources

Negation from a linguistic point of view could be studied through the theoretical work of Horn (1989), since it is considered a masterpiece. From a more computational perspective, the work of Morante and Sporleder (2012) could be a good introduction. Also relevant are the set of papers included in this special issue which can be downloaded at <https://www.mitpressjournals.org/toc/coli/38/2>. Other interesting resources to read are the documentation associated with the tutorials given by Roser Morante at IJCNLP 2011[2] and Noa Cruz at RANLP 2017.[3] The latter is focused on the biomedical domain.

There are six top-level NLP conferences that include negation detection among their topics (all of them are core A conferences, except ACL, which is core A*): ACL – Association for Computational Linguistics,[4] EMNLP – Empirical Methods in Natural Language Processing,[5] NAACL – North American Chapter of the Association for Computational Linguistics,[6] EACL – European Chapter

2. <https://www.aclweb.org/mirror/ijcnlp11/downloads/tutorial/tu3_present.pdf>

3. <https://www.noacruz.com/resources/>

4. <https://www.aclweb.org/portal/>

5. <http://emnlp2018.org/>

6. <http://naacl.org/>

of the Association for Computational Linguistics,[7] COLING – International Conference on Computational Linguistics,[8] CoNLL – Conference on Natural Language Learning.[9] There are also relevant conferences in the related fields of information retrieval, artificial intelligence, machine learning, and data mining: SIGIR – Special Interest Group on Information Retrieval,[10] AAAI – Association for the Advancement of Artificial Intelligence,[11] ICML – International Conference on Machine Learning,[12] ICDM – International Conference on Data Mining.[13]

Computational Linguistics is the leading journal in this field and it is freely available at <http://www.mitpressjournals.org/loi/coli>. Other related journals with a high impact factor according to the *Journal Citation Reports* (Garfield, 1991) are *Language Resources and Evaluation,*[14] *Information, Processing & Management,*[15] *Computer Speech and Language,*[16] *Journal of the Association for Information Science and Technology*[17] and *Natural Language Engineering.*[18]

Freely available datasets provided in past competitions could be used to train and test the negation detection systems and could also serve as a benchmark to compare the performance with that obtained by the rest of the participants: for example, the datasets from the SEM 2012 shared task[19] or the BioNLP'09 Shared Task 3 (J. Kim et al., 2009).

In the case of negation in Spanish, a research group called NEGES <http://www.sepln.org/workshops/neges/index.php?lang=en> emerged in 2017 and is made up of researchers from the fields of computational linguistics and NLP who

7. <http://www.eacl.org/page.php?id=index>
8. <https://coling2018.org/>
9. <http://www.signll.org/conll/>
10. <http://sigir.org/>
11. <http://www.aaai.org/>
12. <https://icml.cc/>
13. <http://icdm2018.org>
14. <https://link.springer.com/journal/10579>
15. <https://journals.elsevier.com/information-processing-and-management/>
16. <https://www.journals.elsevier.com/computer-speech-and-language>
17. <https://onlinelibrary.wiley.com/journal/23301643>
18. <https://www.cambridge.org/core/journals/natural-language-engineering>
19. <https://www.clips.uantwerpen.be/sem2012-st-neg>

aim to contribute to the ongoing research on negation in Spanish in the language technology community. NEGES promotes interest in negation detection in Spanish, provides members with a means of exchanging news of recent research developments and other matters of interest as well as making resources relevant to negation detection in Spanish available, including corpora, annotation guidelines, evaluation scripts, etc. NEGES' activities include the holding of an annual meeting each September at the International Conference of the Spanish Society for Natural Language Processing (SEPLN). In 2018, the workshop includes three different tasks: to reach an agreement on the guidelines to follow for the annotation of negation, to develop a system able to identify all the negation cues present in a document and to evaluate the role of negation in sentiment analysis. If the reader is interested in joining the NEGES group, he/she can simply subscribe to the distribution list l-neg-sp by filling in the details of the form at <https://listas.ujaen.es/mailman/listinfo/l-neg-sp>.

Concerning the detection of negation in Spanish, it is worth noting that one of the biggest challenges is the phenomenon known as double negation (Jespersen, 1917). English does not have double negation (Wang, 2006). To know more about this peculiarity of Spanish, the reader is referred to the work of Martí et al. (2016).

Finally, there are new trends in negation processing of which the reader might want to know more details. For example, van Miltenburg, Morante and Elliott (2016) analyse the descriptions containing negations in the Flickr30K corpus and a categorisation of negation uses, such as the description of an unexpected event in an image. Based on that analysis, the authors provide a set of requirements that an image description system should have in order to generate negation sentences. This work is continued in van Miltenburg, Elliott, and Vossen (2017) in which the authors carry out a cross-linguistic comparison of Dutch, English, and German image descriptions. In addition, alongside the biomedical and review domains, other fields have been investigated recently. Cheng, Baldwin, and Verspoor (2017) follow the methods proposed by Agarwal and Yu (2010) to implement a simple CRF approach for detecting negation cues and scope in the veterinary clinical note domain. The authors train the model over VetCompass training data, the BioScope corpus, or both. VetCompass is a corpus annotated for negation and speculation, constructed from a random sample of one million clinical records from VetCompass UK[20] and annotated following the Bioscope corpus guidelines. This is a novel approach since, as the authors affirm, they are only aware of a few papers that apply NLP in the veterinary domain.

20. <http://www.rvc.ac.uk/VetCOMPASS>

Suggestions for further reading

Cheng, K., Baldwin, T., & Verspoor, K. (2017). Automatic Negation and Speculation Detection in Veterinary Clinical Text. In *Proceedings of the Australasian Language Technology Association Workshop 2017* (pp. 70–78).

Garfield, E. (Ed.). (1991). *Journal citation reports*. Institute for Scientific Information.

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

Jespersen, O. (1917). *Negation in English and other languages*. AF Høst.

Martí, M. A., Taulé, M., Nofre, M., Marsó, L., Martín-Valdivia, M. T., & Jiménez-Zafra, S. M. (2016). La negación en español: análisis y tipología de patrones de negación. *Procesamiento del Lenguaje Natural*, (57), 41–48.

van Miltenburg, E., Morante, R., & Elliott, D. (2016). Pragmatic factors in image description: the case of negations. *arXiv preprint arXiv*:1606.06164.

van Miltenburg, E., Elliott, D., & Vossen, P. (2017). Cross-linguistic differences and similarities in image descriptions. *arXiv preprint arXiv*:1707.01736.

Morante, R., & Sporleder, C. (2012). Modality and negation: An introduction to the special issue. *Computational Linguistics*, *38*(2), 223–260. https://doi.org/10.1162/COLI_a_00095

The tutorial given by Noa Cruz at RANLP 2017[21] focused on negation in the biomedical domain. <https://www.noacruz.com/resources/>.

Wang, C. (2006). La negación en español, chino, inglés y alemán. *Encuentros en Catay*, 20, 129–157.

21. <https://www.noacruz.com/resources/>

CHAPTER 3

Speculation

Chapter 3 goes into detail about the concept of speculation and presents a summary of the most representative works in this field, which includes the different approaches followed by the authors in order to solve the problem highlighted.

3.1 Defining speculation

The phenomenon of speculative language should be studied within the framework of modality since it is related to, among other things (i.e., subjectivity, evidentially, uncertainty, committed belief, and factuality), the related concept of speculation.

Generally speaking, modality is what allows speakers to attach expressions of belief, attitude and obligation to statements. Morante and Sporleder (2012) give a good overview of the concept of modality from which one can conclude that modality can be defined as a philosophical concept, as a subject of study in logic or as a grammatical category. From a philosophical point of view, Von Fintel (2006) defines modality as a category of linguistic meaning having to do with the expression of possibility and necessity. He explains that there are different types of modal meaning (i.e., alethic, epistemic, deontic, bouletic, circumstantial and teleological) which can be conveyed by several types of expressions such as *conditionals*, *adjectives*, *nouns*, *adverbs*, *modal auxiliaries* and *semimodal verbs*. Next, within the modal logic framework, Kratzer (1981) analyses modality in terms of possible-world semantics, where a proposition is identified with the set of possible worlds where it is true. She notes that the interpretation of modals should consider a conversional background, which implies that the meaning of modal expressions is context-dependent. Finally, from a grammatical perspective, Palmer (2001) defines modality as a valid cross-language grammatical category which is similar to aspect or tense, since all three are categories of the clause and are concerned with the event or situation that is conveyed by the statement. He considers that speculation falls within the category of epistemic modality because it is the means by which speakers express judgement about the factual status of a proposition (e.g., *John may be in his office.*)

The notion of speculation, also known as hedging, was first introduced by Lakoff (1972). He defines it as words whose job is to make things fuzzier or less fuzzy. Other definitions are rare in the literature. They include Zuck and Zuck (1986), who define hedging as the process whereby the authors reduce the strength of a statement, and Markkanen and Schröder (1989), who consider it as a manipulative, non-direct sentence strategy of saying less than one means. Hyland (1995) refers to speculation as the expression of tentativeness and possibility in language use. He extensively studies the topic, focusing on scientific texts where statements are rarely made without subjective assessments of truth. In Hyland (1998), he explains that modality can be seen as any linguistics means used to indicate either a lack of complete commitment to the truth value of any accompanying proposition or a desire not to express that commitment categorically. He also argues that speculation is part of epistemic modality because it indicates an unwillingness to make an explicit and complete commitment to the truth of propositions. He establishes a categorisation of hedge cues, dividing them into lexical and non-lexical ones. Lexical cues include modal auxiliaries like *may* and epistemic modality, such as judgment verbs (e.g., *suggest*), evidential verbs (e.g., *appear*), deductive verbs (e.g., *conclude*), adjectives (e.g., *probable*), adverbs (e.g., *possibly*) and nouns (e.g., *suggestion*). Non-lexical hedges usually include reference to limiting experimental conditions, reference to a model or theory or admission to a lack of knowledge. Others authors to study hedging in the scientific domain include Light, Qiu and Srinivasan (2004) and Medlock and Briscoe (2007).

What does seem clear is that, like negation, speculation is a challenging phenomenon from a computational point of view. Two main tasks have been addressed in the computational linguistic community, namely the detection of hedge cues and the resolution of the scope of these cues. For instance, in (2), the speculation cue is the token *could* while its associated scope is the syntagma *happen to him is an industrial accident*.

(2) The best thing that could happen to him is an industrial accident.

As this example shows, speculation cues are linguistic devices that reveal the author's attitude or opinion by presenting information as uncertain or unreliable within the text (Verbeke et al., 2012). Hedge keywords can be expressed by different word classes and by multiword expressions (i.e., expressions that contain more than a word and whose meaning cannot be derived from the individual meanings of the words that constitute the expression) such as *cannot be excluded*. In addition, it becomes crucial to know, at sentence level, which words are affected by the cues.

Finally, just as with negation, speculative language is extensively used. Hyland (1996) reports one hedge in every 50 words of a corpus of research articles. Light et al. (2004) mention that 11% of the sentences in MEDLINE (Miller, Lacroix, & Backus, 2000) contain speculative language. Szarvas et al. (2008) explain that about 18% of the sentences in the abstract section and about 20% of sentences in the full papers sub-collection of the BioScope corpus feature speculation. In the review domain, Konstantinova et al. (2012), show that the proportion of speculative information in the Simon Fraser University (SFU) Review corpus is 22.7%.

3.2 Speculation detection

A fair amount of literature on hedging in scientific texts has been published since the 1990s. For instance, Friedman, Alderson, Austin, Cimino, and Johnson (1994) discuss uncertainty and hedging in radiology reports. Their system assigns one of five levels of certainty (namely *no certainty*, *low certainty*, *moderate certainty*, *high certainty* and *cannot evaluate*) to extracted findings.

However, speculative language from a Natural Language Processing (NLP) perspective has only been studied in the past few years. The first approaches focused on classing speculative sentences according to whether they contain speculation cues or not. Light et al. (2004) introduce the problem of dealing with speculation using their own list of hedge cues to identify speculative sentences in MEDLINE abstracts. They also experiment with automated methods, proposing two different systems: one based on support vector machine (SVM), the other based on substring matching. The latter system marks as speculative those sentences that contain any of the following substrings: *suggest*, *potential*, *likely*, *may*, *at least*, *in part*, *possible*, *potential*, *further investigation*, *unlikely*, *putative*, *insights*, *point toward*, *promise* and *propose*. Both the substring and the SVM system perform well. The SVM classifier results are higher than those yielded by the substring matching method in terms of precision (84% vs. 55%). The opposite is the case in terms of recall, where SVM performs worse (39% vs. 79%).

Medlock and Briscoe (2007) draw on this work and investigate the automatic classification of speculative language using loosely supervised machine learning. They implement a simple probabilistic model for acquiring training data. This learner returns a labelled data set for each class, on which the probabilistic classifier is trained. The training corpus consists of 300,000 randomly selected sentences, while the authors manually annotate six full-text papers from the functional genomics literature relating to *Drosophila melanogaster* (the fruit fly) to form the test corpus. They have since made this dataset publicly available.[1] The system outstrips the baseline classifier described in Light et al. (2004) by 16% in terms of precision/recall break-even point (BEP). Error analysis shows that the model is unsuccessful in identifying assertive statements of knowledge paucity, which are generally marked rather syntactically than lexically. The classifier also

1. The dataset is available at <http://www.benmedlock.co.uk/hedgeclassif.html>.

has difficulties in distinguishing between a speculative affirmation and one relating to a pattern of observed non-universal behaviour. Medlock (2008) extends this work and experiments with additional features, namely part-of-speech (PoS) tags, stems and bigrams. According to the results, they explain that adding PoS and stem features to a bag-of-words input representation can slightly improve accuracy. Adding bigrams produces a statistically significant improvement over a bag-of-words representation. The best result outperforms the results previously obtained in Medlock and Briscoe (2007): 76% vs. 82% precision/recall BEP.

Szarvas (2008) follows Medlock and Briscoe in classifying sentences as either speculative or non-speculative. He extends their research by using a Maximum Entropy classifier which incorporates bigrams and trigrams as features, performs a re-ranking-based feature-selection procedure, and exploits external dictionaries. In the experiments, he uses the dataset gathered by Medlock and Briscoe (2007) as a learning source. At the same time, he makes available the BMC Bioinformatics data set (by annotating four full-text papers from the open-access BMC Bioinformatics website), which is used for evaluation purposes. He also investigates hedging in radiology reports. His best configuration (i.e., performing manual and automatic feature selection consecutively and using external dictionaries) achieves a precision/recall BEP performance of 85.29% and an F-score of 85.08% on the biomedical papers. This configuration yields lower results on radiology reports (82.07% in terms of F-score). The error analysis indicates that more complex features like *dependency structure* and *clausal phrase information* could only help in allocating the scope of hedge cues detected in a sentence, not the detection of any of these cues itself.

Kilicoglu and Bergler (2008) apply a linguistically motivated approach to the same classification task by using knowledge from existing lexical resources and by incorporating syntactic patterns. Additionally, hedge cues are weighted by automatically assigning an information-gain measure and by assigning weights semi–automatically depending on their types and their centrality to hedging. The system is evaluated on two different datasets: the Drosophila data set from Medlock and Briscoe (2007) and the annotated BMC Bioinformatics papers from Szarvas (2008). In the first data set, the authors' approach achieves a competitive precision/recall BEP of 85% using the semi-automatic weighting scheme. On the BMC dataset, it yields a precision/recall BEP of 82%. The results confirm that the selection of hedging devices affects the speculative strength of the sentence, which can be measured accurately by weighting the hedge cues. Error analysis reveals that false-positive errors are caused by the word-sense ambiguity of speculation cues such as *could*, and by weak hedge cues like some adverbs (e.g., *usually*), normalisations (e.g., *implication*) and epistemic deductive verbs (e.g., *conclude*). False-negative errors arise because the method does not address syntactic patterns and fails to identify certain derivational forms of epistemic words.

Shatkay, Pan, Rzhetsky, and Wilbur (2008) introduce a novel task, in which they classify sentence fragments from biomedical texts along five different parameters. One of the parameters is the degree of certainty: each statement is assigned a value between 0 and 3, with 0 indicating no certainty and 3 indicating absolute certainty. Another of the parameters measures polarity, i.e. whether the statement is negated or not. The authors annotate a corpus of 10,000 sentences and sentence fragments selected from full-text articles from different biomedical journals. Using an SVM classifier, the results for certainty vary from 99% for level 3 to 46% for level 2, both in terms of F-score. Results for polarity classification are 95% F-score for the negative class and 100% F-score for the positive.

Ganter and Strube (2009) are the first authors to exploring the following new domain. They develop a system for the automatic detection of Wikipedia sentences that contain weasels. They adopt Wikipedia's definition of weasel words (i.e., words and phrases aimed at creating an impression that something specific and meaningful has been said, when only a vague or ambiguous claim has in fact been made) since they are closely related to hedges and private states. The authors experiment with two different classifiers, one based on word frequency measures and another one based on syntactic patterns. Both approaches perform comparably well (around 70% precision/recall BEP), so word frequency and distance to the weasel tag are sufficient. The experiments also show that syntactic patterns work better when using a broader notion of hedging tested on manual annotations.

In 2008, the availability of the BioScope corpus (Szarvas et al., 2008): the set of clinical free-texts, of biological texts from full papers and of scientific abstracts annotated for negation, speculation and their linguistic scope, facilitated the development of corpus-based statistical systems for negation/hedge detection. Since it was made publicly accessible, many works have been carried out using it as a training and evaluation source. In this sense, the task of resolving the cues and scope of speculation was first introduced in Morante and Daelemans (2009a). They port the system initially designed for negation detection (Morante & Daelemans, 2009b) described in Section 2.3.2, to speculation. In the first phase, hedge cues are identified by a set of classifiers, and in the second stage, another set of classifiers are employed to detect the scope of the speculation keyword. They show that the same scope-finding approach can be applied to both negation and hedging. The F-score of speculation detection for clinical documents is 38.16%, with 26.21% of scopes correctly identified. For papers, the F-score is 59.66%, and 35.92% of scopes are correctly predicted. The F-score for abstracts is 78.54% and the percentage of scopes correctly classified is 65.55%. Özgür and Radev (2009) develop a supervised classifier for identifying speculation cues and a manually compiled list of lexico-syntactic rules for identifying their scopes. For the performance of the rule-based system when identifying speculation scopes, the authors report accuracy of 61.13% and 79.89% for the BioScope full papers and abstracts, respectively.

Using the same corpus, other authors, such as Ballesteros et al. (2012), have also taken into account speculation in their systems which, in most cases, were initially designed for negation. For example, Agarwal and Yu (2010) report F-scores of 88% and 86% when detecting speculation cue phrases and their scope in biological literature and F-scores of 93% and 90% in clinical notes. However, as with negation, their approach is not directly comparable owing to the fact that they use different corpus partitions and evaluation measures. The system developed by Apostolova et al. (2011) reports F-scores of 75.57% for clinical documents, 78.99% for papers and 73.87% for abstracts in the scope-recognition task. This means that their system outperforms the baseline results, as is the case in the negation detection task. Cruz Díaz et al. (2012) report a performance value of 94.9% when detecting the cues and 80.9% resolving the scope (with gold-standard cues) in the clinical sub-collection, both in terms of F-score. The approach presented by Z. Chen et al. (2013) yields F-score values of 84.21% for abstracts, 67.24% for papers and 72.92% for clinical texts in the scope detection phase (using as cues those that appear annotated as such in the corpus). Finally, Qian et al. (2016) show that their convolutional neural network (CNN)-based model gets the best performances for speculation scopes on the abstracts sub-corpus, and achieves comparable performances on clinical texts.

This increased attention for speculation detection is reflected in the fact that it has become a subtask of the BioNLP Shared Task in 2009, and the topic of the Shared Task at CoNLL-2010 (Farkas et al., 2010). The latter comprises two tasks: Task 1 is dedicated to detecting uncertain sentences in two different domains, biological publications and Wikipedia articles. Task 2 aims to resolve in-sentence uncertainty detection, i.e., it automatically annotates the cue phrases and the left and right boundaries of their scope. In this case, the training and evaluation data consist of biological texts.

In Task 1, the best system for Wikipedia data is the system developed by Georgescul (2010). It obtains an F-score of 60.2%. For biological documents, Tang, Wang, Wang, Yuan, and Fan (2010) report a performance of 86.4% in terms of F-score. Both approaches handle the task as a classical sentence-classification problem and essentially employ a bag-of-words feature representation. In addition, neither system derives features from syntactic parsing. However, many authors tackle the task as a word-by-word token classification problem, i.e., they focus on the cue phrases and seek to classify every token if it is a part of a cue phrase, then a sentence is predicted as uncertain if it contains at least one recognised cue phrase. Examples are the approaches of Velldal, Øvrelid and Oepen (2010) and Vlachos and Craven (2010).

Task 2 is implemented by all the authors as a two-stage architecture where the speculation cues are first detected and, then, the scope associated with these cues is predicted. Tang et al. (2010) report the best result on hedge cue recognition, with an F-score of 81.3%. Similarly to Morante and Daelemans (2009a), they set out to label words according to a Begin/Inside/Outside-scheme (BIO-scheme), i.e., determining whether the token is at the beginning, inside or outside of a hedge cue. They use a cascade subsystem in which a conditional random fields (CRF) model and a large margin-based model are trained. Then, another CRF classifier is trained using the result of the first predictions. For scope detection, the best F-score (57.3%) is obtained by Morante, Van Asch and Daelemans (2010). They introduce a number of changes to the approach described in Morante and Daelemans (2009a): they use one classifier per task instead of a metalearner combining three classifiers; information is added from the dependency tree instead of using only shallow features and a better treatment of multiword cues is carried out. Rei and Briscoe (2010) combine a set of manually compiled rules, a CRF classifier, and a sequence of post-processing steps on the same task, obtaining the second best result. Finally, Velldal et al. (2010) develop handcrafted rules based on syntactic information taken from dependency structures. With this approach, they achieve an F-score of 55.3%, the third best for the task.

In terms of both combined ranks and average F_1-score, Øvrelid, Velldal and Oepen (2010) attain the best overall result. They show how the use of syntactic structure enables the resolution of hedge scope in a hybrid, two-stage approach to uncertainty analysis. In the first stage, a Maximum Entropy classifier, combining surface-oriented and syntactic features, identifies cue words. In stage two, a small set of manually created rules operating on dependency representations is used.

In the wake of the CoNLL-2010 Shared Task, many systems using the same corpora for training and evaluation have been created and described in the literature. Velldal (2011) presents a solution which builds on the token classification approach described by Velldal et al. (2010) but is set within the framework of SVM classification instead of Maximum Entropy. He shows how better results can be obtained by approaching the task as a disambiguation problem, restricting the attention only to those tokens whose base forms have previously been observed as hedge cues. Reformulating the problem in this way simplifies the classification task, reducing the number of examples that need to be considered, and thereby also trimming down the relevant feature space to a much more manageable size. The resulting feature space is still huge; however, the author applies the method of random indexing, further reducing the dimensionality of the feature space by two orders of magnitude. The system achieves an F_1-score of 86.64% in Task 1. Read, Velldal, Oepen, and Øvrelid (2011) propose an SVM-based discriminative ranking function for selecting subtrees from head-driven phrase structure grammar (HPSG)-based constituent structures, showing that while this technique achieves good performance on its own, combining it with an existing rule-based system operating on dependency parses improves performance beyond either of them in isolation.

In the cue-detection phase, Velldal et al. (2012) present a greatly simplified method for cue identification using a linear SVM classifier. This is accomplished by treating the set of cue words as a closed class. This means that, in line with Velldal (2011), the classifier only attempts to disambiguate known cue words, ignoring any words not observed as cues in the training data. In the scope-recognition phase, they employ a set of rules on syntactic features and n-gram features of surface forms and lexical information together with a machine-learning system that selects subtrees in constituent structures. The F-score achieved by this system is 59.4%. Kenji and Tanaka-Ishii (2014) provide a comprehensive summary of the methods and results of the system used in the CoNLL-2010 Shared Task. In addition, they propose a simple yet effective cue-selection algorithm which minimises hedging error and does not require cue annotation. Unlike previous works, the proposed method focuses on cue selection, decoupling it from disambiguation and by optimising it over the sentence-hedging error rate. The task performs well in experiments, even for settings with poor disambiguation, without cue annotation and with otherwise unreliable corpora from a machine-learning point of view. Li, Gao, and Shavlik (2014) empirically explore three fundamental issues of uncertainty detection: (1) the predictive ability of different learning methods on this task; (2) whether using unlabelled data can lead to a more accurate model; and (3) whether closed-domain training or cross-domain training is better. For these purposes, the authors adopt two statistical learning approaches: the commonly used bag-of-words model based on Naïve Bayes, and the sequence-labelling approach using a hidden Markov model (HMM). The results are promising: (1) on Wikipedia and biomedical datasets, the HMM model improves on Naïve Bayes by up to 17.4 percentage points and 29.0 percentage points, respectively, in terms of absolute F-score; (2) compared to CoNLL-2010 systems, their best HMM model achieves an F-score of 62.9% with maximum likelihood parameter estimation (MLE) and 64.0% with expectation-maximisation parameter estimation (EM) on the Wikipedia dataset. The results on the biomedical dataset are less impressive; (3) when the expression ability of a model (e.g., Naïve Bayes) is not strong enough, cross-domain training is helpful, and when a model is powerful (e.g., HMM), cross-domain training may produce biased parameters; and (4) under MLE, combining the unlabelled examples with the labelled ones helps.

Moncecchi, Minel and Wonsever (2014) tackle the task using a learning methodology that proposes the use of an iterative, error-based approach in order to improve classification performance. They analyse how the incorporation of syntactic constituent information to the learning and post-processing steps produces a performance improvement of almost 12 points in terms of F-score over previously unseen data. Zhou, Deng, Huang and Zhu (2015) highlight that previous hedge-scope detection methods usually take all tokens in a sentence as candidate boundaries, which inevitably generates a large number of negatives for classifiers. These imbalances mislead classifiers considerably and result in lower performance. Therefore, the authors propose a dependency-based candidate boundary selection method (DCBS), which selects the most likely tokens as candidate boundaries and removes the additional tokens which have less potential to improve the performance, based on the dependency tree. In addition, they employ the composite kernel to integrate lexical and syntactic information and demonstrate the effectiveness of structured syntactic features for hedge-scope detection. Experiments on the CoNLL-2010 Shared Task corpus show that this method achieves an F_1-score of 71.92% on the gold-standard cues.

For sentiment analysis, as mentioned in Section 1.1, distinguishing between objective and subjective facts is crucial since speculation is a linguistic expression that tends to correlate with subjectivity (also known as private state). For instance, authors such as Benamara et al. (2012) have studied the effect of speculation on opinion expressions according to their type (i.e., buletic, epistemic and deontic). They highlight that, as is the case with negation, each of these types has a specific effect on the opinion expression in its scope and this information should be used as a feature in a machine-learning setting for sentence-level opinion classification. However, although it has been proven that speculation has an effect on opinion expressions and it should be taken into account; there are only a small number of works focused on the detection of speculation in the review domain. This is due to the fact that the annotation of a corpus with this kind of information, which would make it possible to tackle this problem efficiently, has only been done recently, in the SFU Review corpus (Konstantinova et al., 2012). Using this corpus, Cruz et al. (2015) present the first attempt to detect speculation in the review domain. They propose a machine-learning system which works in two stages: in the first stage, speculation cues are identified, and in the second, the full scope of these cues is determined. The performance obtained in the cue-prediction phase is close to that obtained by a human carrying out the same task. In scope detection, the results are also promising and represent a substantial improvement on the baseline (up by roughly 10%). Skeppstedt, Schamp-Bjerede, Sahlgren, Paradis, and Kerren (2015) compare an SVM classifier to a lexicon-based approach for the task of detecting the stance categories *speculation*, *contrast* and *conditional* in English consumer reviews. Around 3,000 training instances are required to achieve a stable performance of an F-score of 90%. This outperforms the lexicon-based approach, for which an F-score of just above 80% is achieved. The machine-learning results for the other two categories show a lower average (an approximate F-score of 60% for contrast and 70% for conditional). For detecting sentences with speculation, an SVM model trained on bag-of-words/bigrams performs around 10 points better than a lexicon-matching approach. With 3,000–5,000 training instances, the model performance is stable at an approximate F-score of 90%, which is just above the inter-annotator agreement F-score. For detecting conditional sentences and sentences including contrast, however, the results are lower (an F-score of around 60% for contrast and around 70% for conditional).

Recent research in the detection of speculation has included the use of more than one corpus as a source of training and testing. As with negation, Skeppstedt et al. (2016) train machine-learning models to recognise markers in the BioScope corpus (Szarvas et al., 2008) and the SFU Review corpus (Konstantinova et al., 2012). In both corpora, speculation markers are detected with results close to previously reported annotator-agreement scores. Also, the strategy of training the model on the SFU Review corpus and evaluating it on the BioScope corpus is more successful than the previously explored strategy of training a model on biomedical article texts and applying it on the clinical text genre (Morante & Daelemans, 2009a). There might thus be a greater similarity between how speculation is expressed in consumer reviews and in clinical texts than between clinical and biomedical texts. Also using these two corpora, Adel and Schütze (2016) present novel attention architectures for uncertainty detection: external attention and sequence-preserving attention. They conduct an extensive set of experiments with various configurations along different parameters of attention, including different focuses and sources of attention and sequence-agnostic vs. sequence-preserving attention. They apply recurrent neural networks (RNN) and CNN to this task for the first time. The CNNs with external attention improved on the state-of-the-art by more than 3.5 points in terms of F_1-score on a Wikipedia benchmark and perform similarly to the state-of-the-art model on a biomedical benchmark which uses a large set of linguistic features. The code is publicly available for future research at <http://cistern.cis.lmu.de>.

Jean, Harispe, Ranwez, Bellot, and Montmain (2016) propose an automatic machine-learning method for detecting uncertainty in natural language, using three corpora for training: the BioScope corpus (Szarvas et al., 2008), the WikiWeasel corpus (Farkas et al., 2010) and the SFU corpus (Konstantinova et al., 2012). This method is based on the selection of optimal features to represent a sentence as a concise vector representation. This representation is based on the analysis of global and local features at sentence level. The local features are built from a specific aggregation of different conditional probabilities of n-gram patterns weighted by a confidence score. The sentence length has been used as a global feature. The experimental analyses performed show that this approach obtains good results on all parameters of uncertainty and improve upon the best-known results on several datasets. An important component of the model proposed is the notion of confidence that can be associated with contextual observations. Indeed, in this study, the authors also propose and evaluate several confidence criteria that can be used to integrate statistical observations automatically while carefully considering their semantics to avoid overestimation. The empirical results reported show that the proposed estimator outperforms existing scores.

In the social media domain, Wei et al. (2013) construct the first uncertainty corpus based on tweets. The dataset was collected from Twitter during the summer riots in London of 6–13 August 2011, comprising 326,747 tweets in total. Search criteria include hashtags such as *#ukriots*, *#londonriots*, *#prayforlondon*. The authors further extracted the tweets relating to seven significant events during the riot as identified by UK newspaper *The Guardian* from this set of tweets. Finally, all 4,743 extracted tweets relating to the seven events were annotated. The authors also conduct experiments on the generated tweets corpus to study the effectiveness of different types of features for uncertainty text identification. In addition to n-gram features, they explore the effectiveness of three categories of social-media-specific features including content-based, user-based and Twitter-specific ones. Results show that the three categories of social-media-specific features can improve uncertainty identification. Furthermore, content-based features bring the highest improvement among the three and the presence of uncertain cue-phrase contributes most for content-based features.

As in English, the number of attempts to identify speculation in other languages is lower than those approaches developed to detect negation. In addition, there are languages such as Spanish or German for which there are no works on identifying speculation.

In the case of Swedish, Velupillai, Dalianis, and Kvist (2011) present an annotation model of six factuality levels linked to diagnoses in clinical assessments from an emergency ward (the Stockholm EPR Diagnosis-Factuality corpus), showing a fairly high overall agreement. Velupillai (2011) also proposes an automatic classifier using CRF, which is trained and tested on this corpus. The classifier obtains promising results (the best overall results are 69.9% average F-score using all classes, 76.2% F-score using merged classes), using simple local context features. Preceding context is more useful than posterior, although the best results are obtained using a window size of +/−4. Lower levels of certainty are more problematic than higher levels, which is also the case for the human annotators creating the corpus. Velupillai et al. (2014) port pyConTextNLP (B. E. Chapman et al., 2011) from English to Swedish (pyConTextSwe), as explained in Section 2.3.5.

For Chinese, Ji, Qiu, and Huang (2010), develop a system to detect speculation in Chinese news texts. However, only the speculative sentences are located, with no more fine-grain information such as scope. Z. Chen et al. (2013) develop a supervised machine-learning method with CRFs to detect speculative information in scientific literature. As with negation, they also evaluate the effectiveness of each feature under the character-based and word-based frameworks, as well as of the combination of features. Experimental results show that the single-word feature and the PoS feature are effective, and the combined features improve performance most significantly. This Chinese speculation-detection system achieves an accuracy of 87.10%. Zou et al. (2015) construct a Chinese corpus for negation and speculation identification (CNeSp) and propose a model to detect cues and their related scope, as described in Section 2.3.5. Evaluation shows the appropriateness of the syntactic-structure-based framework which shows significant improvement over the state-of-the-art of speculation identification in Chinese. Zhou et al. (2016) propose a novel syntactic and semantic information exploitation method for scope detection. A composite kernel model is employed to capture lexical and syntactic information. An long short-term memory (LSTM) model is used to explore semantic information. Furthermore, they exploit a hybrid system to integrate a composite kernel and an LSTM model into a unified framework. Experiments show that a composite kernel model effectively captures lexical and syntactic information, an LSTM model captures deep semantic information and their combination further improves the performance of hedge-scope detection. S. Zhang et al. (2016) propose a sequence-labelling-based system for speculation detection, which relies on features from bag-of-characters models, bag-of-words models, character embedding and word embedding. They compare the systems in which word embedding is calculated based on word segmentations given by general and domain-specific segmenters respectively. The systems are able to achieve a performance as high as 92.2% measured by F-score, demonstrating that word segmentation is critical to producing high-quality word embedding to facilitate downstream information extraction applications, and suggesting that a domain-dependent word segmenter can be vital to such a clinical NLP task in Chinese language.

The NTCIR-10 and NTCIR-11 MedNLP Tasks (Morita, Kano, Ohkuma, Miyabe, & Aramaki, 2013) were the first and second shared tasks to evaluate technologies that retrieve important information from medical reports written in Japanese. These tasks include three sub-tasks: a named-entity removal task (the de-identification task), a disease-name extraction task (the complaint and diagnosis task), and a normalisation task (the International Codes for Disease, or ICD, coding task). These sub-tasks include the extraction of modality attributes. Following the success of these MedNLP tasks, the NTCIR-12 MedNLPDoc Task (Aramaki, Morita, Kano, & Ohkuma,) comprises a new, challenging task where participants' systems infer disease names in ICD from textual medical records, including a possible code set which means that at least one coder is utilised.

In Hungarian, there have been a couple of recent attempts to detect uncertainty. Vincze (2014) presented the first machine-learning algorithm that aims to identify linguistic markers of uncertainty in Hungarian texts from two domains: Wikipedia and news media. The system is based on sequence labelling and makes use of a rich feature set including orthographic, lexical, morphological, syntactic and semantic features. Having access to annotated data from two domains, they also focus on the domain specificities of uncertainty detection by comparing results obtained in in-domain and cross-domain settings. The results show that the domain of the text has a significant influence on uncertainty detection. Vincze (2016) also experiments by identifying uncertainty cues in the Hungarian social media texts annotated by Vincze et al. (2014). The results indicate that the idiosyncrasies of social media texts should be accounted for when implementing an uncertainty detector. Also, selecting the training data has a significant effect on learning efficiency, but adding out-domain data to a small set of in-domain data can also contribute to performance. Moreover, differences among uncertainty cue types may also affect the efficiency of uncertainty detection and therefore some types of linguistic uncertainty may require special treatment in uncertainty detection.

Finally, for Arabian, Al-Sabbagh, Girju, and Diesner (2015) develop a machine-learning model for uncertainty distribution in Arabic. They also propose a unified framework to identify and extract uncertainty cues, holders, and scopes in one fell swoop by casting each task as a supervised token sequence labelling problem. The tool yields an F_1-score of 75.9%, averaged across its three machine-learning models.

3.3 Conclusions and chapter summary

Speculation, also known as hedging, can be defined within the framework of modality. It was first introduced by Lakoff (1972) and is used by speakers to present information as uncertain or unreliable. However, it was not until recently that hedging has been investigated in NLP, mainly as a consequence of the availability of corpora annotated with this information and the organisation of several challenges and shared tasks.

This chapter has presented an overview of the concept of speculation as well as a summary of the most representative works in this field, which includes the task of recognising speculation cues and the resolution of their scope.

3.4 Further reading and relevant resources

A good overview of the concept of speculation is provided in the work of Morante and Sporleder (2012), which also covers the aspects of negation as mentioned in Section 2.5. Related concepts to speculation and the sources from which the reader can obtain more detail are *evidentiality* (Von Fintel, 2006), *hedging* (Lakoff, 1972), (Hyland, 1998), (Medlock & Briscoe, 2007), *factuality* (Saurí & Pustejovsky, 2009), *subjectivity* (Wiebe, Bruce, Bell, Martin, & Wilson, 2001; Wiebe, Wilson, Bruce, Bell, & Martin, 2004) and *certainty* (Shanahan, Qu, & Wiebe, 2006).

Most of the resources listed for negation in Section 2.5 include analysis about speculation detection. In order not to repeat information, the reader should go to that section to find more recommended reading as well as a papers and conferences related to this topic.

Freely available datasets of past competitions that could be used to train and test speculation recognition systems are, for example, the datasets from the BioNLP'09 Shared Task[2] (J. Kim et al., 2009) and the CoNLL-2010 (Farkas et al., 2010).[3]

New conceptualisations of uncertainty as an epistemic status of scientific propositions have recently appeared. The reader is referred to the work of Chen, Song and Heo (2018) in order to learn more about the scalable and adaptive method to identify uncertainty cues under the broadened conceptualisation of uncertainty developed by the authors. To be more precise, they propose a method that starts with a small number of representative words as uncertainty cues and then expands to a much larger set of semantically equivalent words by using the latest NLP technologies, i.e., word2vec models (Mikolov, Sutskever, Chen, Corrado, & Dean, 2013) trained on large-scale documents.

Finally, as for negation detection, new domains have recently been investigated. Cheng et al. (2017) apply their approach also for detecting speculation in the veterinary clinical note domain. The results demonstrate that for detecting this phenomenon, in-domain training data is often necessary to attain reasonable performance levels.

---

2. <http://www.nactem.ac.uk/tsujii/GENIA/SharedTask/>

3. <http://rgai.inf.u-szeged.hu/conll2010st/download.html>

Suggestions for further reading

Cheng, K., Baldwin, T., & Verspoor, K. (2017). Automatic Negation and Speculation Detection in Veterinary Clinical Text. In *Proceedings of the Australasian Language Technology Association Workshop 2017* (pp. 70–78).

Chen, C., Song, M., & Heo, G. E. (2018). A scalable and adaptive method for finding semantically equivalent cue words of uncertainty. *Journal of Informetrics*, 12(1), 158–180.

Ghosh, S., Johansson, R., Riccardi, G., & Tonelli, S. (2011). *Proceedings of the 5th International Joint Conference on Natural Language Processing (IJCNLP 2011)*.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (pp. 3111–3119).

Miller, N., Lacroix, E. M., & Backus, J. E. (2000). MEDLINEplus: building and maintaining the National Library of Medicine's consumer health Web service. *Bulletin of the Medical Library Association*, 88(1), 11.

Morante, R., & Sporleder, C. (2012). Modality and negation: An introduction to the special issue. *Computational Linguistics*, 38(2), 223–260. https://doi.org/10.1162/COLI_a_00095

Shanahan, J. G., Qu, Y., & Wiebe, J. (Eds.). (2006). *Computing attitude and affect in text: Theory and applications* (Vol. 20). Dordrecht: Springer.

The tutorial given by Roser Morante at IJCNLP 2011. <https://www.aclweb.org/mirror/ijcnlp11/downloads/tutorial/tu3_present.pdf>.

Wiebe, J., Bruce, R., Bell, M., Martin, M., & Wilson, T. (2001, September). A corpus study of evaluative and speculative language. In *Proceedings of the Second SIGdial Workshop on Discourse and Dialogue-Volume 16* (pp. 1–10). Association for Computational Linguistics.

Wiebe, J., Wilson, T., Bruce, R., Bell, M., & Martin, M. (2004). Learning subjective language. *Computational linguistics*, 30(3), 277–308.

CHAPTER 4

Applications

This chapter is an in-depth description of the applications for which information about negation and speculation has proven to be useful. It presents several examples of tasks where accurate negation or speculation identification improves the results of the task in question, from information extraction to other less usual ones such as text watermarking detection.

4.1 Information extraction

Information extraction aims at extracting factual information from texts. As Prabhakaran, Rambow, and Diab (2010) point out, there is more to meaning than just propositional context. They also argue that text cannot be seen as a repository of propositions about the world since language provides cues for the discourse participants to model cognitive state (e.g., *beliefs*, *desires*, and *intentions*).

As examples, in the biomedical domain, Light et al. (2004) explore the use of speculative language in MEDLINE abstracts, focusing on expressions of levels of belief (e.g., *hypotheses*, *tentative conclusions*, *hedges*, and *speculations*). They discuss how beneficial it could be to detect this kind of information in this task. For instance, extracting tables of protein–protein interactions would be made more efficient by the knowledge of which interactions are speculative and which are definite. They add that, in the context of knowledge discovery (KD), current speculative statements about a topic of interest can be used as a seed for the automated KD process. For his part, Medlock (2008) affirms that interactive bioinformation systems that take account of hedging can render a significantly more effective service to curators and researchers alike. Meystre, Savova, Kipper-Schuler, and Hurdle (2008) are aware that the surrounding text of the concepts extracted from narrative-text documents plays a critical role. This relevant contextual information includes negation (e.g., *denies any chest pain*), temporality (e.g., *fracture of the tibia 2 years ago*), and the event-subject identification (e.g., *his mother has diabetes*). What is more, the most popular information extraction systems such as *cTakes* (Savova et al., 2010) incorporate negation detection.

There are numerous examples of tasks carried out during information extraction that can be made more efficient by the detection of negation and speculation.

The extraction of drug–drug interactions from literature is another concrete case where detecting this linguistic phenomenon improves the main task. Bokharaeian, Díaz Esteban and Ballesteros Martínez (2013) achieve better results in drug–drug interactions in the DDI 2011 corpus when considering annotations for negations. In addition, the highest-scoring team in the Semeval 2013-DDI Extraction 2013 challenge (Segura Bedmar, Martínez, & Herrero Zazo, 2013) also use negation information in their system (M. F. M. Chowdhury & Lavelli, 2013). A recent study related to this task explores clause-dependency-related features alongside linguistic-based negation scope and cues to overcome the complexity of the sentences, showing that performance improves when employing these proposed features combined with a bag-of-words kernel (Bokharaeian, Diaz, & Chitsaz, 2016).

4.2 Sentiment analysis and opinion mining

Sentiment analysis and opinion mining are ongoing fields of research consisting of the computational treatment of opinion, sentiment, and subjectivity in text (Pang & Lee, 2008). Common applications of sentiment analysis include the automatic determination of whether a review posted online (of a movie, a book, or a consumer product) is positive or negative toward the item being reviewed (Taboada, 2016).

Many authors have studied the role of negation in sentiment analysis since they are aware of the fact that handling negation in this task is a very important step (Kolchyna, Souza, Treleaven, & Aste, 2015) and that it is still a research challenge that needs to be addressed (Ragini & Anand, 2016). In fact, as Liu (2012) affirms, negation words are the most important class of sentiment shifters. Wiegand, Balahur, Roth, Klakow and Montoyo (2010) present a study of the role of negation in sentiment analysis, proving that this common linguistic construction is highly appropriate for this task. The authors explain that an effective negation model for sentiment analysis usually requires the knowledge of polarity expressions. Negation is not only conveyed by common negation words but also other lexical units, such as *diminishers*. Negation expressions are ambiguous, i.e., in some contexts they do not function as a negation and, therefore, need to be disambiguated. A negation does not negate every word in a sentence; therefore, using syntactic knowledge to model the scope of negation expressions is useful.

As described in Section 2.3.2, Councill et al. (2010) explain that, as they expected, the performance of their sentiment-analysis system is improved dramatically by introducing negation-scope detection. In a more recent work, Dadvar et al. (2011) investigate the problem of determining the polarity of sentiments in film reviews when negation words, such as *not* and *hardly*, occur in sentences. The authors observe significant improvements in the classification of the documents after applying negation detection. Hogenboom, van Iterson, Heerschop, Frasincar and Kaymak (2011) show that properly accounting for negation when analysing sentiment in natural-language texts may help improve the classification of unseen natural-language text as carrying either a positive or a negative sentiment. Lapponi et al. (2012) present a system for negation resolution and use it as a component in a simple negation-aware testbed for sentiment classification. Results show that all negation-aware configurations are beneficial in terms of the combined F_1-score. Asmi and Ishaya (2012) propose a framework for automatic identification of opinions in textual data, including rules for negation recognition and calculation especially designed to improve sentiment text analysis. For ChandraKala and Sindhu (2012), negation detection is one of the most important pre-processing steps in identifying opinions efficiently. Reitan et al. (2015) confirm that taking negation into account improves sentiment-classification performance significantly on Twitter. Recently, Ohana, Tierney, and Delany (2016) investigate whether the treatment of negative sentiments in negated text can improve the performance of sentiment classification tasks. They propose a novel adjustment factor based on negation occurrences as a proxy for negative sentiment polarity. This shows statistically significant performance improvements on all domains tested. Pröllochs, Feuerriegel, and Neumann (2016) examine how detecting negation scopes can improve the accuracy of sentiment analysis for financial news, leading to an improvement of up to 10.63% in the correlation between news sentiment and stock market returns. This reveals negation-scope detection as a crucial leverage in making decisions based on sentiment analysis. Hussein, Doaa Mohey El-Din Mohamed (2016) find that negation is the most important challenge with the greatest impact on any sentiment analysis. They come to this conclusion through a comparison between the 41 papers in sentiment-analysis challenges. Diamantini, Mircoli, and Potena (2016) experiment with different datasets, proving that their proposed negation-handling algorithm based on dependency-based parse trees achieves better sentiment-analysis accuracy. Sharif, Samsudin, Deris, and Naseem (2016) detect the effect of negation on consumer reviews which appear positive but are in fact completely negative in meaning. Their proposed negation approach presents a way of calculating negation identification that helps to improve review classification accuracy. Farooq, Mansoor, Nongaillard, Ouzrout, and Qadir (2017) show that their proposed negation-handling method improves the accuracy of both negation-scope identification and overall sentiment analysis.

In sentiment analysis, as in other areas, the impact of speculation has not been studied as much as it has been in negation. However, it is clear that identifying speculative information is also crucial in sentiment analysis, where, as Saurí and Pustejovsky (2009) explain, the same situation can be presented as a fact in the real world, a mere possibility or as a counterfactual according to different sources.

In fact, Pang and Lee (2004) show how subjectivity detection in the review domain helps to improve polarity classification. Wilson, Wiebe, and Hoffmann (2005) also suggest that the identification of speculation in reviews can be used for opinion mining since it provides a measure of the reliability of the opinion contained.

Carrillo-de-Albornoz and Plaza (2013) also show that negation, intensifiers, and modality are common linguistic constructions that can modify the emotional meaning of the text and therefore, need to be taken into consideration in sentiment analysis. (Mohammad, 2016) also affirms that certain terms such as *negations* and *modals* impact the sentiment of a sentence, without the words themselves having strong sentiment associations. Cruz et al. (2015) measure the practical impact of accurate negation and speculation detection in sentiment analysis. They prove that the accurate detection of cues and scopes is of paramount importance to the sentiment detection task since performance is improved by identifying this kind of information. At the same time, the experiments indicate that simplistic approaches to negation and speculation are insufficient for sentiment classification. Kiritchenko and Mohammad (2016) observe that negations, modals, and degree adverbs can significantly affect the sentiment of the words they modify. They explain that the most change in sentiment is caused by negation where this negation consistently lowers the scores of positive words, and increases the scores of negative words. Modals also tend to lower the scores of positive words, and increase the scores of negative words, though to a much smaller extent than negations.

4.3 Recognising textual entailment

Since 2005, recognising textual entailment has generated much interest in the natural language research community as a result of the PASCAL recognising textual entailment challenge (Dagan, Glickman, & Magnini, 2006). Textual entailment is defined as a directional relationship between pairs of text expressions, denoted by *t* (i.e., the entailing "text") and *h* (i.e., the entailed "hypothesis"). It can be said that *t* entails *h* if the meaning of *h* can be inferred from the meaning of *t*, as would typically be interpreted by the average person (Gaona, Gelbukh, & Bandyopadhyay, 2010) .

Factuality-related information has been taken as a basic feature in some systems using the data from the PASCAL RTE challenges (De Marneffe et al., 2006; Snow, Vanderwende, & Menezes, 2006). For example, de Marneffe et al. (2006) show how negation influences some patterns of entailment. They focus on contexts that reverse monotonicity, such as *negations* and *quantifiers*. Snow et al. (2006) describe a heuristic based on negation and modality mismatch, which allows them to predict false entailment. Androutsopoulos and Malakasiotis (2009) discuss the need to be careful with negations and other expressions that do not preserve truth values. Also in this context, Sammons, Vydiswaran, and Roth (2010) believe that recognising key negation phenomena correctly and consistently could significantly improves the overall accuracy of the system. Rios, Specia, Gelbukh, and Mitkov (2014) propose a statistical relational learning approach to recognising textual entailment. They represent an entailment decision problem with a range of different features that try to capture entailment and non-entailment by focusing on negations and quantifiers. Lai and Hockenmaier (2014) present a system that employs features that depend on the degree of word overlap and alignment between the two sentences, the presence of negation, and the semantic similarities of the words and substrings that are not shared across the two sentences. Jimenez et al. (2014) are other authors that are aware that negations play an important role in this task. Zhao, Zhu, and Lan (2014) use a predefined list of negation words designed to deal with contradiction-entailment relationships. Lien and Kouylekov (2015) develop a special rule-based contradiction module focused on negation since it is the most frequent contradiction indicator. Other examples of textual-entailment detectors that include negation detection are those built by Sharma et al. (2015) and Beltagy, Roller, Cheng, Erk, and Mooney (2017).

4.4 Machine translation

Machine translation is one of the oldest subfields of artificial intelligence research and it consists in automatically converting one natural language into another, preserving the meaning of the input text, and producing fluent text in the output language.

In recent years, there has been increasing interest in improving the quality of machine translation systems over a wide range of linguistic phenomena such as *negation* and *modality*. However, initial considerations about their inclusion in this task began to develop in the 1980s. An example is the research developed by van Munster (1988).

More recent examples are the works of Collins, Koehn, and Kučerová (2005) and Baker et al. (2010). The former includes a treatment of negation in translation from German to English, since it is a phenomenon that leads to differing word order between these two languages (e.g., with *not* in English and *nicht* in German). The latter introduces modality identification in a machine translation application. They show how using a structure-based tagger to annotate English modalities on an English-Urdu training corpus improves the translation quality score for Urdu. They conclude that speculation is very important for a correct representation of events and likewise for translation. Wetzel and Bond (2012) report that negation might pose a problem to statistical machine translation systems. They alleviate these difficulties in translation from Japanese to English by automatically expanding the training data with negated sentence pairs. The additional data is obtained by rephrasing existing data based on the semantic structure of the input. The results show improvements in translation quality. Baker et al. (2012) develop a modality/negation lexicon and a set of automatic modality/negation taggers whose produced labels are appended to the nodes in the syntactic tree input, in order to build the translation models. As a result, Urdu-English translation improves by 0.5 BLEU points over a syntax-only baseline. Fancellu and Webber (2014) present an approach to translating negative sentences from Chinese to English that is based on the application of the semantics of negation to n-best list re-ranking. More precisely, they identify the core semantic elements of negation (i.e., cue, event and scope) in a source-side dependency parse. Then, they re-ranking hypotheses on the n-best list produced after decoding according to the extent to which a hypothesis realises these elements. This method shows a considerable improvement over the baseline. Finally, Fancellu and Webber (2015) provide an analysis of the errors involved in Chinese-to-English translation which could guide future work on improving the translation of negative sentences.

4.5 Information retrieval

The task of information retrieval is to select, from a collection of textual documents, a subset that is relevant to a particular query, based on keyword search and possibly augmented by the use of a thesaurus. The ranked list of documents returned does not provide any detailed information on the content of those documents. This differs from information extraction whose goal is not to rank or select documents, but to extract from the documents important facts about pre-specified types of events, entities, or relationships, with the aim of building more meaningful and more vivid representations of their semantic content (Piskorski & Yangarber, 2013).

Koopman, Bruza, Sitbon and Lawley (2010) study the effects of negation on information retrieval, concluding that, overall, negation does not have a major impact on retrieval and that specific methods of dealing with negation would only be required in specific domains such as medical data, where negation is prevalent and can pose problems in the quality of results retrieved. An example of a context-sensitive medical information retrieval system that includes negation detection is the one developed by Averbuch et al. (2004). The authors explain that the context of negation, a negative finding, is of special relevance because many of the most frequently described findings are those denied by the patient or subsequently ruled out. Hence, if negation is not taken into account in this task, many of the retrieved documents will be irrelevant. Denny, Miller, Waitman, Arrieta, and Peterson (2009) identify QT[1] interval prolongation from electrocardiogram (ECG) impressions using a general-purpose natural language processor. In this work, the authors apply a modified version of the NegEx algorithm W. W. Chapman et al. (2001b) to identify the negation. They assert that natural language processing (NLP) with negation detection can extract concepts from ECG impressions with high accuracy. Denny et al. (2012) investigate how NLP improves recognition of colorectal cancer (CRC) testing in an electronic medical record. As part of its NLP, they identify Unified Medical Language System (UMLS) concepts found in each sentence along with information on its relevant context and information about whether or not the concept is negated. Moreover, an algorithm identifies negated phrases as well as common verbs and other modifiers that change the status of CRC-related testing (e.g., *refused*, *declined*). The results show that applying NLP to an Electronic Health Records (EHR) detects more CRC tests than either a manual chart review or a billing-records review (i.e., queries based on the billing code) alone. For Koopman and Zuccon (2014), assigning a negative weight to negated content is more effective than the common practise of removing or ignoring this content (Voorhees & Hersh, 2012). However, on an individual query level, negated content can be useful and therefore negated content within a document should not be ignored. Kuhn and Eickhoff (2016) describe a way to improve the quality of bio-medical information retrieval by drawing implicit negative feedback from negated information in noisy natural language search queries.

4.6 Other tasks

Negation recognition can also improve other tasks. For instance, Fiszman, Rindflesch and Kilicoglu (2006) report that one of the main causes of failure highlighted by their summarisation system is that of missed negation. Therefore, negative information should be taken into account. Su, Huang, and Chen (2010) explore how linguistically encoded information of the nature of evidence for a statement can contribute to the prediction of trustworthiness (distinguishing truth from lies) in NLP. Their experimental results report improvements of up to 14.85% over the baseline. This confirms that the nature of evidence is an important clue for trustworthiness detection. In the classification of citations task, authors such as Di Marco, Kroon and Mercer (2006) show that identifying the nature of the exact

1. A measure of the time between the start of the Q wave and the end of the T wave in the heart's electrical cycle.

relationship between a citing and cited paper requires an understanding of the rhetorical relations within the argumentative context in which a citation is placed. To determine these relations automatically, the use of hedging to modify the effect of a scientific claim will be significant. They also explain that hedging is a relevant aspect of the rhetorical structure of citation contexts and that the pragmatics of hedges may help in determining the rhetorical purpose of citations. Speculation detection is also beneficial in the field text-structure identification. For example, Grabar and Hamon (2009) study how the use of speculation markers in scientific writing can be useful in ascertaining whether these markers are regularly spread across biomedical articles and then in establishing the logical structure of articles. More precisely, they compute associations between article sections and speculation markers, before coming to the conclusion that speculation is governed by observable usage rules in scientific articles and may help their structuring. Szarvas and Gurevych (2013) investigate the application of uncertainty detection to text watermarking. The aim of this problem is to produce individually identifiable copies of a source text via small manipulations to the text (e.g. *synonym substitutions*). They demonstrate that uncertainty cues are promising for this task since they can be accurately disambiguated and their substitution with other cues has only a marginal impact to the meaning of the text.

4.7 Conclusions and chapter summary

This chapter shows that the treatment of modality and negation is highly relevant for all NLP applications that involve deep text understanding. This includes applications that need to discriminate between factual and non-factual information. Hence, the adequate modelling of these phenomena is of crucial importance since many applications can benefit from identifying this kind of information. In addition, this chapter presents several examples of tasks, from *information extraction* to other less usual ones such as *text watermarking detection*, where accurate negation or speculation identification improves the results of the task in question.

4.8 Further reading and relevant resources

A useful work to better understand the task of recognising textual entailment is the tutorial given by Ido Dagan, DanRoth, and Fabio Massimo Zanzotto at ACL 2007. It includes negation and modality as an aspect of the logical structure. To know more about sentiment analysis and the implications that negation has in this task, the reader is referred to the book written by Liu (2015). Challenges related to sentiment analysis in which detecting negation is also crucial and about which the reader might want to know more are: 1) recognising contradiction and contrast (Harabagiu et al., 2006; J. Kim, Zhang, Park, & Ng, 2005), 2) detecting irony (Reyes & Rosso, 2014) and 3) identifying sarcasm (Joshi, Bhattacharyya, & Carman, 2017).

Suggestions for further reading

Harabagiu, S., Hickl, A., & Lacatusu, F. (2006). Negation, contrast and contradiction in text processing. *Paper presented at the AAAI*, 6, 755–762.

Joshi, A., Bhattacharyya, P., & Carman, M. J. (2017). Automatic sarcasm detection: A survey. *ACM Computing Surveys (CSUR)*, 50(5), 73. https://doi.org/10.1145/3124420

Kim, J. J., Zhang, Z., Park, J. C., & Ng, S. K. (2005). BioContrasts: extracting and exploiting protein–protein contrastive relations from biomedical literature. *Bioinformatics*, 22(5), 597–605.

Liu, B. (2015). *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge University Press. https://doi.org/10.1017/CBO9781139084789

Mohammad, S. M. (2016). Sentiment analysis: Detecting valence, emotions, and other affectual states from text. In *Emotion measurement* (pp. 201–237).

Reyes, A., & Rosso, P. (2014). On the difficulty of automatically detecting irony: Beyond a simple case of negation. *Knowledge and Information Systems*, 40(3), 595–614. https://doi.org/10.1007/s10115-013-0652-8

The tutorial given by Ido Dagan, DanRoth and Fabio Massimo Zanzotto at ACL 2007. <www.cs.biu.ac.il/~dagan/TE-Tutorial-ACL07.ppt>

CHAPTER 5

Resources

Chapter 5 presents a set of relevant resources for any researcher or developer interested in the problem: the main publicly available corpora annotated with information about negation and speculation for different domains and languages, the set of available negation/speculation recognition tools developed for English that can be integrated in other NLP applications and a description of the evaluation metrics used by the NLP community to evaluate the detectors built to tackle this task.

5.1 Annotated corpora

Two of the first and most notable attempts to annotate negation and speculation in English were the Genia Event corpus (J. Kim, Ohta, & Tsujii, 2008) and the BioInfer corpus (Pyysalo et al., 2007). In the Genia Event corpus,[1] which is a subset of the original GENIA corpus (J. D. Kim et al., 2003), biological events in 1,000 MEDLINE abstracts are annotated with negation and three levels of uncertainty: *certain*, *probable* and *doubtful*. The attribute *certain* is chosen when the existence of the event cannot be questioned. Events are marked as *probable* if their existence cannot be stated for certain and *doubtful* is used when the existence of the event is unlikely or the event forms part of a hypothesis. In addition, the GENIA corpus has been continually enriched with various levels of syntactic, semantic and discourse-level annotation (Thompson, Ananiadou, & Tsujii, 2017).

In the BioInfer (Bio Information Extraction Resource) corpus,[2] biological relations are annotated for negation. The corpus is 1,100 sentences in size. In these two corpora, biological terms (relations and events) are annotated for both negation and hedging, but linguistic cues are not annotated.

1. The Genia Event corpus can be downloaded from <http://www.nactem.ac.uk/meta-knowledge/>.

2. The resource is publicly available at <http://mars.cs.utu.fi/BioInfer/>.

In the biomedical domain, the existence of the BioScope corpus[3] (Szarvas et al., 2008), in which both negative/speculative keywords and their scope are annotated, has facilitated the development of corpus-based statistical systems for negation/hedge detection and resolution. It consists of more than 20,000 sentences which are split into three collections: clinical documents used for the CMC clinical coding challenge (Farkas & Szarvas, 2008), 9 scientific papers (five scientific papers from FlyBase and four scientific papers from the open access BMC Bioinformatics repository) and scientific abstracts from the GENIA corpus.

In a text, only sentences with one or more instances of speculative or negative language are considered for annotation. The annotation is based on linguistic principles, i.e., parts of sentences which do not contain any biomedical terms are also annotated if they assert the non-existence/uncertainty of something. Each negated/speculated sentence is annotated with information about the keyword and the scope. The annotation of Bioscope followed a min-max strategy where the minimal unit that expresses negation/speculation is considered the cue (min strategy) and the scope is extended to the largest syntactic unit possible (max strategy). The cue is always included in the scope.

Other corpora freely available that are annotated with negation and speculation in this domain are the THYME corpus (Styler IV et al., 2014) and the NegDDI-DrugBank 2013 corpus (Bokharaeian, Diaz, Neves, and Francisco (2013). The THYME corpus[4] consists of 1,254 de-identified clinical reports from Mayo Clinic. The reports summarise the interactions between physicians and patients in two distinct fields within oncology: brain cancer and colon cancer. Each note is annotated with temporal events, temporal relations and clinical concepts. The event properties include contextual modality attributes that include the values *actual*, *hypothetical*, *hedged* and *generic*, where *actual* covers events which have actually happened, *hypothetical* is referred to conditionals and possibilities, *hedged* is for situations where doctors proffer a diagnosis, but do so cautiously in to avoid legal liability for an incorrect diagnosis or for overlooking the actual one and, in contrast, *generic* events do not refer to a particular patient's illness or treatment, but instead discuss illness or treatment in general.

The NegDDI-DrugBank 2013 corpus[5] is specially designed for the drug–drug interaction detection task since it has been proven that detecting negation

3. The corpus is publicly available for educational and research purposes at <http://rgai.inf.u-szeged.hu/index.php?lang=en&page=bioscope>.

4. The annotated data are available at <http://thyme.healthnlp.org>..

5. The corpus is available for public use at <http://nil.fdi.ucm.es/sites/default/files/NegDDI_DrugBank.zip>.

improves results (Bokharaeian et al., 2013; M. F. M. Chowdhury & Lavelli, 2013). It is an extended version of the DDI-DrugBank 2013 corpus (Herrero-Zazo, Segura-Bedmar, Martínez, & Declerck, 2013) annotated with information about negation. It consists of 6,648 sentences extracted from the DDI-DrugBank database, where 1,448 of the sentences contain at least one negation scope. The annotation is based on Bioscope's guidelines and consists of adding two new tags: the cue and the scope of the negation.

In other domains, the FactBank corpus[6] (Saurí & Pustejovsky, 2009) contains 208 documents from newswire and broadcast news reports in which event mentions are annotated for polarity (when a sentence is affirmative or negative) and certainty. Three levels of certainty are distinguished in the database: *certain*, *probable* and *possible*. The tag *underspecified* is used for events for which there is not enough evidence to attribute any of the former labels. The original corpus does not contain annotation for cues; it is only predicates denoting events that are marked.

Some publicly available corpora have appeared as a result of some of the shared tasks organised over the years. An example is the sets of documents used in the CoNLL-2010[7] (Farkas et al., 2010):

– A biological dataset. The training dataset consists of the biological part of the BioScope corpus annotated manually for hedge cues and their scopes. The evaluation dataset is based on 15 biomedical articles downloaded from the PubMedCentral database, including five random articles taken from the BMC Bioinformatics journal in October 2009, five random articles to which the drosophila MeSH term was assigned and five random articles with the MeSH terms human, blood cells and transcription factor. These texts are manually annotated for hedge cues and their scope. To annotate the training and the evaluation datasets, the same annotation principles are applied.
– A Wikipedia dataset. A total of 2,186 paragraphs collected from Wikipedia archives are used as training data and 2,346 for evaluation. The Wikipedia paragraphs are selected using the weasel tags added by the editors of the encyclopaedia. Each sentence is annotated manually for weasel cues, based on the principle that sentences should be treated as uncertain if they contain at least one weasel cue, i.e., the scope of weasel words is the entire sentence.

6. The FactBank corpus is freely available at <https://catalog.ldc.upenn.edu/ldc2009t23>.

7. The datasets are freely available for further benchmark experiments at <http://rgai.inf.u-szeged.hu/conll2010st/download.html>.

Another publicly available corpus that has appeared as a result of a shared task is that used in the SEM 2012 Shared Task (Morante & Blanco, 2012), in which two different datasets[8] are provided:

- A subset of the ConanDoyle-neg corpus (Morante & Daelemans, 2012). This dataset includes two stories for training and development and another for testing.
  The ConanDoyle-neg corpus was released in conjunction with this shared task. It is a corpus of Arthur Conan Doyle's stories manually annotated with negation keywords and their scope. The negation cue, its scope and the negated event are annotated in each sentence which contains negation statements. The annotation is inspired by the guidelines of Bioscope, but with several differences. Among the most important differences are the following: the negated event is annotated; negation cues are not included in the scope; scopes can be discontinuous; affixal cues are annotated; and if the scope of a negation cue is not explicit, the negation cue is marked as such, but the scope is not annotated.
- A set of 3,993 sentences of the *Wall Street Journal* section of the Penn TreeBank marked with MNEG role in PropBank annotated with the focus of negation (Blanco & Moldovan, 2011a).

The focus of negation is also annotated in the DeepTutor Negation corpus[9] (DT-Neg corpus) (Banjade & Rus, 2016) which contains texts extracted from tutorial dialogues where students interacted with an intelligent tutoring system to solve conceptual physics problems. In total, the corpus is composed of 1,088 instances which contain annotated negations in student responses with scope and focus marked according to the context of the dialogue.

In the sentiment analysis domain, the Simon Fraser University (SFU) Review corpus (Taboada, 2008) is annotated with negation and speculation information. This corpus is extensively used in opinion mining (Martınez-Cámara, Martın-Valdivia, Molina-González, & Urena-López, 2013; Rushdi Saleh, Martín-Valdivia, Montejo-Ráez, & Ureña-López, 2011; Taboada, Brooke, Tofiloski, Voll, & Stede, 2011) and consists of 400 documents of film, book, and consumer product reviews (50 of each type) from the website Epinions.com. The corpus has several annotated versions (e.g., *for appraisal and rhetorical relations*), including one where all 400 documents are annotated at the token level with negative and speculative cues

8. Both datasets can be obtained from <https://www.clips.uantwerpen.be/sem2012-st-neg/data.html>.

9. The DT-Neg corpus is available for research purposes at <http://deeptutor.memphis.edu/resources.htm>.

and at sentence level with their linguistic scope[10] (Konstantinova et al., 2012). The entire corpus has been annotated by one linguist, adapting the existing Bioscope corpus guidelines in order to fit the needs of the review domain. A second linguist has annotated 10% of the documents, which were randomly selected, in a stratified way, with the aim of measuring inter-annotator agreement.

The SFU Opinion and Comments Corpus (SOCC)[11] (Kolhatkar et al., 2018) is another corpus annotated for negation in the sentiment analysis domain. In total, 1,043 comments written in response to a subset of the articles published in the Canadian newspaper "The Globe and Mail" in the five-year period between 2012 and 2016, annotated with three layers of annotations: constructiveness, negation, and appraisal. With respect to the negation annotation, the authors develop extensive and detailed guidelines for the annotation of negative keywords, scope and focus.

Negation and speculation detection is becoming an important task in other languages than English, so the natural language processing (NLP) community is developing annotated resources for many domains in other languages. However, few of these resources annotated with negation/speculation are freely available. Examples of downloadable corpora are the UAM Spanish Treebank[12] (Moreno, López, Sánchez, & Grishman, 2003) and the SFU ReviewSP-NEG[13] (Jiménez-Zafra, Taulé, Martín-Valdivia, Ureña-López, & Martí, 2017) for Spanish; the hUnCertainty Corpus[14] (Vincze, Simkó, & Varga, 2014) for Hungarian; the Cemr corpus[15] (B. He et al., 2017) and the CNeSp corpus[16] (Zou, Zhou, & Zhu, 2016) for Chinese; and the EMC Dutch clinical corpus[17] (Afzal et al., 2014) for Dutch.

Finally, one should note that, despite the publicly available corpora annotated for negation and speculation, the problem is that the lack of unified annotation principles leads to the impossibility of direct comparison of the corpora. This means that each of the negation and speculation detectors is optimised for the

10. This version of the corpus is freely available at <https://www.sfu.ca/~mtaboada/research/SFU_Review_Corpus.html>.

11. A complete version of the corpus is accesible via <https://github.com/sfu-discourse-lab/SOCC>.

12. The corpus is available at <http://www.lllf.uam.es/~sandoval/UAMTreebank.html>.

13. It is available at <http://sinai.ujaen.es/sfu-review-sp-neg-2/>.

14. It is downloadable at <http://rgai.inf.u-szeged.hu/index.php?lang=en&page=uncertainty>.

15. The corpus is available at <https://github.com/WILAB-HIT/Resources>.

16. It is available at <http://nlp.suda.edu.cn/corpus/CNeSp/>.

17. It is available at <http://biosemantics.org/index.php/resources/emc-dutch-clinical-corpus>.

corpus or domain on which it was trained, i.e., existing detectors can hardly be used across domains, and creating new resources and tools for each domain is time consuming and costly. Instead, a unified, comprehensive approach would be optimal, which could be adapted to the specific needs of each domain without additional effort. Language independence of the model would also be desirable (Vincze, 2015).

5.2 Tools

The majority of the negation/speculation detection systems reviewed in this book are not publicly available. Among the systems that recognise this type of linguistic phenomenon in English are the following:

- NegEx (W. W. Chapman et al., 2001b). A stand-alone algorithm that can be integrated with any application that indexes clinical conditions from text. In fact, NegEx is a frequently applied negation algorithm in biomedical informatics systems due to its simplicity, availability, and generalisability to various NLP applications. NegEx locates trigger terms indicating that a clinical condition is negated or possible. It also determines which text falls within the scope of the trigger terms. Therefore, the input is a sentence with optionally indicated clinical conditions from the sentence. The output could be the value of indexed conditions if the user indicates the conditions whose negation status he/she is unsure of (i.e., NegEx returns negated or possible for those conditions within the scope of negation terms) or the text within the scope of a trigger term (this is a more generalised output since it is not necessary to predetermine conditions of interest).
- ConText (Harkema et al., 2009). An extension of the NegEx algorithm for determining the values of two additional contextual features other than negation, namely temporality and experiencer. ConText's input is a sentence with indexed clinical conditions. The output for each indexed condition is the value for contextual features or modifiers. As mentioned before, the initial version of ConText determines values for three modifiers: *negation* (affirmed or negated), *temporality* (recent, historical, or hypothetical) and *experiencer* (patient or other). A newer version (pyConText) is more extendable and can include user-defined modifiers such as *uncertainty.*[18]

18. The java and python versions of NegEx and ConText are publicly available at <https://code.google.com/archive/p/negex/>. More information about the algorithms can be also obtained at <http://toolfinder.chpc.utah.edu/content/contextnegex>.

- DEEPEN[19] (Mehrabi et al., 2015). An algorithm developed with the aim of reducing the number of NegEx's false positives (FP) by taking into account the dependency relationship between negation words and concepts within a sentence using the Stanford Dependency Parser (SDP). The input is a sentence with an indicated clinical condition. The output is a string that identifies the negation status of the concept in the sentence. The possible negation statuses are: *affirmed*, which means NegEx considers the concept negated; *affirmed confirmed by SDP*, used when NegEx considers the concept negated but DEEPEN considers it affirmed; and *negation confirmed by SDP* if NegEx considers the concept negated and DEEPEN confirms that.
- More recently, (Enger, Velldal, & Øvrelid, 2017) provide developers and researchers with an open-source toolkit[20] for negation detection which identifies negation cues and their corresponding scope in either raw or parsed text using maximum-margin classification. The tool is trained using the ConanDoyle-neg corpus (Morante & Daelemans, 2012). However, users are able to train their own models, too. The input is a raw running text or parsed data in the CoNLL-X format (Buchholz & Marsi, 2006). The output is a file where the first eight columns are identical to the input file, and the columns thereafter include cues and scopes encoded in the ConanDoyle format.

Systems that identify negation and/or speculation in other languages and that are freely available to the NLP community have not been identified.

5.3 Evaluation

Older studies have measured the performance of the negation and speculation detection systems in terms of precision/recall break-even point (BEP), an evaluation measure originally introduced in the field of information retrieval to evaluate retrieval systems. It is calculated as the average of precision and recall when the difference between the two is minimal.

However, nowadays, negation and speculation detectors are evaluated by the NLP community, following the model of the evaluation scheme established by the

19. DEEPEN is written in Java and is freely available for researchers to use at <http://svn.code.sf.net/p/ohnlp/code/trunk/DEEPEN/>.

20. The source code is available at <https://github.com/marenger/negtool>.

CoNLL-2010 Shared Task (Farkas et al., 2010) on speculation detection,[21] also applying this when evaluating results for the negation task.

5.3.1 Evaluation measures for cue identification

For the approaches for cue detection: precision, recall, and their harmonic mean F_1-score (Rijsbergen, 1979) are used to evaluate the systems at three different levels (sentence-level, token-level, and cue-level).

The sentence-level scores correspond to Task 1 in the CoNLL-2010 Shared Task, that is, correctly identifying whether a sentence contains uncertainty or not.

The scores at the token level measure the number of individual tokens within the span of a cue annotation that the classifier has correctly labelled as a cue where:

$$\text{Precision (P)} = \frac{\text{\# tokens correctly negated / speculated by the system}}{\text{\# tokens negated / speculated by the system}}$$

$$\text{Recall (R)} = \frac{\text{\# tokens correctly negated / speculated by the system}}{\text{\# tokens negated / speculated in the test collection}}$$

$$F_1 = \frac{2PR}{P + R}$$

Finally, the stricter cue-level scores measure how well a classifier succeeds in identifying entire cues (which will in turn provide the input for the downstream components that later try to resolve the scope of the speculation or negation within the sentence). A true positive (TP) at the cue level requires that the predicted cue exactly match the annotation in its entirety (full multiword cues included).

5.3.2 Evaluation measures for scope resolution

The method of evaluating scope used in CoNLL-2010 is rather strict (Morante & Blanco, 2012): a TP requires an exact match for both the entire cue and the entire scope. On the other hand, a FP can be incurred by any one of three different events: (1) incorrect cue labelling with correct scope boundaries; (2) correct cue labelling with incorrect scope boundaries; or (3) incorrectly labelled cue and scope. Moreover, conditions (1) and (2) will give a double penalty, in the sense that they also count as false negatives (FN) given that the gold-standard cue or scope is

21. The scorer tools are publicly available at <http://rgai.inf.u-szeged.hu/conll2010st/download.html>.

missed. Finally, FNs are of course also incurred by cases where the gold-standard annotations specify a scope but the system makes no such prediction.

Therefore, it is common to use another method of evaluating the scope resolution. That is, to obtain the system performance, two different tests have been carried out: token-level evaluation and cue-level evaluation.

In the token-level evaluation, a token is correctly classified if it has been properly classified as being inside or outside the scope of all negation or speculation cues that appear in the sentence. This means that if there is more than one negation or speculation cue in the sentence, the token is correctly assigned a class for each of these cues. The evaluation takes the token as a unit. The same measures as in the cue detection task have been employed. In this case:

$$\text{Precision (P)} = \frac{\text{\# tokens belonging to some scope correctly identified by the system}}{\text{\# tokens belonging to some scope identified by the system}}$$

$$\text{Recall (R)} = \frac{\text{\# tokens belonging to some scope correctly identified by the system}}{\text{\# tokens belonging to some scope identified in the test collection}}$$

F_1-score is calculated using the same expression as in the cue detection task.

On the other hand, also in the scope recognition task, the percentage of scopes correctly classified is evaluated. This is a cue-level evaluation and therefore takes the cue as a unit. In this case, the scope associated with a cue is correct when all the tokens of a sentence have been correctly classified as inside or outside the scope of the cue.

5.4 Conclusions and chapter summary

This chapter is dedicated to relevant resources for any researcher or developer interested in the negation/speculation detection problem: corpora, tools and evaluation metrics. It includes the main publicly available corpora annotated with information about negation and speculation for different domains and languages. It also lists the set of available negation/speculation recognition tools developed for English that can be integrated in other NLP applications. Finally, this chapter describes the evaluation metrics used by the NLP community to evaluate the detectors built to tackle this task.

5.5 Further reading and relevant resources

A reference textbook that introduces the classic metrics of precision, recall and F-score is Baeza-Yates and Ribeiro-Neto (1999). The work of Sokolova, Japkowicz and Szpakowicz (2006) is another relevant reference that describes the evaluation measures in use today, focused on a classifier's ability to identify classes correctly. The authors also suggest new measures which take into account properties such as the avoidance or class discrimination.

As explained in this chapter, negation and speculation annotation is a complex task on which there is no consensus. In the recommended bibliography, there are some works that could help the reader to better understand its difficulties, pitfalls and challenges (Blanco & Moldovan, 2011b; Light et al., 2004; Morante, 2010; Pustejovsky & Stubbs, 2012; Vincze, 2010; Wilbur, Rzhetsky, & Shatkay, 2006; Wu et al., 2014).

Suggestions for further reading

Baeza-Yates, R., & Ribeiro-Neto, B. (1999). Modern information retrieval. New York, NY: ACM Press.

Kolhatkar, V., Wu, H., Cavasso, L., Francis, E., Shukla, K., & Taboada, M. (2018). The SFU Opinion and Comments Corpus: A Corpus for the Analysis of Online News Comments.

Morante, R. (2010). Descriptive analysis of negation cues in biomedical texts.

Pustejovsky, J., & Stubbs, A. (2012). *Natural Language Annotation for Machine Learning: A guide to corpus-building for applications*. " O'Reilly Media, Inc.".

Sokolova, M., Japkowicz, N., & Szpakowicz, S. (2006). Beyond accuracy, F-score and ROC: A family of discriminant measures for performance evaluation. *Paper presented at the Australasian Joint Conference on Artificial Intelligence*, 1015–1021.

Vincze, V. (2010, July). Speculation and negation annotation in natural language texts: what the case of BioScope might (not) reveal. In *Proceedings of the workshop on negation and speculation in natural language processing* (pp. 28–31). Association for Computational Linguistics.

Wilbur, W. J., Rzhetsky, A., & Shatkay, H. (2006). New directions in biomedical text annotation: definitions, guidelines and corpus construction. *BMC bioinformatics*, 7(1), 356.

CHAPTER 6

Future trends and discussion

6.1 Future trends

Future trends indicate that recent research in negation and speculation detection has included the use of more than one corpus as a source of training and testing. For example, models have been trained to recognise markers in one corpus and evaluate it on another.

In addition, efforts are also focused on detecting negation and speculation in other languages than English. Until now, most efforts have involved the creation or adaptation of regular-expression-based algorithms. But now research is focused on annotating corpora in other languages with negation and speculation information. The increase of these corpora will lead to the creation of machine learning systems.

From our analysis of the state of the art in negation and speculation detection in Chapters 2 and 3, we find that most of the projects have been conducted on the clinical and opinion domains. Therefore, forthcoming studies are likely to be conducted on different areas since, as described in Chapter 4, negation and speculation information can be extremely useful in many natural language processing (NLP) tasks.

NLP is gaining importance in the medical domain where negation and speculation annotation is now related to the temporality of the event (e.g., THYME corpus (Styler IV et al., 2014)). The reason behind this is that identifying which events are negated/speculated is of the utmost importance when determining the symptoms or effects of an illness or treatment.

Finally, as mentioned in Section 5.1, it is necessary to make progress in creating unified annotation principles that open the possibility of direct comparison between corpora. The first challenge is reaching all the elements that express negation/speculation since there are some markers such as verbs, nouns or affixes that entail complex decisions and have not been addressed by many authors. Secondly, the elements affected by the cues should be analysed more thoroughly in order to reach a consensus on the extent of the scopes.

6.2 Discussion

This book describes previous and ongoing work on negation and speculation detection with the aim of offering a comprehensive overview of the field. It starts with an introduction to the definition of negation and speculation from different perspectives; at the same time it provides an explanation of the basic notions that one must understand in order to begin to tackle the problem. Then, it goes into detail about the concepts of negation and speculation and it provides a description of the most relevant negation and speculation detection systems found in the literature. Finally, a set of relevant resources for any researcher or developer interested in the problem is provided.

As can be seen from the different chapters of this book, negation and speculation detection is a challenging task, spanning many different areas and applications. Most of the research on negation and speculation has been done for the English language. Among the research carried out for English, two very different domains have received particular attention: the biomedical and the review domains. Comparing both of them, they differ in many aspects. The percentage of negative and speculative information in the review domain is higher than in the biomedical one because this kind of information is widely used to express opinions in the reviews which lends the text a greater degree of complexity. In addition, clinical documents are characterised by short sentences and medical terminology. In reviews, sentence length is much longer than in the clinical data and the style of the text is more literary, thereby allowing for a greater degree of linguistic richness. The review domain also tends to feature misspelling mistakes. These differences generally make negation/speculation detection more difficult in the review domain and cause the results yielded by the systems developed for this area to be lower than those obtained by the approaches for the biomedical domain.

It also highlights how the results for the cue-detection phase are higher than those for the scope recognition. This is due to the difficulty of the task of identifying the scope, where, as many authors agree, lexical information is necessary but not sufficient on its own; syntactic features must be considered also.

Although most of the work on negation and speculation is focused on the English language, the NLP community is adapting the existing tools and resources for English into other languages as well as creating other new resources which will be able to extend the catalogue of tools and corpora.

An important body of research has been published in this field in the past decade but there are still many challenges to be overcome. However, given the high number of academic projects, shared tasks and business initiatives that currently exist and that continue to be created with the objective of tackling this problem, it can be stated as a near certainty that solutions involving this task will become more and more relevant to the business community, helping the NLP community to take timely and informed decisions.

6.3 Final remarks

As mentioned in Section 1.1, and as discussed in more detail throughout the book, negation and speculation detection present a real challenge to researchers because, although they might seem easy to deal with on the surface, negation and speculation pose additional issues that derive from their interaction with many other phenomena. This means that it is a problem that is gaining relevance in recent years and that efforts are being made to improve the results obtained for this task. However, much still remains to be done since scope detector performance, among other things, remains far from the level of well-established tasks such as *parsing*.

The motivation for writing this book came from the lack of relevant manuscripts and survey texts comprehensively presenting information related to this task which will be useful for students of NLP subjects who are interested in understanding this problem in more depth as well as for researchers in this field. The main advantage of this book is that it will not only provide an overview of the state of the art in negation and speculation recognition, but that it also introduces newly developed data sets and scripts.

Glossary

Abstract	a summary of the contents of a book, article, or speech.
Affective computing	the study and development of systems and devices that can recognise, interpret, process, and simulate human affects. It is sometimes called artificial emotional intelligence or emotion AI.
Analytic negation	negative particles that are normally associated with the main verb of the clause.
Anaphora	co-reference of one expression with its antecedent. The antecedent provides the information necessary for the interpretation of the expression. This is often understood as an expression “referring” back to the antecedent.
Annotation guide-lines	instructions provided to humans for annotating linguistic features, relationships or structures in text.
Appraisal	the ways that writers or speakers express approval or disapproval for things or ideas.
Artificial intelligence	the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings. The term is frequently applied to the project of developing systems endowed with the intellectual processes characteristic of humans, such as the ability to reason, discover meaning, generalise, or learn from past experience.
Artificial neural network	computing systems loosely inspired by the biological neural networks that constitute animal brains.
Bag-of-words model	a simplifying representation used in NLP and information retrieval. In this model, a text (such as a sentence or a document) is represented as the bag (multiset) of its words, disregarding grammar and even word order but keeping multiplicity.
Baseline	a method that uses heuristics, simple summary statistics, randomness, or machine learning to provide a point of comparison for the more advanced methods.
Belief	the state of mind in which a person thinks something to be the case with or without there being empirical evidence to prove that something is the case with factual certainty.
Bigram	a sequence of two adjacent elements from a string of tokens, which are typically letters, syllables, or words.

BIO scheme	(short for beginning, inside, outside) a common tagging format for tagging tokens in a chunking task in computational linguistics. The B- prefix before a tag indicates that the tag is the beginning of a chunk, and an I- prefix before a tag indicates that the tag is inside a chunk. An O tag indicates that a token belongs to no chunk.
BLEU	(bilingual evaluation understudy) an algorithm for evaluating the quality of text which has been machine-translated from one natural language to another. Quality is considered to be the correspondence between a machine's output and that of a human.
Break-even point	the average of precision and recall when the difference between the two is minimal.
Certainty	firm conviction that something is the case.
Chunk	a sequence of words in the input that constitutes an elementary grouping of a particular syntactic category.
Chunking	(also shallow parsing) an analysis of a sentence which first identifies constituent parts of sentences (nouns, verbs, adjectives, etc.) and then links them to higher order units that have discrete grammatical meanings (noun groups or phrases, verb groups, etc.).
Classification	the problem of identifying to which of a set of categories (sub-populations) a new observation belongs, on the basis of a training set of data containing observations (or instances) whose category membership is known.
Clausal negation	a type of negation that negates an entire proposition.
Clinical note	an entry into a medical or health record made by a physician, nurse, lab technician or any other member of a patient's healthcare team.
Clinical practice guideline	systematically developed statements intended to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances.
Computational linguistics	an interdisciplinary field concerned with the statistical or rule-based modelling of natural language from a computational perspective, as well as the study of appropriate computational approaches to linguistic questions.
Computer science	the study of the theory, experimentation, and engineering that form the basis for the design and use of computers. It is the scientific and practical approach to computation and its applications and the systematic study of the feasibility, structure, expression, and mechanisation of the methodical procedures (or algorithms) that underlie the acquisition, representation, processing, storage, communication of, and access to, information.
Confidence	a state of being certain either that a hypothesis or prediction is correct or that a chosen course of action is the best or most effective.
Constituent negation	a type of negation that negates some constituent or clause.
Convolutional neural network	a class of deep, feed-forward artificial neural network that has successfully been applied to analysing visual imagery. It uses a variation of multilayer perceptrons designed to require minimal pre-processing.

Corpus	a body of linguistic data, usually in naturally occurring data in machine-readable form, which has been gathered according to a principled sampling method.
Conditional random fields	a class of statistical modelling method often applied in pattern recognition and machine learning and used for structured prediction.
Contextual valence shifter	lexical phenomena that can cause the valence of a lexical item to shift from one pole to the other or, less forcefully, to modify the valence towards a more neutral position (e.g., negation, intensifier and diminisher).
Cross-validation	a model-validation technique for assessing how the results of a statistical analysis will generalise to an independent data set.
Cue	a word that has negative or speculative meaning.
Database	an organized collection of data.
Data set	a collection of related sets of information that is composed of separate elements but can be manipulated as a unit by a computer.
Deep learning	a subset of machine learning in artificial intelligence that has networks capable of learning unsupervised from data that is unstructured or unlabelled. It imitates the workings of the human brain in processing data and creating patterns for use in decision making.
Denial	see negation.
Dependency grammar	a type of generative grammar in which grammatical structure is determined by the relationship between a governor and its dependents.
Dependency parsing	the task of recognising a sentence and assigning a syntactic structure to it.
Derivational negation	the use of negative morphemes in the derivation of lexical item.
Dictionary	a collection of words and phrases with information about them. Traditional dictionaries contain spellings, pronunciations, inflections, pronunciations, word classes, definitions, etymologies, and usage guides. A dictionary for computational purposes (sometimes called a lexicon) rarely says anything about word origin, and may say nothing about meaning or pronunciation either.
Diminishers	degree adverbs which decrease the effect of the modified item.
Disambiguation	the process of determining which sense of a word is being used in a particular context.
Discharge summary	a clinical report prepared by a physician or other health professional at the conclusion of a hospital stay or series of treatments.
Discourse	a unit of language longer than a single sentence; also refers to the use of spoken or written language in a social context.
Drug-drug interaction	a change in the effects of one drug by the presence of another drug.
Electronic health record	the systematised collection of patients' and populations' health information electronically stored in a digital format.
Entity	something that exists apart from other things, having its own independent existence.

Euphemism	a generally innocuous word or expression used in place of one that may be found offensive or suggest something unpleasant.
Epistemic modality	a sub-type of linguistic modality that deals with a speaker's evaluation or judgment of, degree of confidence in, or belief in the knowledge upon which a proposition is based.
Evaluation	a systematic determination of a subject's merit, worth and significance, using criteria governed by a set of standards.
Event	something that happens, especially when it is unusual or important.
Evidence-based medicine	an approach to medical practice intended to optimise decision-making by emphasising the use of evidence from well-designed and well-conducted research.
Evidentiality	the indication of the nature of evidence for a given statement, that is, whether evidence exists for the statement and, if so, what kind.
Expectation-maximisation algorithm	an iterative method to find maximum likelihood or maximum *a posteriori* estimates of parameters in statistical models, where the model depends on unobserved latent variables.
Factuality	the quality of being factual.
False-negative error	to falsely infer the absence of something that is present.
False-positive error	to falsely infer the presence of something that is absent.
Falsity	the state of being false or untrue.
Feature	a measurable property or characteristic of a phenomenon being observed.
Focus	the part of the scope that is most prominently or explicitly negated.
F-score	(also F_1-score or F-measure) a measure of a test's accuracy. It considers both the precision p and the recall r of the test to compute the score: p is the number of correct positive results divided by the number of all positive results returned by the classifier, and r is the number of correct positive results divided by the number of all relevant samples (all samples that should have been identified as positive). The F_1-score is the harmonic average of the precision and recall.
F_1-score	see F-score.
Gold standard	for a given task, the set of "correct" answers as created by one or more humans doing the task.
Head-driven phrase structure grammar	a highly lexicalised, constraint-based grammar. It is a type of phrase structure grammar, as opposed to a dependency grammar, and it is the immediate successor to generalised phrase structure grammar.
Hedges	linguistic tools that allow authors to indicate that they cannot justify their opinions with facts.
Hedging	see speculation.
Heuristic	a technique designed to solve a problem more quickly, used when classic methods are too slow, or to find an approximate solution, when classic methods fail to find any exact solution.
Hidden Markov model	a statistical Markov model in which the system being modelled is assumed to be a Markov process with unobserved (i.e., hidden) states.

Hypothesis	a proposed explanation for a phenomenon.
Indirect speech	a means of expressing the content of statements, questions or other utterances, without quoting them explicitly as is done in direct speech.
Information extraction	the task of automatically extracting structured information from unstructured and/or semi-structured machine-readable documents.
Information retrieval	the activity of obtaining information resources relevant to an information need from a collection of information resources.
Inter-annotator agreement	a measurement for the reliability of annotation project design as well as the resultant dataset.
Irony	the use of words to convey a meaning that is the opposite of the literal meaning of the words.
Keyword	see cue
Knowledge discovery	the process of automatically searching large volumes of data for patterns that can be considered knowledge about the data.
Lexical negation	the use of the negation morpheme (ne-) for creating negative forms of words.
Lexicon	see dictionary.
Linguistics	the study of language which involves an analysis of language form, language meaning, and language in context.
Logic	a subject concerned with the most general laws of truth that is now generally held to consist of the systematic study of the form of valid inference.
Long short-term memory model	a special kind of recurrent neural network, capable of learning long-term dependencies.
Machine learning	a field of computer science that uses statistical techniques to give computer systems the ability to "learn" (i.e., to progressively improve performance on a specific task) with data, without being explicitly programmed to do so.
Machine translation	the task of automatically converting one natural language into another, preserving the meaning of the input text, and producing fluent text in the output language.
Market research	any organised effort to gather information about target markets or customers.
Maximum entropy classifier	a probabilistic classifier which belongs to the class of exponential models.
Maximum likelihood estimation	a method of estimating the parameters of statistical models.
MeSH	(Medical Subject Headings) a comprehensive controlled vocabulary for the purpose of indexing journal articles and books in the life sciences; it serves as a thesaurus that facilitates searching.
Metalinguistic negation	the negation of what is mentioned rather than used in a sentence.
Modality	linguistic forms that allow speakers to attach expressions of belief, attitude and obligation to statements.

Morpheme	the smallest grammatical unit in a language.
Morphological negation	the negation of the verbal predicate of a declarative sentence, excluding the negation of a noun phrase, negative pronouns, and negative adverbials.
Multiword expression	a lexeme made up of a sequence of two or more lexemes that has properties that are not predictable from the properties of the individual lexemes or their normal mode of combination.
Naïve Bayes	a family of simple "probabilistic classifiers" based on applying Bayes' theorem with strong (naive) independence assumptions between the features.
Named entity recognition	a subtask of information extraction that seeks to locate and classify named entities in text into pre-defined categories such as the names of persons, organisations, locations, expressions of times, quantities, monetary values, percentages, etc.
Natural language processing	the subfield of computer science concerned with using computational techniques to learn, understand, and produce human language content.
Negation	a linguistic phenomenon that denies or rejects statements, transforming a positive sentence into a negative one.
Negative predictive value	the proportion of negative results in statistics and diagnostic tests that are true negative results.
N-gram	sequences of words of length n.
Normalisation	a process that converts a list of words to a more uniform sequence.
Opinion mining	the use of natural language processing, text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information.
Optimisation	the selection of the best element from a set of available alternatives.
Paraphrase	a restatement of the meaning of a text or passage using other words.
Part-of-speech tagging	the process of marking up a word in a text (corpus) as corresponding to a particular part-of-speech, based on both its definition and its context.
Performance	the process of calculating space and time complexity of an algorithm.
Phrase structure grammar	a type of generative grammar in which constituent structures are represented by phrase structure rules or rewrite rules.
Polarity	the grammatical systems associated with distinguishing between positive and negative clauses.
Positive predictive value	see precision.
Precision	(also called positive predictive value) the fraction of relevant instances among the retrieved instances.
Protein-protein interaction	the physical contacts of high specificity established between two or more protein molecules as a result of biochemical events steered by electrostatic forces including the hydrophobic effect.
Query	a precise request for information retrieval with database and information systems.

Recall	see sensitive.
Recognising textual entailment	the task of deciding, given two text fragments, whether the meaning of one text is entailed (can be inferred) from another text.
Recommendation system	a subclass of information filtering system that seeks to predict the "rating" or "preference" a user would give to an item.
Recurrent neural network	a class of artificial neural network where connections between units form a directed graph along a sequence. This allows the network to exhibit dynamic temporal behaviour for a time sequence. Unlike feed-forward neural networks, recurrent neural networks can use their internal state (memory) to process sequences of inputs.
Regular expression	a sequence of characters that define a search pattern. Usually this pattern is then used by string-searching algorithms for "find" or "find and replace" operations on strings, or for input validation.
Reinforcement learning	an area of machine learning inspired by behaviourist psychology, concerned with how software agents ought to take actions in an environment so as to maximise some notion of cumulative reward.
Relationship extraction	the detection and classification of semantic relationship mentions within a set of artefacts, typically from text or XML documents. The task is very similar to that of information extraction, but information extraction additionally requires the removal of repeated relations (disambiguation) and generally refers to the extraction of many different relationships.
Rhetorical relation	(or discourse relation) a description of how two segments of discourse are logically connected to one another.
Ruled-based system	a system which involves human-crafted or curated rule sets.
Scope	a text fragment governed by the corresponding cue in a sentence.
Script	a file containing a list of user commands, allowing them to be invoked once to execute in sequence.
Semantics	the linguistic and philosophical study of meaning in language, in programming languages, in formal logics, and in semiotics. It is concerned with the relationship between signifiers, like words, phrases, signs, and symbols, and what they stand for (their denotation).
Sentence boundary detection	the problem in natural language processing of deciding where sentences begin and end.
Sentiment analysis	see opinion mining.
Sensitivity	(also called true positive rate, recall, or probability of detection) the proportion of positives that are correctly identified as such.
Shallow parsing	see chunking.
Shared task	meetings organised to tackle specific problems which are challenging to tackle by lone research groups for various reasons.
Shifter	a word whose meaning changes depending on the situation, e.g. deictic words.
Specificity	(also called true negative rate) the proportion of negatives that are correctly identified as such.

Speculation	a linguistic phenomenon used to express a statement without attributing certainty to it.
Statement	a definite or clear expression of something in speech or writing.
State-of-the-art	the highest level of general development, as of a device, technique, or scientific field achieved at a particular time.
Statistical machine learning	the development of algorithms and techniques that learn from observed data by constructing stochastic models that can be used for making predictions and decisions.
Subjectivity	the perceptions, experiences, expectations, personal or cultural understanding, and beliefs specific to a person.
Supervised learning	the machine learning task of learning a function that maps an input to an output based on example input-output pairs. It infers a function from labelled training data consisting of a set of training examples. In supervised learning, each example is a pair consisting of an input object (typically a vector) and a desired output value (also called the supervisory signal). A supervised learning algorithm analyses the training data and produces an inferred function, which can be used for mapping new examples. An optimal scenario will allow for the algorithm to correctly determine the class labels for unseen instances.
Support vector machine	supervised learning models with associated learning algorithms that analyse data used for classification and regression analysis. A support vector machine constructs a hyperplane or set of hyperplanes in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks like outlier detection.
Syntactic parsing	the task of recognising a sentence and assigning a syntactic structure to it.
Syntagma	a syntactic unit or a word or phrase forming a syntactic unit.
Synthetic negation	a type of negation in which the negation words also have other functions in the sentence.
Tagger	a piece of software that reads text in a language and assigns parts-of-speech to each word (and other token), such as noun, verb, adjective, etc., although generally computational applications use more fine-grained part-of-speech tags like “noun-plural”.
Taxonomy	a model used to organise and index knowledge (stored as documents, articles, videos, etc.) so that users can find the information they are searching for.
Temporality	the state of existing within or having some relationship with time.
Test dataset	a set of examples used only to assess the performance (e.g., generalisation) of a fully specified classifier.
Text mining	the processing of unstructured (textual) information, extracting meaningful numeric indices from the text, and, thus, making the information contained in the text accessible to the various data mining (statistical and machine learning) algorithms.
Text watermarking detection	the task of producing individually identifiable copies of a source text via small manipulations to the text.

Textual entailment	a directional relationship between pairs of text expressions.
Thesaurus	a book, software program, or online service that provides synonyms for a word.
Token	an instance of a sequence of characters in some particular document that are grouped together as a useful semantic unit for processing.
Tokenisation	the process of demarcating and possibly classifying sections of a string of input characters. The resulting tokens are then passed on to some other form of processing. The process can be considered a sub-task of parsing input.
Training dataset	a dataset of examples used for learning, that is, to fit the parameters of a classifier.
Trigger	procedural code that is automatically executed in response to certain events on a particular table or view in a database.
Trustworthiness prediction	the task of distinguishing truth from lies.
Uncertainty	a situation involving ambiguous and/or unknown information.
Unified medical language system	a set of files and software that brings together many health and biomedical vocabularies and standards to enable interoperability between computer systems.
Unsupervised learning	the machine learning task of inferring a function to describe the hidden structure from "unlabelled" data (a classification or categorisation is not included in the observations).
Weasel	word or phrase aimed at creating an impression that something specific and meaningful has been said, when only a vague or ambiguous claim has in fact been made.
Word embedding	the collective name for a set of language-modelling and feature-learning techniques in natural language processing where words or phrases from the vocabulary are mapped onto vectors of real numbers. Conceptually it involves a mathematical embedding from a space with one dimension per word to a continuous vector space with a much lower dimension.
Word sense disambiguation	an open problem of natural language processing and ontology. The aim is to identify which sense of a word (i.e., meaning) is used in a sentence, when the word has multiple meanings.
Workflow	a view or representation of real work. The flow being described may refer to a document, service, or product that is being transferred from one step to another. In a sequential workflow, each step is dependent on the occurrence of the previous step; in a parallel workflow, two or more steps can occur concurrently.

References

Abu-Jbara, A., & Radev, D. (2012). Umichigan: A conditional random field model for resolving the scope of negation. *Proceedings of the First Joint Conference on Lexical and Computational Semantics, Vol. 1: Proceedings of the Main Conference and the Shared Task, and Vol. 2: Proceedings of the Sixth International Workshop on Semantic Evaluation* (pp. 328–334).

Adel, H., & Schütze, H. (2016). Exploring different dimensions of attention for uncertainty detection. *arXiv Preprint arXiv:1612.06549*.

Afzal, Z., Pons, E., Kang, N., Sturkenboom, M. C., Schuemie, M. J., & Kors, J. A. (2014). ContextD: An algorithm to identify contextual properties of medical terms in a dutch clinical corpus. *BMC Bioinformatics*, 15(1), 373. https://doi.org/10.1186/s12859-014-0373-3

Agarwal, S., & Yu, H. (2010). Biomedical negation scope detection with conditional random fields. *Journal of the American Medical Informatics Association: JAMIA*, 17(6), 696–701. https://doi.org/10.1136/jamia.2010.003228; https://doi.org/10.1136/jamia.2010.003228

AL-Khawaldeh, F. T. (2015). A study of the effect of resolving negation and sentiment analysis in recognizing text entailment for arabic. *World of Computer Science and Information Technology Journal (WCSIT)*, 5(7), 124–128.

Al-Sabbagh, R., Girju, R., & Diesner, J. (2015). A unified framework to identify and extract uncertainty cues, holders, and scopes in one fell-swoop. In A. Gelbukh (Ed.), *Computational Linguistics and Intelligent Text Processing. CICLing 2015, Cairo, Egypt, April 20–24* (Lecture Notes in Computer Science, Vol. 9041) (pp. 310–334). Cham, Switzerland: Springer.

Androutsopoulos, I., & Malakasiotis, P. (2009). A survey of paraphrasing and textual entailment methods. *arXiv Preprint arXiv:0912.3747. Published in Journal of Artificial Intelligence*, 38(1), 135–187, 2010.

Apostolova, E., Tomuro, N., & Demner-Fushman, D. (2011). Automatic extraction of lexico-syntactic patterns for detection of negation and speculation scopes. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers, Vol. 2*, (pp. 283–287).

Aramaki, E., Morita, M., Kano, Y., & Ohkuma, T. (2016). Overview of the NTCIR-12 MedNLP-Doc task. *Proceedings of NTCIR-12*, 71–75.

Asmi, A., & Ishaya, T. (2012). Negation identification and calculation in sentiment analysis. *The Second International Conference on Advances in Information Mining and Management, Venice, Italy, 25–29 June* (pp. 1–7).

Attardi, G., Cozza, V., & Sartiano, D. (2015). Detecting the scope of negations in clinical notes. In C. Bosco, S. Tonelli & F. M. Zanotto (Eds.), *Proceedings of the Second Italian Conference on Computational Linguistics (CLIC-it 2015), Trento, 3–4 December, 2015* (pp. 14–19). Torino: Academia University Press. https://doi.org/10.4000/books.aaccademia.1286

Averbuch, M., Karson, T., Ben-Ami, B., Maimon, O., & Rokach, L. (2004). Context-sensitive medical information retrieval. In M. Fieschi et al. (Eds.), *Proceedings of the 11th World Congress on Medical Informatics (MEDINFO-2004)* (pp. 1–8). Amsterdam: IOS Press.

Baker, K., Bloodgood, M., Dorr, B. J., Filardo, N. W., Levin, L. S., & Piatko, C. D. (2010). A modality lexicon and its use in automatic tagging. *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC'10)* (pp. 1402–1407).

Baker, K., Bloodgood, M., Dorr, B. J., Callison-Burch, C., Filardo, N. W., Piatko, C., Leven, L. & Miller, S. (2012). Modality and negation in simt use of modality and negation in semantically-informed syntactic MT. *Computational Linguistics*, 38(2), 411–438. https://doi.org/10.1162/COLI_a_00099

Ballesteros, M., Francisco, V., Díaz, A., Herrera, J., & Gervás, P. (2012). Inferring the scope of negation in biomedical documents. In A. Gelbukh (Ed.), *Computational Linguistics and Intelligent Text Processing. 14th International Conference, CICLing 2013, Karlovasi, Samos, Greece, March 24–30, 2013, Proceedings, Part II* (Lecture Notes in Computer Science, Vol. 7181) (pp. 363–375). Heidelberg: Springer.

Banjade, R., & Rus. (2016) DT-neg: Tutorial dialogues annotated for negation scope and focus in context. *LREC 2016 Portorož Proceedings*.

Beltagy, I., Roller, S., Cheng, P., Erk, K., & Mooney, R. J. (2017). Representing meaning with a combination of logical and distributional models. *Computational Linguistics*, 42(4), 763–808. https://doi.org/10.1162/COLI_a_00266

Benamara, F., Chardon, B., Mathieu, Y. Y., Popescu, V., & Asher, N. (2012). How do negation and modality impact opinions? *Proceedings of the Workshop on Extra-Propositional Aspects of Meaning in Computational Linguistics, Jeju, Republic of Korea, 13 July* (pp. 10–18). Stroudsburg, PA: Association for Computational Linguistics (ACL).

Blanco, E., & Moldovan, D. (2011a). Semantic representation of negation using focus detection. *ACL-HILT 2011. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Portland, OR, 19–24 June, 2011, Vol. 1: Long Papers* (pp. 581–589). Stroudsburg, PA: Assocation for Computational Lingusitics (ACL).

Blanco, E., & Moldovan, D. I. (2011b). Some issues on detecting negation from text. In R. C. Murray & P. M. McCarthy (Eds.), *Proceedings of the Twenty-Fourth International Florida Artificial Intelligence Research Society Conference (FLAIRS)*. Palo Alto CA: Association for the Advancement of Artificial Intelligence.

Bokharaeian, B., Díaz Esteban, A., & Ballesteros Martínez, M. (2013). Extracting drug-drug interaction from text using negation features. *Procesamieto de Lenguage Natural*, 51, 49–56.

Bokharaeian, B., Diaz, A., & Chitsaz, H. (2016). Enhancing extraction of drug-drug interaction from literature using neutral candidates, negation, and clause dependency. *PloS One*, 11(10), e0163480. https://doi.org/10.1371/journal.pone.0163480

Bokharaeian, B., Diaz, A., Neves, M., & Francisco, V. (2014). Exploring negation annotations in the DrugDDI corpus. *Fourth Workshop on Building and Evaluating Resources for Health and Biomedical Text Processing (BIOTxtM 2014)*.

Buchholz, S., & Marsi, E. (2006). CoNLL-X shared task on multilingual dependency parsing. *CoNLL-X '06. Proceedings of the Tenth Conference on Computational Natural Language Learning, New York City, NY, 8–9 June* (pp. 149–164). Stroudsburg, PA: Association for Computational Linguistics (ACL). https://doi.org/10.3115/1596276.1596305

Cairns, B. L., Nielsen, R. D., Masanz, J. J., Martin, J. H., Palmer, M. S., Ward, W. H., & Savova, G. K. (2011). The MiPACQ clinical question answering system. *AMIA Annual Symposium Proceedings* (pp. 171–180).

Carrillo-de-Albornoz, J., & Plaza, L. (2013). An emotion-based model of negation, intensifiers, and modality for polarity and intensity classification. *Journal of the Association for Information Science and Technology*, 64(8), 1618–1633.

ChandraKala, S., & Sindhu, C. (2012). Opinion mining and sentiment classification: A survey. *ICTACT Journal on Soft Computing* 3(1), 421–427.

Chang, C., & Lin, C. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3), 27.

Chapman, B. E., Lee, S., Kang, H. P., & Chapman, W. W. (2011). Document-level classification of CT pulmonary angiography reports based on an extension of the ConText algorithm. *Journal of Biomedical Informatics*, 44(5), 728–737. https://doi.org/10.1016/j.jbi.2011.03.011

Chapman, W. W., Bridewell, W., Hanbury, P., Cooper, G. F., & Buchanan, B. G. (2001a). Evaluation of negation phrases in narrative clinical reports. *Proceedings of the American Medical Informatics Association Symposium* (pp. 105–109).

Chapman, W. W., Hilert, D., Velupillai, S., Kvist, M., Skeppstedt, M., Chapman, B. E., Conway, M. Tharp, M., Mowery, D. L., & Deleger, L. (2013). Extending the NegEx lexicon for multiple languages. In C. U. Lehmann, E. Ammenwerth & C. Nøhr (Eds.), *MEDINFO 2013. Proceedings of the 14th World Congress on Medical and Health Informatics* (Studies in Health Technology and Informatics 192) (pp. 677–681). Amsterdam: IOS Press.

Chapman, W. W., Bridewell, W., Hanbury, P., Cooper, G. F., & Buchanan, B. G. (2001b). A simple algorithm for identifying negated findings and diseases in discharge summaries. *Journal of Biomedical Informatics*, 34, 301–310. https://doi.org/10.1006/jbin.2001.1029

Chen, Z., Zou, B., Zhu, Q., & Li, P. (2013). Chinese negation and speculation detection with conditional random fields. In G. Zhou, J. Li, D. Zhao, & Y. Feng (Eds.), *Natural Language Processing and Chinese Computing. Second CCF Conference, NLPCC 2013, Chongqing, China, November 15–19, 2013. Proceedings* (Communications in Computer and Information Science, Vol. 400) (pp. 30–40). Heidelberg: Springer.

Chowdhury, M. F. M., & Lavelli, A. (2013). Exploiting the scope of negations and heterogeneous features for relation extraction: A case study for drug-drug interaction extraction. *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Atlanta, 09–15 June* (pp. 765–771).

Chowdhury, M., & Mahbub, F. (2012). FBK: Exploiting phrasal and contextual clues for negation scope detection. *Proceedings of the First Joint Conference on Lexical and Computational Semantics, Vol. 1: Proceedings of the Main Conference and the Shared Task, and Vol. 2: Proceedings of the Sixth International Workshop on Semantic Evaluation* (pp. 340–346).

Collins, M., Koehn, P., & Kučerová, I. (2005). Clause restructuring for statistical machine translation. *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, University of Michigan, 25–30 June* (pp. 531–540). Stroudsburg, PA: Association for Computational Linguistics (ACL).

Costumero, R., Lopez, F., Gonzalo-Martín, C., Millan, M., & Menasalvas, E. (2014). An approach to detect negation on medical documents in Spanish. In D. Ślęzak, A. H. Tan, J. F. Peters, & L. Schwabe (Eds.), *Brain Informatics and Health (BIH 2014)* (Lecture Notes in Computer Science, Vol. 8609) (pp. 366–375). Cham, Switzerland: Springer.

Cotik, V., Roller, R., Xu, F., Uszkoreit, H., Budde, K., & Schmidt, O. D. (2016). Negation detection in clinical reports written in German. *Proceedings of the Fifth Workshop on Building and Evaluating Resources for Biomedical Text Mining (BioTxtM 2016), Osaka, Japan, December 12th 2016*, 115–124.

Cotik, V., Stricker, V., Vivaldi, J., & Rodriguez, H. (2016). Syntactic methods for negation detection in radiology reports in spanish. *BioNLP 2016, Proceedings of the 15th Workshop on Biomedical Natural Language Processing* (pp. 156–165). Stroudsburg, PA: Association for Computational Linguistics (ACL). https://doi.org/10.18653/v1/W16-2921

Councill, I. G., McDonald, R., & Velikovich, L. (2010). What's great and what's not: Learning to classify the scope of negation for improved sentiment analysis. In *Proceedings of the Workshop on Negation and Speculation in National Language Processing, NeSp-NLP* (pp. 51–59). Stroudsburg, PA: Association for Computational Linguistics (ACL).

Cruz Díaz, N. P., Maña López, M. J., Vázquez, J. M., & Álvarez, V. P. (2012). A machine-learning approach to negation and speculation detection in clinical texts. *Journal of the American Society for Information Science and Technology*, 63(7), 1398–1410. https://doi.org/10.1002/asi.22679

Cruz, N. P., Taboada, M., & Mitkov, R. (2015). A machine-learning approach to negation and speculation detection for sentiment analysis. *Journal of the Association for Information Science and Technology*, 67(9), 526–558.

Dadvar, M., Hauff, C., & de Jong, F. (2011). Scope of negation detection in sentiment analysis. In *DIR 2011. Proceedings of the Dutch-Belgian Information Retrieval Workshop, University of Amsterdam* (pp. 16–20).

Daelemans, W., Van Den Bosch, A., & Weijters, T. (1997). IGTree: Using trees for compression and classification in lazy learning algorithms. *Artificial Intelligence Review*, 11(1–5), 407–423. https://doi.org/10.1023/A:1006506017891

Dagan, I., Glickman, O., & Magnini, B. (2006). The PASCAL recognising textual entailment challenge. In J. Quiñonero-Candela, I. Dagan, B. Magnini & F. d'Alché-Buc (Eds.), *Machine Learning Challenges. Evaluating Predictive Uncertainty, Visual Object Classification, and Recognising Tectual Entailment. First PASCAL Machine Learning Challenges Workshop, MLCW 2005, Southampton, UK, April 11–13, 2005, Revised Selected Papers* (Lecture Notes in Artificial Intelligence 3944) (pp. 177–190). Heidelberg: Springer.

Dalianis, H., Hassel, M., Henriksson, A., & Skeppstedt, M. (2012). *Stockholm EPR Corpus: A Clinical Database Used to Improve Health Care*. Stockholm: DSV, Stockholm University.

de Albornoz, J. C., Plaza, L., Díaz, A., & Ballesteros, M. (2012). Ucm-i: A rule-based syntactic approach for resolving the scope of negation. *Proceedings of the First Joint Conference on Lexical and Computational Semantics, Vol. 1: Proceedings of the Main Conference and the Shared Task, and Vol. 2: Proceedings of the Sixth International Workshop on Semantic Evaluation* (pp. 282–287).

De Haan, F. (1997). *The Interaction of Modality and Negation: A Typological Study*. London: Routledge.

de Marneffe, M., MacCartney, B., Grenager, T., Cer, D., Rafferty, A., & Manning, C. D. (2006). Learning to distinguish valid textual entailments. In *Proceedings of the Second Pascal Recognising Textual Entailment Challenge Workshop (RTE-2) , Venice, Italy*.

De Marneffe, M., MacCartney, B., & Manning, C. D. (2006). Generating typed dependency parses from phrase structure parses. *Proceedings of LREC, 6* (pp. 449–454).

Deléger, L., & Grouin, C. (2012). Detecting negation of medical problems in french clinical notes. *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium, Miami, FL, 28–30 January* (pp. 697–702). https://doi.org/10.1145/2110363.2110443

Denny, J. C., Miller, R. A., Waitman, L. R., Arrieta, M. A., & Peterson, J. F. (2009). Identifying QT prolongation from ECG impressions using a general-purpose natural language processor. *International Journal of Medical Informatics*, 78, S34–S42. https://doi.org/10.1016/j.ijmedinf.2008.09.001

Denny, J. C., Choma, N. N., Peterson, J. F., Miller, R. A., Bastarache, L., Li, M., & Peterson, N. B. (2012). Natural language processing improves identification of colorectal cancer testing in the electronic medical record. *Medical Decision Making: An International Journal of the Society for Medical Decision Making*, 32(1), 188–197. https://doi.org/10.1177/0272989X11400418; https://doi.org/10.1177/0272989X11400418

Di Marco, C., Kroon, F. W., & Mercer, R. E. (2006). Using hedges to classify citations in scientific articles. In J. G. Shanahan, Y. Qu, & J. Wiebe (Eds.), *Computing Attitude and Affect in Text: Theory and Applications* (The Information Retrieval Series 20) (pp. 247–263). Dordrecht, Netherlands: Springer. https://doi.org/10.1007/1-4020-4102-0_19

Diamantini, C., Mircoli, A., & Potena, D. (2016). A negation handling technique for sentiment analysis. *International Conference on Collaboration Technologies and Systems (CTS), Orlando, FL, 31 October – 4 November* (pp. 188–195).

Dowty, D. (1994). The role of negative polarity and concord marking in natural language reasoning. In M. Harvey & L. Santelmann (Eds.), *Proceedings of SALT, 4. Proceedings of the 4th Semantics and Linguistic Theory Conference, held May 6–8, 1994, at the University of Rochester* (pp. 114–144).

Elazhary, H. (2017). NegMiner: An automated tool for mining negations from electronic narrative medical documents. *I.J. Intelligent Systems and Applications*, 4, 14–22. https://doi.org/10.5815/ijisa.2017.04.02

Elkin, P. L., Brown, S. H., Bauer, B. A., Husser, C. S., Carruth, W., Bergstrom, L. R., & Wahner-Roedler, D. L. (2005). A controlled trial of automated classification of negation from clinical notes. *BMC Medical Informatics and Decision Making*, 5(1), 13. https://doi.org/10.1186/1472-6947-5-13

Enger, M., Velldal, E., & Øvrelid, L. (2017). An open-source tool for negation detection: A maximum-margin approach. *Proceedings of the Workshop Computational Semantics Beyond Events and Roles (SemBEaR 2017), Valencia, Spain, 3–4 April* (pp. 64–69). Stroudsburg, PA: Association for Computational Linguistics (ACL). https://doi.org/10.18653/v1/W17-1810

Fancellu, F., Lopez, A., & Webber, B. L. (2016). Neural networks for negation scope detection. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, Berlin, Germany, 7–12 August, Vol. 1* (pp. 495–504). Stroudsburg, PA: Association for Computational Linguistics (ACL).

Fancellu, F., & Webber, B. (2015). Translating negation: A manual error analysis. *Second Workshop on Extra-Propositional Aspects of Meaning in Computational Semantics (ExProM 2015), Denver, CO, 5 June*, 1–10. Stroudsburg, PA: Association for Computational Linguistics (ACL).

Fancellu, F., & Webber, B. L. (2014). Applying the semantics of negation to SMT through n-best list re-ranking. *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2014, April 26–30, 2014, Gothenburg, Sweden* (pp. 598–606). Stroudsburg, PA: Association for Computational Linguistics (ACL).

Farkas, R., & Szarvas, G. (2008). Automatic construction of rule-based ICD-9-CM coding systems. *BMC Bioinformatics*, 9(3), S10. https://doi.org/10.1186/1471-2105-9-S3-S10

Farkas, R., Vincze, V., Móra, G., Csirik, J., & Szarvas, G. (2010). The CoNLL-2010 shared task: Learning to detect hedges and their scope in natural language text. *Proceedings of the Fourteenth Conference on Computational Natural Language Learning – Shared Task, Uppsala, Sweden, 15–16 July* (pp. 1–12).

Farooq, U., Mansoor, H., Nongaillard, A., Ouzrout, Y., & Qadir, M. A. (2017). Negation handling in sentiment analysis at sentence level. *JCP*, 12(5), 470–478.

Fiszman, M., Rindflesch, T. C., & Kilicoglu, H. (2006). Summarizing drug information in medline citations. *AMIA Annual Symposium Proceedings 2006* (pp. 254–258).

Friedman, C., Alderson, P. O., Austin, J. H., Cimino, J. J., & Johnson, S. B. (1994). A general natural-language text processor for clinical radiology. *Journal of the American Medical Informatics Association: JAMIA*, 1(2), 161–174. https://doi.org/10.1136/jamia.1994.95236146

Fujikawa, K., Seki, K., & Uehara, K. (2013). NegFinder: A web service for identifying negation signals and their scopes. *Information and Media Technologies*, 8(3), 884–889.

Ganter, V., & Strube, M. (2009). Finding hedges by chasing weasels: Hedge detection using wikipedia tags and shallow linguistic features. *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers, Suntec, Singapore, 4 August* (pp. 173–176). https://doi.org/10.3115/1667583.1667636

Gaona, M. A. R., Gelbukh, A. F., & Bandyopadhyay, S. (2010). Recognizing textual entailment with statistical methods. *MCPR 2010. Second Mexican Conference on Pattern Recognition, Puebla, Mexico, 27–29 September* (pp. 372–381).

Georgescul, M. (2010). A hedgehop over a max-margin framework using hedge cues. *Proceedings of the Fourteenth Conference on Computational Natural Language Learning – Shared Task, Uppsala, Sweden, 15–16 July* (pp. 26–31).

Gindl, S., Kaiser, K., & Miksch, S. (2008). Syntactical negation detection in clinical practice guidelines. 187–192. In S. Andersen, G. Klein, S. Schulz, J. Aarts & M. Mazzoleni (Eds.), *eHealth Beyond the Horizon – Get IT There* (Studies in Health Technology and Informatics, 136) (pp. 187). Amsterdam: IOS Press.

Goldin, I., & Chapman, W. W. (2003). Learning to detect negation with 'not'in medical texts. *Proceedings of the SIGIR '03 Workshop on Text Analysis and Search for Bioinformatics, Toronto, Canada, 28 July – 1 August*.

Goryachev, S., Sordo, M., Zeng, Q. T., & Ngo, L. (2006). *Implementation and Evaluation of Four Different Methods of Negation Detection*. Boston, MA: DSG.

Grabar, N., & Hamon, T. (2009). Exploitation of speculation markers to identify the structure of biomedical scientific writing. *AMIA Annual Symposium Proceedings 2009* (pp. 203–207).

Gros, O., & Stede, M. (2013). Determining negation scope in german and english medical diagnoses. In M. Taboada & R. Trnavac (Eds.), *Nonveridicality and Evaluation. Theoretical, Computational and Corpus Approaches* (pp. 113–126). Leiden: Brill.

Gyawali, B., & Solorio, T. (2012). UABCoRAL: A preliminary study for resolving the scope of negation. *Proceedings of the First Joint Conference on Lexical and Computational Semantics, Vol. 1: Proceedings of the Main Conference and the Shared Task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation* (pp. 275–281).

Harabagiu, S., Hickl, A., & Lacatusu, F. (2006). Negation, contrast and contradiction in text processing. *Proceedings of the National Conference on Artificial Intelligence (AAAI 6),Boston, MA* (pp. 755–762).

Harkema, H., Dowling, J. N., Thornblade, T., & Chapman, W. W. (2009). ConText: An algorithm for determining negation, experiencer, and temporal status from clinical reports. *Journal of Biomedical Informatics*, 42(5), 839–851. https://doi.org/10.1016/j.jbi.2009.05.002

He, B., Dong, B., Guan, Y., Yang, J., Jiang, Z., Yu, Q., … Qu, C. (2017). Building a comprehensive syntactic and semantic corpus of Chinese clinical texts. *Journal of Biomedical Informatics*, 69, 203–217. https://doi.org/10.1016/j.jbi.2017.04.006

He, H., Fancellu, F., & Webber, B. (2017). Neural networks for negation cue detection in Chinese. *Proceedings of the Workshop Computational Semantics Beyond Events and Roles (SemBEaR 2017), Valencia, Spain, 3–4 April* (pp. 59–63). Stroudsburg, PA: Association for Computational Linguistics (ACL). https://doi.org/10.18653/v1/W17-1809

Herrero-Zazo, M., Segura-Bedmar, I., Martínez, P., & Declerck, T. (2013). The DDI corpus: An annotated corpus with pharmacological substances and drug–drug interactions. *Journal of Biomedical Informatics*, 46(5), 914–920. https://doi.org/10.1016/j.jbi.2013.07.011

Hintikka, J. (2002). Negation in logic and in natural language. *Linguistics and Philosophy*, 25(5–6), 585–600. https://doi.org/10.1023/A:1020895229197

Hirschberg, J., & Manning, C. D. (2015). Advances in natural language processing. *Science* 349(6245), 261–266. https://doi.org/10.1126/science.aaa8685

Hogenboom, A., van Iterson, P., Heerschop, B., Frasincar, F., & Kaymak, U. (2011). Determining negation scope and strength in sentiment analysis. *IEEE International Conference on Systems, Man, and Cybernetics (SMC), Anchorage, AK, 9–12 October* (pp. 2589–2594).

Horn, L. R., & Kato, Y. (2000). *Negation and Polarity: Syntactic and Semantic Perspectives: Syntactic and Semantic Perspectives* Oxford: Oxford University Press.

Horn, L. R. (1989). *A Natural History of Negation*. Chicago, IL: University of Chicago Press.

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. *Proceedings of the 19th National Conference on Artifical Intelligence, July 25–29, 2004, San Jose, CA*, 755–760.

Huang, Y., & Lowe, H. J. (2007). A novel hybrid approach to automated negation detection in clinical radiology reports. *Journal of the American Medical Informatics Association*, 14(3), 304–311. https://doi.org/10.1197/jamia.M2284

Huddleston, R., & Pullum, G. K. (2002). *The Cambridge Grammar of the English Language* (pp.1–23). Cambridge: Cambridge University Press. https://doi.org/10.1017/9781316423530

Hussein, Doaa Mohey El-Din Mohamed. (2016). A survey on sentiment analysis challenges. *Journal of King Saud University – Engineering Sciences.* https://doi.org/10.1016/j.jksues.2016.04.002

Hyland, K. (1995). The author in the text: Hedging scientific writing. *Hong Kong Papers in Linguistics and Language Teaching*, 18, 33–42.

Hyland, K. (1996). Talking to the academy forms of hedging in science research articles. *Written Communication*, 13(2), 251–281. https://doi.org/10.1177/0741088396013002004

Hyland, K. (1998). *Hedging in Scientific Research Articles* (Pragmatics & Beyond New Series 54). Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.54

Jean, P., Harispe, S., Ranwez, S., Bellot, P., & Montmain, J. (2016). Uncertainty detection in natural language: A probabilistic model. *WIMS '16. Proceedings of the 6th International Conference on Web Intelligence, Mining and Semantics, Nîmes, France, 13–15 June*. New York, NY: ACM.

Ji, F., Qiu, X., & Huang, X. (2010). Exploring uncertainty sentences in chinese. *Proceedings of the 16th China Conference on Information Retrieval* (pp. 594–601).

Jia, L., Yu, C., & Meng, W. (2009). The effect of negation on sentiment analysis and retrieval effectiveness. *CIKM '09. Proceedings of the 18th ACM Conference on Information and Knowledge Management, Hong Kong, China, 2–6 November* (pp. 1827–1830). New York, NY: ACM.

Jia, Z., Li, H., Ju, M., Zhang, Y., Huang, Z., Ge, C., & Duan, H. (2014). A finite-state automata based negation detection algorithm for chinese clinical documents. *PIC 2014. Proceedings of 2014 IEEE International Conference on Progress in Informatics and Computing, Shanghai, 16–18 May*, (pp. 128–132).

Jimenez, S., Duenas, G., Baquero, J., Gelbukh, A. F., Bátiz, A. J. D., & Mendizábal, A. (2014). UNAL-NLP: Combining soft cardinality features for semantic textual similarity, relatedness and entailment. *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)* (pp. 732–742). https://doi.org/10.3115/v1/S14-2131

Jiménez-Zafra, S. M., Taulé, M., Martín-Valdivia, M. T., Ureña-López, L. A., & Martí, M. A. (2017). SFU ReviewSP-NEG: A spanish corpus annotated with negation for sentiment analysis. A typology of negation patterns. *Language Resources and Evaluation* (pp. 1–37).

Jurafsky, D., & Martin, J. H. (2008). *Speech and Language Processing*. New York, NY: Pearson Education.

Kang, T., Zhang, S., Xu, N., Wen, D., Zhang, X., & Lei, J. (2017). Detecting negation and scope in chinese clinical notes using character and word embedding. *Computer Methods and Programs in Biomedicine*, 140, 53–59. https://doi.org/10.1016/j.cmpb.2016.11.009

Kenji, A., & Tanaka-Ishii, K. (2014). Sentence hedge detection without cue annotation: A heuristic cue selection approach. *Information and Media Technologies*, 9(2), 173–186.

Kennedy, A., & Inkpen, D. (2006). Sentiment classification of movie and product reviews using contextual valence shifters. *Computational Intelligence*, 22(2), 110–125. https://doi.org/10.1111/j.1467-8640.2006.00277.x

Kilicoglu, H., & Bergler, S. (2008). Recognizing speculative language in biomedical research articles: A linguistically motivated perspective. *BMC Bioinformatics*, 9(Suppl 11), S10. https://doi.org/10.1186/1471-2105-9-S11-S10

Kim, J., Ohta, T., Pyysalo, S., Kano, Y., & Tsujii, J. (2009). Overview of BioNLP '09 shared task on event extraction. *BioNLP '09. Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing – Shared Task, Boulder, CO, 5 June* (pp. 1–9). Stroudsburg, PA: Association for Computational Linguistics (ACL).

Kim, J., Ohta, T., & Tsujii, J. (2008). Corpus annotation for mining biomedical events from literature. *BMC Bioinformatics*, 9(1), 10. https://doi.org/10.1186/1471-2105-9-10

Kim, J. D., Ohta, T., Tateisi, Y., & Tsujii, J. (2003). GENIA corpus-Semantically annotated corpus for bio-textmining. *Bioinformatics* 19(Suppl 1), i180–i182. https://doi.org/10.1093/bioinformatics/btg1023

Kiritchenko, S., & Mohammad, S. (2016). The effect of negators, modals, and degree adverbs on sentiment composition. *Proceedings of the Workshop on Computation Approaches to Subjectivity and Sentiment Analysis (Wassa), San Diego, CA, 12–17 June* (pp. 43–52). Stroudsburg, PA: Association for Computiational Linguistics (ACL).

Klima, E. S. (1964). Negation in english. In J. A. Fodor & J. J. Katz (Eds.), *The Structure of Language* (pp. 245–323). Englewood Cliffs, NJ: Prentice-Hall.

Kolchyna, O., Souza, T. T., Treleaven, P., & Aste, T. (2015). Twitter sentiment analysis: Lexicon method, machine learning method and their combination. *arXiv Preprint arXiv:1507.00955*.

Konstantinova, N., & de Sousa, S. C. (2011). Annotating negation and speculation: The case of the review domain. *Proceedings of the Second Student Research Workshop associated with RANLP 2011*, RANLP 2011 Organising Committee (pp. 139–144). <http://aclweb.org/anthology/R11-2022>

Konstantinova, N., de Sousa, S., Cruz, N., Maña, M. J., Taboada, M., & Mitkov, R. (2012). A review corpus annotated for negation, speculation and their scope. *LREC* (pp. 3190–3195).

Koopman, B., Bruza, P. D., Sitbon, L., & Lawley, M. J. (2010). Analysis of the effect of negation on information retrieval of medical data. *Proceedings of 15th Australasian Document Computing Symposium (ADCS)*. <http://koopman.id.au/papers/negation_ir.pdf>

Koopman, B., & Zuccon, G. (2014). Understanding negation and family history to improve clinical information retrieval. *Proceedings of the 37th Annual International ACM SIGIR Conference on Research & Development in Information Retrieval* (pp. 971–974).

Kratzer, A. (1981). The notional category of modality. In H-J. Elkmeyer & H. Rieser (Eds.), *Words, Worlds, and Contexts* (pp. 38–74). Berlin: De Gruyter.

Kuhn, L., & Eickhoff, C. (2016). Implicit negative feedback in clinical information retrieval. *arXiv Preprint arXiv:1607.03296*.

Lai, A., & Hockenmaier, J. (2014). Illinois-LH: A denotational and distributional approach to semantics. *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), Dublin, Ireland, 23–23 August* (pp. 329–334). https://doi.org/10.3115/v1/S14-2055

Laka, I. (2013). Negation in syntax: On the nature of functional categories and projections. *Anuario Del Seminario De Filología Vasca“ Julio De Urquijo*”, 25(1), 65–136.

Lakoff, G. (1972). Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Papers from the Eight Regional Meeting of the Chicago Linguistics Society* (pp. 183–228). Chicago, IL: CLS.

Lapponi, E., Read, J., & Ovrelid, L. (2012). Representing and resolving negation for sentiment analysis. In J. Vreeken, C. Ling, M. J. Zaki, A. Siebes, J. X. Ju, B. Goethals, G. I. Webb, & X. Wu (Eds.), *Data Mining Workshops (ICDMW)* (pp. 687–692). Washington DC: IEEE Computer Society.

Lapponi, E., Velldal, E., Øvrelid, L., & Read, J. (2012). UiO 2: Sequence-labeling negation using dependency features. *Proceedings of the First Joint Conference on Lexical and Computational Semantics – Vol. 1: Proceedings of the Main Conference and the Shared Task, and Vol. 2: Proceedings of the Sixth International Workshop on Semantic Evaluation* (pp. 319–327).

Lawler, J. (2010). Negation and negative polarity. In P. C. Hogan (Ed.), *The Cambridge Encyclopedia of the Language Sciences* (pp. 554–555). Cambridge: Cambridge University Press.

Lazib L., Zhao Y., Qin B., Liu T. (2016) Negation scope detection with recurrent neural networks models in review texts. In W. Che et al. (Eds.), *Social Computing. ICYCSEE 2016* (Communications in Computer and Information Science, Vol. 623). Singapore: Springer.

Li, X., Gao, W., & Shavlik, J. W. (2014). Detecting semantic uncertainty by learning hedge cues in sentences using an HMM. *Proceedings of Workshop on Semantic Matching in Information Retrieval, Queensland, Australia, 11 July* (pp. 30–37).

Lien, E., & Kouylekov, M. (2015). Semantic parsing for textual entailment. *Proceedings of the 14th International Conference on Parsing Technologies*, Bilbao, Spain, 22–24 July, 2014 (pp. 40–72). Stroudsburg, PA: Association for Computational Linguistics (ACL). https://doi.org/10.18653/v1/W15-2205.

Light, M., Qiu, X. Y., & Srinivasan, P. (2004). The language of bioscience: Facts, speculations, and statements in between. *Proceedings of BioLink 2004 Workshop on Linking Biological Literature, Ontologies and Databases: Tools for Users* (pp. 17–24).

Lindberg, D. A., Humphreys, B. L., & McCray, A. T. (1993). The unified medical language system. *Methods Inf. Med.*, 32, 281–291. https://doi.org/10.1055/s-0038-1634945

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1), 1–167. https://doi.org/10.2200/S00416ED1V01Y201204HLT016

Macdonald, C., & Ounis, I. (2006). The TREC Blogs06 collection: Creating and analysing a blog test collection. *Department of Computer Science, University of Glasgow Tech Report TR-2006-224, 1*, 3.1–4.1.

Markkanen, R., & Schröder, H. (1989). Hedging as a translation problem in scientific texts. In C. Laurén & M. Nordman (Eds.), *Special Languages: From Human Thinking to Thinking Machines* (pp. 171–175). Clevedon: Multilingual Matters.

Martınez-Cámara, E., Martın-Valdivia, M., Molina-González, M., & Urena-López, L. (2013). Bilingual experiments on an opinion comparable corpus. *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, Atlanta, GA, 14 June*, (pp. 87–93).

Medlock, B. (2008). Exploring hedge identification in biomedical literature. *Journal of Biomedical Informatics*, 41(4), 636–654. https://doi.org/10.1016/j.jbi.2008.01.001

Medlock, B., & Briscoe, T. (2007). Weakly supervised learning for hedge classification in scientific literature. *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL'07)', Association for Computational. Linguistics, Prague* (pp. 992–999). Stroudsburg, PA: Association for Computational Linguistics (ACL).

Mehrabi, S., Krishnan, A., Sohn, S., Roch, A. M., Schmidt, H., Kesterson, J., … Liu, H. (2015). DEEPEN: A negation detection system for clinical text incorporating dependency relation into NegEx. *Journal of Biomedical Informatics*, 54, 213–219. https://doi.org/10.1016/j.jbi.2015.02.010

Meystre, S. M., Savova, G. K., Kipper-Schuler, K. C., & Hurdle, J. F. (2008). Extracting information from textual documents in the electronic health record: A review of recent research. *Yearb Med Inform*, 35(128), 44

Miller, G. A. (1995). WordNet: A lexical database for english. *Communications of the ACM*, 38(11), 39–41. https://doi.org/10.1145/219717.219748

Mitchell, K. J. (2004). Implementation and evaluation of a negation tagger in a pipeline-based system for information extraction from pathology reports. *Medical Informatics*, 107, 663–667.

Mizuno, J., Kruengkrai, C., Ohtake, K., Hashimoto, C., Torisawa, K., Kloetzer, J., & Inui, K. (2015). Recognizing complex negation on twitter. *PACLIC 2015. 29th Pacific Asia Conference on Language, Information and Computation, Shanghai, 30 October – 1 November* (pp. 544–552).

Mohammad, S. M. 2016. Sentiment analysis: Detecting valence, emotions, and other affectual states from text. In H. Meiselman (Ed.) *Emotion Measurement* (pp. 201–238). Amsterdam: Elsevier. https://doi.org/10.1016/B978-0-08-100508-8.00009-6

Moncecchi, G., Minel, J., & Wonsever, D. (2014). The influence of syntactic information on hedge scope detection. In A. L. C. Bazzan & K. Pichara (Eds.), *Advances in Artificial Intelligence – IBERAMIA 2014. 14th Ibero-American Conference on Artificial Intelligence, Santiago de Chile, Chile, 24–27 November* (Lecture Notes in Computer Science, Vol. 8864) (pp. 83–94). Heidelberg: Springer.

Morante, R., & Blanco, E. (2012). * SEM 2012 shared task: Resolving the scope and focus of negation. *Proceedings of the First Joint Conference on Lexical and Computational Semantics – Vol. 1: Proceedings of the Main Conference and the Shared Task, and Vol. 2: Proceedings of the Sixth International Workshop on Semantic Evaluation* (pp. 265–274).

Morante, R., & Daelemans, W. (2009a). Learning the scope of hedge cues in biomedical texts. *BioNLP '09. Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing, Boulder, CO, 4–5 June*, (pp. 28–36). Stroudsburg, PA: Association for Computational Linguistics (ACL).

Morante, R., & Daelemans, W. (2009b). A metalearning approach to processing the scope of negation. *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL)* (pp. 21–29). Stroudsburg, PA: Association for Computational Linguistics (ACL). https://doi.org/10.3115/1596374.1596381

Morante, R., & Daelemans, W. (2012). ConanDoyle-neg: Annotation of negation in conan doyle stories. *LREC 2012. Proceedings of the Eighth International Conference on Language Resources and Evaluation, Istanbul, Turkey, 21–27 May*.

Morante, R., Liekens, A., & Daelemans, W. (2008). Learning the scope of negation in biomedical texts. *EMNLP '08. Proceedings of the Conference on Empirical Methods in Natural Language Processing, Honolulu, HI, 25–27 October* (pp. 715–724). Stroudsburg, PA: Association for Computational Linguistics (ACL). https://doi.org/10.3115/1613715.1613805

Morante, R., & Sporleder, C. (2012). Modality and negation: An introduction to the special issue. *Computational Linguistics*, 38(2), 223–260. https://doi.org/10.1162/COLI_a_00095

Morante, R., Van Asch, V., & Daelemans, W. (2010). Memory-based resolution of in-sentence scopes of hedge cues. *CoNLL '10. Shared Task Proceedings of the Fourteenth Conference on Computational Natural Language Learning – Shared Task, Uppsala, Sweden, 15–16 July* (pp. 40–47). Stroudsburg, PA: Association for Computational Linguistics (ACL).

Moreno, A., López, S., Sánchez, F., & Grishman, R. (2003). Developing a syntactic annotation scheme and tools for a spanish treebank. In A. Abeillé (Ed.), *Treebanks. Building and Using Parsed Corpora* (pp. 149–163) Dordrecht: Kluwer.

Morita, M., Kano, Y., Ohkuma, T., Miyabe, M., & Aramaki, E. (2013). Overview of the NTCIR-10 MedNLP task. *Proceedings of NTCIR-10, Tokyo, Japan, 18–21 June*.

Mowery, D. L., Velupillai, S., South, B. R., Christensen, L., Martinez, D., Kelly, L., … Savova, G. (2014). Task 2: ShARe/CLEF eHealth evaluation lab 2014. *CLEF 2014 Online Working Notes, 1180*, 31–42.

Mutalik, P. G., Deshpande, A., & Nadkarni, P. M. (2001). Use of general-purpose negation detection to augment concept indexing of medical documents a quantitative study using the umls. *Journal of the American Medical Informatics Association*, 8(6), 598–609. https://doi.org/10.1136/jamia.2001.0080598

Nadkarni, P. M., Ohno-Machado, L., & Chapman, W. W. (2011). Natural language processing: An introduction. *Journal of the American Medical Informatics Association: JAMIA*, 18(5), 544–551. https://doi.org/10.1136/amiajnl-2011-000464

Nawaz, R., Thompson, P., & Ananiadou, S. (2010). Evaluating a meta-knowledge annotation scheme for bio-events. *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing (NeSp-NLP 2010)* (pp. 69–77).

Ohana, B., Tierney, B., & Delany, S. J. (2016). Sentiment classification using negation as a proxy for negative sentiment. *FLAIRS 2016. The 29th International FLAIRS Conference, Key Largo, FL, 16–18 May* (pp. 316–321).

Øvrelid, L., Velldal, E., & Oepen, S. (2010). Syntactic scope resolution in uncertainty analysis. *Proceedings of the 23rd International Conference on Computational Linguistics, COLING 2010* (pp. 1379–1387).

Özgür, A., & Radev, D. R. (2009). Detecting speculations and their scopes in scientific text. *EMNLP '11. Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, Vol. 3, Singapore, 6–7 August* (pp. 1398–1407). Stroudsburg, PA: Association for Computational Linguistics (ACL).

Packer, A. L. (1999). SCIELO-an electronic publishing model for developing countries. In J. Smith, A. Ardö & P. Linde (Eds.), *ELPUB*. Washington, DC: ICCC Press.

Packer, A. L. (1998). SciELO: Uma metodologia para publicação eletrônica. *Ciência Da Informação*, 27(2), 109–121. https://doi.org/10.1590/S0100-19651998000200002

Palmer, F. R. (2001). *Mood and Modality* Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139167178

Pang, B., & Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. *Proceedings of ACL 2004* (pp. 271–278). Stroudsburg, PA: Association for Computational Linguistics (ACL).

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1–2), 1–135. https://doi.org/10.1561/1500000011

Payne, T. E. (1997). *Describing Morphosyntax: A Guide for Field Linguists*. Cambridge: Cambridge University Press.

Piskorski, J., & Yangarber, R. (2013). Information extraction: Past, present and future. In T. Poibeau et al. (Eds.), *Multi-source, Multilingual Information Extraction and Summarization. Theory and Applications of Natural Language Processing* (pp. 23–49). Heidelberg: Springer. https://doi.org/10.1007/978-3-642-28569-1_2

Prabhakaran, V., Rambow, O., & Diab, M. (2010). Automatic committed belief tagging. *Proceedings of the 23rd International Conference on Computational Linguistics (COLING): Posters* (pp. 1014–1022).

Pröllochs, N., Feuerriegel, S., & Neumann, D. (2016). Negation scope detection in sentiment analysis: Decision support for news-driven trading. *Decision Support Systems*, 88, 67–75. https://doi.org/10.1016/j.dss.2016.05.009

Pröllochs, N., Feuerriegel, S., & Neumann, D. (2017). Understanding negations in information processing: Learning from replicating human behavior. *Arxiv.org*.

Pyysalo, S., Ginter, F., Heimonen, J., Björne, J., Boberg, J., Järvinen, J., & Salakoski, T. (2007). BioInfer: A corpus for information extraction in the biomedical domain. *BMC Bioinformatics*, 8(1), 50. https://doi.org/10.1186/1471-2105-8-50

Qian, Z., Li, P., Zhu, Q., Zhou, G., Luo, Z., & Luo, W. (2016). Speculation and negation scope detection via convolutional neural networks. *The 2016 Conference on Empirical Methods on Natural Language Processing*. Stroudsburg, PA: Association for Computational Linguistics (ACL).

Ragini, J. R., & Anand, P. R. (2016). Sentiment analysis: A comprehensive overview and the state of art research challenges. *Indian Journal of Science and Technology*, 9(47).

Rea, S., Pathak, J., Savova, G., Oniki, T. A., Westberg, L., Beebe, C. E., … Huff, S. M. (2012). Building a robust, scalable and standards-driven infrastructure for secondary use of EHR data: The SHARPn project. *Journal of Biomedical Informatics*, 45(4), 763–771. https://doi.org/10.1016/j.jbi.2012.01.009

Read, J., Velldal, E., Oepen, S., & Øvrelid, L. (2011). Resolving speculation and negation scope in biomedical articles with a syntactic constituent ranker. *Proceedings of the Fourth International Symposium on Languages in Biology and Medicine, Singapore (LBM 2011)*.

Read, J., Velldal, E., Øvrelid, L., & Oepen, S. (2012). UiO 1: Constituent-based discriminative ranking for negation resolution. *Proceedings of the First Joint Conference on Lexical and Computational Semantics, Vol. 1: Proceedings of the Main Conference and the Shared Task, and Vol. 2: Proceedings of the Sixth International Workshop on Semantic Evaluation* (pp. 310–318).

Rei, M., & Briscoe, T. (2010). Combining manual rules and supervised learning for hedge cue and scope detection. *Proceedings of the Fourteenth Conference on Computational Natural Language Learning – Shared Task* (pp. 56–63).

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

Rijsbergen, C. J. V. (1979). *Information retrieval* (2nd ed.). Newton, MA: Butterworth-Heinemann.

Rios, M., Specia, L., Gelbukh, A., & Mitkov, R. (2014). Statistical relational learning to recognise textual entailment. In A. Gelbukh (Ed.), *International Conference on Intelligent Text Processing and Computational Linguistics, CICLing 2014* (Lecture Notes in Computer Science, Vol. 8403) (pp. 330–339). Heidelberg: Springer.

Rushdi Saleh, M., Martín-Valdivia, M. T., Montejo-Ráez, A., & Ureña-López, L. (2011). Experiments with SVM to classify opinions in different domains. *Expert Systems with Applications*, 38(12), 14799–14804. https://doi.org/10.1016/j.eswa.2011.05.070

Sammons, M., Vydiswaran, V., & Roth, D. (2010). Ask not what textual entailment can do for you… *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, Uppsala, Sweden, 11–16 July* (pp. 1199–1208). Stroudsburg, PA: Association for Computational Linguistics (ACL).

Saurí, R., & Pustejovsky, J. (2009). FactBank: A corpus annotated with event factuality. *Language Resources and Evaluation*, 43(3), 227–268. https://doi.org/10.1007/s10579-009-9089-9

Savova, G. K., Masanz, J. J., Ogren, P. V., Zheng, J., Sohn, S., Kipper-Schuler, K. C., & Chute, C. G. (2010). Mayo clinical text analysis and knowledge extraction system (cTAKES): Architecture, component evaluation and applications. *Journal of the American Medical Informatics Association: JAMIA*, 17(5), 507–513. https://doi.org/10.1136/jamia.2009.001560

Segura Bedmar, I., Martínez, P., & Herrero Zazo, M. (2013). Semeval-2013 task 9: Extraction of drug-drug interactions from biomedical texts (ddiextraction 2013). *Second Joint Conference on Lexical and Computional Semantics (*SEM), Vol. 2* (pp. 341–350).

Seifert, S., & Welte, W. (1987). *A Basic Bibliography on Negation in Natural Language*. Tübingen: Gunter Narr.

Sharif, W., Samsudin, N. A., Deris, M. M., & Naseem, R. (2016). Effect of negation in sentiment analysis. *Innovative Computing Technology (INTECH), 2016 Sixth International Conference on* (pp. 718–723). https://doi.org/10.1109/INTECH.2016.7845119

Sharma, N., Sharma, R., & Biswas, K. K. (2015). Recognizing textual entailment using dependency analysis and machine learning. *HLT-NAACL 2015. The 2015 Conference of the North American Association for Computational Linguistics: Human Language Technologies. Proceedings of the Conference, Denver, CO, 31 May – 5 June* (pp. 147–153).

Shatkay, H., Pan, F., Rzhetsky, A., & Wilbur, W. J. (2008). Multi-dimensional classification of biomedical text: Toward automated, practical provision of high-utility text to diverse users. *Bioinformatics* 24(18), 2086–2093. https://doi.org/10.1093/bioinformatics/btn381; https://doi.org/10.1093/bioinformatics/btn381

Shivade, C., de Marneffe, M., Fosler-Lussier, E., & Lai, A. M. (2015). Extending NegEx with kernel methods for negation detection in clinical text. *ExProM 2015. Extra-Propositinal Aspects of Meaning in Computational Linguistics, Denver, CO, 5 June*.

Skeppstedt, M. (2011). Negation detection in swedish clinical text: An adaption of NegEx to swedish. *Journal of Biomedical Semantics*, 2(3), S3. https://doi.org/10.1186/2041-1480-2-S3-S3

Skeppstedt, M., Paradis, C., & Kerren, A. (2016). Marker words for negation and speculation in health records and consumer reviews. In M. Neves, F. Rinaldi, G. Nenadic, & D. Rebholz-Schumann (Eds.), *SMBM 2016. Semantic Mining in Biomedicine. Proceedings of the 7th International Symposium on Semantic Mining in Biomedicine, Potsdam, Germany, 4–5 August, 2016* (pp. 64–69).

Skeppstedt, M., Schamp-Bjerede, T., Sahlgren, M., Paradis, C., & Kerren, A. (2015). Detecting speculations, contrasts and conditionals in consumer reviews. In A. Balahur, E. van der Groot, P. Vossen, & A. Montoyo (Eds.), *6th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis, WASSA 2015: Workshop proceedings* (pp. 162–168). Stroudsburg, PA: Association for Computational Linguistics (ACL).

Snow, R., Vanderwende, L., & Menezes, A. (2006). Effectively using syntax for recognizing false entailment. *Proceedings of the Main Conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics, New York, NY* (pp. 33–40). Stroudsburg, PA: Association for Computational Linguistics (ACL).

Stricker, V., Iacobacci, I., & Cotik, V. (2015). Negated findings detection in radiology reports in Spanish: An adaptation of NegEx to Spanish. *IJCAI-Workshop on Replicability and Reproducibility in Natural Language Processing: Adaptative Methods, Resources and Software, Buenos Aires, Argentina, 25–27 July*.

Styler IV, W. F., Bethard, S., Finan, S., Palmer, M., Pradhan, S., de Groen, P. C., … Savova, G. (2014). Temporal annotation in the clinical domain. *Transactions of the Association for Computational Linguistics*, 2, 143–154.

Su, Q., Huang, C., & Chen, H. K. (2010). Evidentiality for text trustworthiness detection. *Proceedings of the 2010 Workshop on NLP and Linguistics: Finding the Common Ground (NLPLING 2010), Uppsala, Sweden, 16 July* (pp. 10–17). Stroudsburg, PA: Association for Computational Linguistics (ACL).

Szarvas, G. (2008). Hedge classification in biomedical texts with a weakly supervised selection of keywords. *Proceedings of 46th Meeting of the Association for Computational Linguistics* (pp. 281–289). Stroudsburg, PA: Assocation for Computational Linguistics (ACL).

Szarvas, G., & Gurevych, I. Uncertainty detection for natural language watermarking. *Proceedings of the International Joint Conference on Natural Language Processing (IJCNLP 2013), Nagoya, Japan*, 14–18 October (pp. 1188–1194).

Szarvas, G., Vincze, V., Farkas, R., & Csirik, J. (2008). The BioScope Corpus: Biomedical texts annotated for uncertainty, negation and their scopes. *BMC Bioinformatics*, 9(11), S9.

Taboada, M. (2016). Sentiment analysis: An overview from linguistics. *Annual Review of Linguistics*, 2, 325–347. https://doi.org/10.1146/annurev-linguistics-011415-040518

Taboada, M. (2008). *SFU Review Corpus*. Simon Fraser University. <http://www.sfu.ca/~mtaboada/research/SFU_Review_Corpus.html>

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational Linguistics*, 37(2), 267–307. https://doi.org/10.1162/COLI_a_00049

Taboada, M., Gillies, M. A., & McFetridge, P. (2006). Sentiment classification techniques for tracking literary reputation. *LREC Workshop: Towards Computational Models of Literary Analysis, Genoa, Italy, 22 May* (pp. 36–43).

Tang, B., Wang, X., Wang, X., Yuan, B., & Fan, S. (2010). A cascade method for detecting hedges and their scope in natural language text. *Proceedings of the Fourteenth Conference on Computational Natural Language Learning – Shared Task, Uppsala, Sweden, 15–16 July* (pp. 13–17).

Thompson, P., Ananiadou, S., & Tsujii, J. (2017). The GENIA corpus: Annotation levels and applications. In N. Ide & J. Pustejovsky (Eds.), *Handbook of Linguistic Annotation* (pp. 1395–1432) Berlin: Springer. https://doi.org/10.1007/978-94-024-0881-2_54

Tottie, G. (1991). *Negation in English Speech and Writing: A Study in Variation*. New York, NY: Academic Press.

Uzuner, O., South, B. R., Shen, S., & DuVall, S. L. (2011). 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. *Journal of the American Medical Informatics Association: JAMIA*, 18(5), 552–556. https://doi.org/10.1136/amiajnl-2011-000203

Valencia, V. M. S. (1991). *Studies on Natural Logic and Categorial Grammar*. Valencia: VM Sanchez.

Van der Wouden, T. (2002). *Negative Contexts: Collocation, Polarity and Multiple Negation*. London: Routledge. https://doi.org/10.4324/9780203437636

van Munster, E. (1988). The treatment of scope and negation in rosetta. In D. Vargha (Ed.), *COLING '88. Proceedings of the 12th Conference on Computational Linguistics, Vol. 2, Budapest, Hungary, 22–27 August*, (pp. 442–447). Stroudsburg, PA: Association for Computational Linguistics (ACL).

Velldal, E. (2011). Predicting speculation: A simple disambiguation approach to hedge detection in biomedical literature. *Journal of Biomedical Semantics*, 2(5), S7. https://doi.org/10.1186/2041-1480-2-S5-S7

Velldal, E., Øvrelid, L., & Oepen, S. (2010). Resolving speculation: MaxEnt cue classification and dependency-based scope rules. *Proceedings of the Fourteenth Conference on Computational Natural Language Learning – Shared Task, Uppsala, Sweden, 15–16 July* (pp. 48–55).

Velldal, E., Øvrelid, L., Read, J., & Oepen, S. (2012). Speculation and negation: Rules, rankers, and the role of syntax. *Computational Linguistics*, 38(2), 369–410. https://doi.org/10.1162/COLI_a_00126

Velupillai, S. (2011). Automatic classification of factuality levels: A case study on swedish diagnoses and the impact of local context. *Fourth International Symposium on Languages in Biology and Medicine, LBM 2011, Singapore, 14–15 December*.

Velupillai, S., Dalianis, H., & Kvist, M. (2011). Factuality levels of diagnoses in swedish clinical text. *MIE* (pp. 559–563).

Velupillai, S., Skeppstedt, M., Kvist, M., Mowery, D., Chapman, B. E., Dalianis, H., & Chapman, W. W. (2014). Cue-based assertion classification for swedish clinical text – Developing a lexicon for pyConTextSwe. *Artificial Intelligence in Medicine*, 61(3), 137–144. https://doi.org/10.1016/j.artmed.2014.01.001

Verbeke, M., Frasconi, P., Van Asch, V., Morante, R., Daelemans, W., & De Raedt, L. (2012). Kernel-based logical and relational learning with kLog for hedge cue detection. In S. H. Muggleton, A. Tamaddoni-Nezhad, F. A. Lisi (Eds.), *Inductive Logic Programming* (Lecture Notes in Computer Science, Vol. 7207) (pp. 347–357). Heidelberg: Springer. https://doi.org/10.1007/978-3-642-31951-8_29

Vincze, V. (2014). Uncertainty detection in hungarian texts. *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, Dublin, Ireland, August 23–29* (pp. 1844–1853).

Vincze, V. (2015). Uncertainty Detection in Natural Language Texts (Unpublished PhD dissertation). Unversity of Szeged.

Vincze, V. (2016). Detecting uncertainty cues in Hungarian social media texts. *ExProM 2016. Extra-Propositional Aspects of Meaning in Computational Linguistics. Proceedings of the Workshop, Osaka, Japan, 12 December* (pp. 11–21).

Vincze, V., Simkó, K. I., & Varga, V. (2014). Annotating uncertainty in hungarian webtext. *The 8th Linguistic Annotation Workshop in conjunction with COLING 2014. Proceedings of the Workshop, 23–24 August, Dublin, Ireland* (pp. 64–69).

Vlachos, A., & Craven, M. (2010). Detecting speculative language using syntactic dependencies and logistic regression. *Proceedings of the Fourteenth Conference on Computational Natural Language Learning – Shared Task, Uppsala, Sweden, 15–16 July* (pp. 18–25).

Von Fintel, K. (2006). Modality and language. In D. M. Borchert (Ed.), *Encyclopedia of Philosophy* (2nd ed.). Detroit, MI: MacMillan Reference USA. Most recent version online at <http://mit.edu/fintel/www/modality.pdf>.

Voorhees, E. M., & Hersh, W. R. (2012). Overview of the TREC 2012 medical records track. *The Twenty-First Text REtrieval Conference (TREC 2012) Proceedings* (NIST Special Publication: SP 500–298.) <http://trec.nist.gov/pubs/trec21/t21.proceedings.html>

Wei, Z., Chen, J., Gao, W., Li, B., Zhou, L., He, Y., & Wong, K. (2013). An empirical study on uncertainty identification in social media context. *ACL (2013)* (pp. 58–62). Stroudsburg, PA: Assocation for Computitational Linguistics (ACL).

Wetzel, D., & Bond, F. (2012). Enriching parallel corpora for statistical machine translation with semantic negation rephrasing. In M. Carpuat & L. Specia (Eds.), *SSST-6 '12. Proceedings of the Sixth Workshop on Syntax, Semantics and Structure in Statistical Translation, Jeju, Republic of Korea, 12 July* (pp. 20–29). Stroudsburg, PA: Association for Computational Linguistics (ACL).

White, J. P. (2012). UWashington: Negation resolution using machine learning methods. *Proceedings of the First Joint Conference on Lexical and Computational Semantics, Vol. 1: Proceedings of the Main Conference and the Shared Task, and Vol. 2: Proceedings of the Sixth International Workshop on Semantic Evaluation* (pp. 335–339).

Wiegand, M., Balahur, A., Roth, B., Klakow, D., & Montoyo, A. (2010). A survey on the role of negation in sentiment analysis. *Negation and Speculation in Natural Language Processing (NeSp-NLP 2010), Proceedings of the Workshop, Uppsala, Sweden, 10 July* (pp. 60–68). Stroudsburg, PA: Association for Computational Linguistics (ACL).

Wilson, T., Wiebe, J., & Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. *Proceedings of HLT-EMNLP 2005*.

Wu, S., Miller, T., Masanz, J., Coarr, M., Halgrim, S., Carrell, D., & Clark, C. (2014). Negation's not solved: Generalizability versus optimizability in clinical natural language processing. *PloS One*, 9(11), e112774. https://doi.org/10.1371/journal.pone.0112774

Zhang, S., Kang, T., Zhang, X., Wen, D., Elhadad, N., & Lei, J. (2016). Speculation detection for chinese clinical notes: Impacts of word segmentation and embedding models. *Journal of Biomedical Informatics*, 60, 334–341. https://doi.org/10.1016/j.jbi.2016.02.011

Zhang, Y., Jiang, K., Dong, J., Qian, D., Wu, H., Geng, X., & Wang, L. (2014). Negation detection in chinese electronic medical record based on rules and word co-occurrence. In S. Li, Q. Jin, X. Jiang, J. H. Park (Eds.), *Frontier and Future Development of Information Technology in Medicine and Education, ITME 2013* (Lecture Notes in Electrical Enginering 269) (pp. 2215–2220). Heidelberg: Springer. https://doi.org/10.1007/978-94-007-7618-0_270

Zhao, J., Zhu, T., & Lan, M. (2014). ECNU: One stone two birds: Ensemble of heterogenous measures for semantic relatedness and textual entailment. *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), Dublin, Ireland, 23–24 August* (pp. 271–277). https://doi.org/10.3115/v1/S14-2044

Zhou, H., Deng, H., Huang, D., & Zhu, M. (2015). Hedge scope detection in biomedical texts: An effective dependency-based method. *PloS One*, 10(7), e0133715. https://doi.org/10.1371/journal.pone.0133715

Zhou, H., Xu, J., Yang, Y., Deng, H., Chen, L., & Huang, D. (2016). Chinese hedge scope detection based on structure and semantic information. In M. Sun, X. Huang, H. Lin, Z. Liu & Y. Liu (Eds.), *Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data* (Lecture Notes in Computer Science, Vol 10035), (pp. 204–215). Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-47674-2_18

Zhu, Q., Li, J., Wang, H., & Zhou, G. (2010). A unified framework for scope learning via simplified shallow semantic parsing. *EMNLP 2010. Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, Cambridge, MA, 9–11 October* (pp. 714–724).

Zou, B., Zhou, G., & Zhu, Q. (2013). Tree kernel-based negation and speculation scope detection with structured syntactic parse features. *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, Seattle, Washington, 18–21 October* (pp. 968–976).

Zou, B., Zhou, G., & Zhu, Q. (2016). Research on chinese negation and speculation: Corpus annotation and identification. *Frontiers of Computer Science*, 10(6), 1039–1051. https://doi.org/10.1007/s11704-015-5101-2

Zou, B., Zhu, Q., & Zhou, G. (2015). Negation and speculation identification in chinese language. *ACL 2015* (pp. 656–665). Stroudsburg, PA: Association for Computational Linguistics (ACL).

Zuck, J. G., & Zuck, L. V. (1986). Hedging in newswriting. In A. M. Cornu, J. van Parijs, M. Delahaye, & L. Baten (Eds.), *Beads or Bracelets. How do we Approach LSP* (pp. 172–180). Oxford: Oxford University Press.

Subject index

Negation and speculation detection is an emerging topic that has attracted the attention of many researchers, and there is clearly a lack of relevant textbooks and survey texts. This book aims to define negation and speculation from a natural language processing perspective, to explain the need for processing these phenomena, to summarise existing research on processing negation and speculation, to provide a list of resources and tools, and to speculate about future developments in this research area. An advantage of this book is that it will not only provide an overview of the state of the art in negation and speculation detection, but will also introduce newly developed data sets and scripts. It will be useful for students of natural language processing subjects who are interested in understanding this task in more depth and for researchers with an interest in these phenomena in order to improve performance in other natural language processing tasks.

ISBN 978 90 272 0217 8

9 789027 202178

JOHN BENJAMINS PUBLISHING COMPANY