Modality and Negation: An Introduction to the Special Issue

Roser Morante[*]
University of Antwerp

Caroline Sporleder[**]
Saarland University

Traditionally, most research in NLP has focused on propositional aspects of meaning. To truly understand language, however, extra-propositional aspects are equally important. Modality and negation typically contribute significantly to these extra-propositional meaning aspects. Although modality and negation have often been neglected by mainstream computational linguistics, interest has grown in recent years, as evidenced by several annotation projects dedicated to these phenomena. Researchers have started to work on modeling factuality, belief and certainty, detecting speculative sentences and hedging, identifying contradictions, and determining the scope of expressions of modality and negation. In this article, we will provide an overview of how modality and negation have been modeled in computational linguistics.

1. Introduction

Modality and **negation** are two grammatical phenomena that have been studied for a long time. Aristotle was the initial main contributor to the analysis of negation from a philosophical perspective. Since then, thousands of studies have been performed, as illustrated by the *Basic Bibliography of Negation in Natural Language* (Seifert and Welte 1987). One of the first categorizations of modality is proposed by Otto Jespersen (1924) in the chapter about *Mood*, where the grammarian distinguishes between "categories containing an element of will" and categories "containing no element of will." His grammar devotes also a chapter to negation.

   In contrast to the substantial number of theoretical studies, the computational treatment of modality and negation is a newly emerging area of research. The emergence of this area is a natural consequence of the consolidation of areas that focus on the computational treatment of propositional aspects of meaning, like semantic role labeling, and a response to the need for processing extra-propositional aspects of meaning as a further step towards text understanding. That there is more to meaning than just propositional content is a long-held view. Prabhakaran, Rambow, and Diab (2010)

---

* CLiPS, University of Antwerp, Prinsstraat 13, B-2000 Antwerpen, Belgium.
  E-mail: `roser.morante@ua.ac.be`.
** Computational Linguistics, Saarland University, Postfach 15 11 50, D-66041 Saarbrücken, Germany.
  E-mail: `csporled@coli.uni-sb.de`.

illustrate this statement with the following examples, where the event LAY_OFF(GM, workers) is presented with different extra-propositional meanings:

(1) a. GM will lay off workers.

   b. A spokesman for GM said GM will lay off workers.

   c. GM may lay off workers.

   d. The politician claimed that GM will lay off workers.

   e. Some wish GM would lay of workers.

   f. Will GM lay off workers?

   g. Many wonder whether GM will lay off workers.

Generally speaking, modality is a grammatical category that allows the expression of aspects related to the attitude of the speaker towards her statements in terms of degree of certainty, reliability, subjectivity, sources of information, and perspective. We understand modality in a broad sense, which involves related concepts like "subjectivity", "hedging", "evidentiality", "uncertainty", "committed belief," and "factuality". Negation is a grammatical category that allows the changing of the truth value of a proposition. A more detailed definition of these concepts with examples will be presented in Sections 2 and 3.

Modality and negation are challenging phenomena not only from a theoretical perspective, but also from a computational point of view. So far two main tasks have been addressed in the computational linguistics community: (i) the detection of various forms of negation and modality and (ii) the resolution of the scope of modality and negation cues. Whereas modality and negation tend to be lexically marked, the class of markers is heterogeneous, especially in the case of modality. Determining whether a sentence is speculative or whether it contains negated concepts cannot be achieved by simple lexical look-up of words potentially indicating modality or negation. Modal verbs like *might* are prototypical modality markers, but they can be used in multiple senses. Multiword expressions can also express modality (e.g., *this brings us to the largest of all mysteries* or *little was known*). Modality and negation interact with mood and tense markers, and also with each other. Finally, discourse factors also add to the complexity of these phenomena.

Incorporating information about modality and negation has been shown to be useful for a number of applications such as recognizing textual entailment (de Marneffe et al. 2006; Snow, Vanderwende, and Menezes 2006; Hickl and Bensley 2007), machine translation (Baker et al. 2010), trustworthiness detection (Su, Huang, and Chen 2010), classification of citations (Di Marco, Kroon, and Mercer 2006), clinical and biomedical text processing (Friedman et al. 1994; Szarvas 2008), and identification of text structure (Grabar and Hamon 2009).

This overview is organized as follows: Sections 2 and 3 define modality and negation, respectively. Section 4 gives details of linguistic resources annotated with various aspects of negation and modality. We also discuss properties of the different annotation schemes that have been proposed. Having discussed the linguistic basis as well as the available resources, the remainder of the article then provides an overview of automated methods for dealing with modality and negation. Most of the work in this area has been carried out at the sentence or predicate level. Section 5 discusses various methods for detecting speculative sentences. This is only a first step, however. For a more fine-

grained analysis, it is necessary to deal with modality and negation on a sub-sentential (i.e., predicate) level.

This is addressed in Section 6, which also discusses various methods for the important task of scope detection. Section 7 then moves on to work on detecting negation and modality at a discourse level, that is, in the context of recognizing contrasts and contradictions. Section 8 takes a closer look at dealing with positive and negative opinions and summarizes studies in the field of sentiment analysis that have explicitly modeled modality and negation. Section 9 provides an overview of the articles in this special issue. Finally, Section 10 concludes this article by outlining some of the remaining challenges.

Some notational conventions should be clarified. In the literature, the affixes, words or multiword expressions that express modality and negation have been referred to as triggers, signals, markers, and cues. Here, we will refer to them as **cues** and we will mark them in bold in the examples. The boundaries of their scope will be marked with square brackets.

2. Modality

From a theoretical perspective, modality can be defined as a philosophical concept, as a subject of the study of logic, or as a grammatical category. There are many definitions and classifications of modal phenomena. Even if we compiled an exhaustive and precise set of existing definitions, we would still be providing a limited view on what modality is, because, as Salkie, Busuttil, and van der Auwera (2009, page 7) put it:

> ...modality is a big intrigue. Questions erstwhile considered solved become open questions again. New observations and hypotheses come to light, not least because the subject matter is changing.

Defining modality from a computational linguistics perspective for this special issue becomes even more difficult because several concepts are used to refer to phenomena that are related to modality, depending on the task at hand and the specific phenomena that the authors address. To mention some examples, research focuses on categorizing modality, on committed belief tagging, on resolving the scope of hedge cues, on detecting speculative language, and on computing factuality. These concepts are related to the attitude of the speaker towards her statements in terms of degree of certainty, reliability, subjectivity, sources of information, and perspective. Because this special issue focuses on the computational treatment of modality, we will provide a general theoretical description of modality and the related concepts mentioned in the computational linguistics literature at the cost of offering a simplified view of these concepts.

Jespersen (1924, page 329) attempts to place all moods in a logically consistent system, distinguishing between "categories containing an element of will" and "categories containing no element of will"—later named as **propositional modality** and **event modality** by Palmer (1986). Lyons (1977, page 793) describes **epistemic modality** as concerned with matters of knowledge and belief, "the speaker's opinion or attitude towards the proposition that the sentence expresses or the situation that the proposition describes." Palmer (1986, page 8) distinguishes propositional modality, which is "concerned with the speaker's attitude to the truth-value or factual status of the proposition"

as in Example (2a), and event modality, which "refers to events that are not actualized, events that have not taken place but are merely potential" as in Example (2b):

(2) a. Kate must be at home now.

b. Kate must come in now.

Within propositional modality, Palmer defines two types: **epistemic**, used by speakers "to express their judgement about the factual status of the proposition," and **evidential**, used "to indicate the evidence that they have for its factual status" (Palmer 1986, 8–9). He also defines two types of event modality: **deontic**, which relates to obligation or permission and to conditional factors "that are external to the relevant individual," and **dynamic**, where the factors are internal to the individual (Palmer 1986, pages 9–13). Additionally, Palmer indicates other categories that may be marked as **irrealis** and may be found in the mood system: **future**, **negative**, **interrogative**, **imperative-jussive**, **presupposed**, **conditional**, **purposive**, **resultative**, **wishes**, and **fears**. Palmer explains how modality relates to tense and aspect: The three categories are concerned with the event reported by the utterance, whereas tense is concerned with the time of the event and aspect is "concerned with the nature of the event in terms of its internal temporal constituency" (Palmer 1986, pages 13–16).

From a philosophical standpoint, von Fintel (2006) defines modality as "a category of linguistic meaning having to do with the expression of possibility and necessity." In this sense "a modalized sentence locates an underlying or prejacent proposition in the space of possibilities." Von Fintel describes several types of modal meaning (**alethic**, **epistemic**, **deontic**, **bouletic**, **circumstantial**, and **teleological**), some of which are introduced by von Wright (1951), and shows that modal meaning can be expressed by means of several types of expressions, such as modal auxiliaries, semimodal verbs, adverbs, nouns, adjectives, and conditionals.

Within the modal logic framework several authors provide a more technical approach to modality. Modal logic (von Wright 1951; Kripke 1963) attempts to represent formally the reasoning involved in expressions of the type *it is necessary that ...* and *it is possible that ...* starting from a weak logic called K (Garson 2009). Taken in a broader sense, modal logic also aims at providing an analysis for expressions of deontic, temporal, and doxastic logic. Within the modal logic framework, modality is analyzed in terms of **possible worlds semantics** (Kratzer 1981). The initial idea is that modal expressions are considered to express quantification over possible worlds.

Kratzer (1981, 1991), however, argues that modal expressions are more complex than quantifiers and that their meaning is context-dependent. Recent work on modality in the framework of modal logic is presented by Portner (2009, pages 2–8), who groups modal forms into three categories: **sentential modality** ("the expression of modal meaning at the level of the whole sentence"); **sub-sentential modality** ("the expression of modal meaning within constituents smaller than a full clause"); and **discourse modality** ("any contribution to meaning in discourse which cannot be accounted for in terms of a traditional semantic framework").

From a typological perspective, the study of modality seeks to describe how the languages of the world express different types of modality (Palmer 1986; van der Auwera and Plungian 1998). Knowing how modality is expressed across languages is relevant for the computational linguistics community, not only because it is essential for developing automated systems for languages other than English, but also because

it throws some light on the underlying phenomena that might be beneficial for the development of novel methods for dealing with modality.

Concepts related to modality that have been studied in computational linguistics are: hedging, evidentiality, uncertainty, factuality, and subjectivity. The term **hedging** is originally due to Lakoff (1972, page 195), who describes hedges as "words whose job is to make things more or less fuzzy." Lakoff starts from the observation that "natural language concepts have vague boundaries and fuzzy edges and that, consequently, natural language sentences will very often be neither true, nor false, nor nonsensical, but rather true to a certain extent and false to a certain extent, true in certain aspects and false in certain aspects" (Lakoff 1972, page 183). In order to deal with this aspect of language, he extends the classical propositional and predicate logic to fuzzy logic and focuses on the study of hedges. Hyland (1998) studies hedging in scientific texts. He proposes a pragmatic classification of hedge expressions based on an exhaustive analysis of a corpus. The catalogue of hedging cues includes modal auxiliaries, epistemic lexical verbs, epistemic adjectives, adverbs, nouns, and a variety of non-lexical cues.

Evidentiality is related to the expression of the information source of a statement. As Aikhenvald (2004, page 1) puts it:

> In about a quarter of the world's languages, every statement must specify the type of source on which it is based [...]. This grammatical category, whose primary meaning is information source, is called 'evidentiality'.

This grammatical category was already introduced by Boas (1938), and has been studied afterwards, although less than modality. There is no agreement on whether it should be a subcategory of modality (Palmer 1986; de Haan 1995) or a category by itself (de Haan 1999; Aikhenvald 2004). A broader definition relates evidentiality to the expression of the speaker's attitude towards the information being presented (Chafe 1986). Ifantidou (2001, page 5) considers that the function of evidentials is to indicate the source of knowledge (observation, hearsay, inference, memory) on which a statement is based and the speaker's degree of certainty about the proposition expressed.

Certainty is a type of subjective information that can be conceived of as a variety of epistemic modality (Rubin, Liddy, and Kando 2005). Here we take their definition (page 65):

> . . . certainty is viewed as a type of subjective information available in texts and a form of epistemic modality expressed through explicitly-coded linguistic means. Such devices [...] explicitly signal presence of certainty information that covers a full continuum of writer's confidence, ranging from uncertain possibility and withholding full commitment to statements.

Factuality involves polarity, epistemic modality, evidentiality, and mood. It is defined by Saurí (2008, page 1) as:

> . . . the level of information expressing the commitment of relevant sources towards the factual nature of eventualities in text. That is, it is in charge of conveying whether eventualities are characterized as corresponding to a fact, to a possibility, or to a situation that does not hold in the world.

Factuality can be expressed by several linguistic means: negative polarity particles, modality particles, event-selecting predicates which project factuality information on the events denoted by their arguments (*claim*, *suggest*, *promise*, etc.), and syntactic

constructions involving subordination. The factuality of a specific event can change during the unfolding of the text. As described in Saurí and Pustejovsky (2009), depending on the polarity, events are depicted as either facts or counterfacts. Depending on the level of uncertainty combined with polarity, events will be presented as possibly factual (3a) or possibly counterfactual (3b).

(3) a. The United States may extend its naval quarantine to Jordan's Red Sea port of Aqaba.

b. They may not have enthused him for their particular brand of political idealism.

The term **subjectivity** is introduced by Banfield (1982). Work on subjectivity in computational linguistics is initially due to Wiebe, Wilson, and collaborators (Wiebe 1994; Wiebe et al. 2004; Wiebe, Wilson, and Cardie 2005; Wilson 2008; Wilson et al. 2005; Wilson, Wiebe, and Hwa 2006) and focuses on learning subjectivity from corpora. As Wiebe et al. (2004, page 279) put it:

> Subjective language is language used to express private states in the context of a text or conversation. *Private state* is a general covering term for opinions, evaluations, emotions, and speculations.

Subjectivity is expressed by means of linguistic expressions of various types from words to syntactic devices that are called **subjective elements**. Subjective statements are presented from the point of view of someone, who is called the **source**. As Wiebe et al. (2004) highlight, subjective does not mean not true. For example, in Example (4a), *criticized* expresses subjectivity, but the events CRITICIZE and SMOKE are presented as being true. Not all events contained in subjective statements need to be true, however. Modal expressions can be used to express subjective language, as in Example (4b), where the modal cue *perhaps* combined with the future tense is used to present the event FORGIVE as non-factual.

(4) a. John criticized Mary for smoking.

b. Perhaps you'll forgive me for reposting his response.

Modality and evidentiality are grammatical categories, whereas certainty, hedging, and subjectivity are pragmatic positions, and event factuality is a level of information. In this special issue we will use the term *modality* in a broad sense, similar to the **extended modality** of Matsuyoshi et al. (2010), which they use to refer to "modality, polarity, and other associated information of an event mention." Subjectivity in the general sense and opinion are beyond the scope of this special issue, however, because research in these areas focuses on different topics and already has a well defined framework of reference.

Modality-related phenomena are not rare. According to Light, Qiu, and Srinivasan (2004), 11% of sentences in MEDLINE contain speculative language. Vincze et al. (2008) report that around 18% of sentences occurring in biomedical abstracts are speculative. Nawaz, Thompson, and Ananiadou (2010) find that around 20% of the events in a biomedical corpus belong to speculative sentences and that 7% of the events are expressed with some degree of speculation. Szarvas (2008) notes that a significant proportion of the gene names mentioned in a corpus of biomedical articles appear in

speculative sentence (638 occurences out of a total of 1,968). This means that approximately 1 in every 3 genes should be excluded from the interaction detection process. Rubin (2006) reports that 59% of the sentences in a corpus of 80 articles from *The New York Times* were identified as epistemically modalized.

3. Negation

Negation is a complex phenomenon that has been studied from many perspectives, including cognition, philosophy, and linguistics. As described by Lawler (2010, page 554), cognitively, negation "involves some comparison between a 'real' situation lacking some particular element and an 'imaginal' situation that does not lack it." In the logic formalisms, "negation is the only significant monadic functor," whose behavior is described by the Law of Contradiction that asserts that no proposition can be both true and not true. In natural language, negation functions as an operator, like quantifiers and modals. A main characteristic of operators is that they have a **scope**, which means that their meaning affects other elements in the text. The affected elements can be located in the same clause (5a) or in a previous clause (5b).

(5) a. We did**n't** find the book.

b. We thought we would find the book. This was **not** the case.

The study of negation in philosophy started with Aristotle, but nowadays is still a topic that generates a considerable number of publications in the field of philosophy, logic, psycholinguistics, and linguistics. Horn (1989) provides an extensive description of negation from a historic perspective and an analysis of negation in relation to semantic and pragmatic phenomena. Tottie (1991) studies negation as a grammatical category from a descriptive and quantitative point of view, based on the analysis of empirical material. She defines two main types of negation in natural language: **rejections** of suggestions and **denials** of assertions. Denials can be **explicit** and **implicit**.

Languages have devices for negating entire propositions (**clausal negation**) or constituents of clauses (**constituent negation**). Most languages have several grammatical devices to express clausal negation, which are used with different purposes like negating existence, negating facts, or negating different aspects, modes, or speech acts (Payne 1997). As described by Payne (page 282):

> . . . a negative clause is one that asserts that some event, situation, or state of affairs does *not* hold. Negative clauses usually occur in the context of some presupposition, functioning to negate or counter-assert that presupposition.

van der Wouden (1997) defines what a negative context is, showing that negation can be expressed by a variety of grammatical categories. We reproduce some of his examples in Example (6).

(6) a. Verbs: We want to **avoid** doing any look-up, if possible.

b. Nouns: The positive degree is expressed by the **absence** of any phonic sequence.

c. Adjectives: It is **pointless** to teach any of the vernacular languages as a subject in schools.

d. Adverbs: I've **never** come across anyone quite as brainwashed as your student.

e. Prepositions: You can exchange **without** any problem.

f. Determiners: This fact has **no** direct implications for any of the two methods of font representation.

g. Pronouns: **Nobody** walks anywhere in Tucson.

h. Complementizers: Leave the door ajar, **lest** any latecomers should find themselves shut out.

i. Conjunctions: But neither this article **nor** any other similar review I have seen then had the methodological discipline to take the opposite point of view.

Negation can also be expressed by affixes, as in *motionless* or *unhappy*, and by changing the intonation or facial expression, and it can occur in a variety of syntactic constructions.

Typical negation problems that persist in the study of negation are determining the scope when negation occurs with quantifiers (7a), neg-raising (7b), the use of polarity items (7c) (*any*, *the faintest idea*), double or multiple negation (7d), and affixal negation (Tottie 1991).

(7) a. All the boys didn't leave.

b. I don't think he is coming.

c. I didn't see anything.

d. I don't know nothing no more.

Like modality, negation is a frequent phenomenon in texts. Tottie reports that negation is twice as frequent in spoken text (27.6 per 1,000 words) as in written text (12.8 per 1,000 words). Elkin et al. (2005) find that 1,823 out of 14,792 concepts in 41 Health Records from Johns Hopkins University are identified as negated by annotators. Nawaz, Thompson, and Ananiadou (2010) report that more than 3% of the biomedical events in 70 abstracts of the GENIA corpus are negated. Councill, McDonald, and Velikovich (2010) annotate a corpus of product reviews with negation information and they find that 19% of the sentences contain negations (216 out of 1,135).

3.1 Negation versus Negative Polarity

Negation and negative polarity are interrelated concepts, but it is important to notice that they are different. Negation has been defined as a grammatical phenomenon used to state that some event, situation, or state of affairs does not hold, whereas **polarity** is a relation between semantic opposites. As Israel (2004, page 701) puts it, "as such polarity encompasses not just the logical relation between negative and affirmative propositions, but also the conceptual relations defining contrary pairs like *hot–cold*, *long–short*, and *good–bad*." Israel defines three types of polar oppositions: **contradiction**, a relation in which one term must be true and the other false; **contrariety**, a relation in which only one term may be true, although both can be false; and **reversal**, which

involves an opposition between scales (⟨necessary, likely, possible⟩ ⟨impossible, unlikely, uncertain⟩.). The relation between negation and polarity lies in the fact that negation can reverse the polarity of an expression.

In this context, **negative polarity items** (NPIs) "are expressions with a limited distribution, part of which includes negative sentences" (Hoeksema 2000, page 115), like *any* in Example (8a) or *ever* in Example (8b). Lawler (2010, page 554) defines NPI as "a term applied to lexical items, fixed phrases, or syntactic construction types that demonstrate unusual behavior around negation." NPIs felicitously occur only in the scope of some negative element, such as *didn't* in Example (8b). If this element is removed, the sentence becomes agrammatical, as shown in Example (8c). The presence of an NPI in a context does not guarantee that something is being negated, however, because NPIs can also occur in certain grammatical circumstances, like interrogatives as in Example (8d).

(8) a. I didn't read any book.

b. He didn't ever read the book.

c. * He ever read the book.

d. Do you think I could ever read this book?

Polarity is a discrete category that can take two values: positive and negative. Determining the polarity of words, and phrases is a central task in sentiment analysis, in particular, disambiguating the contextual polarity of words (Wilson, Wiebe, and Hoffman 2009). Thus, in the context of sentiment analysis positive and negative polarity refers to positive and negative opinions, emotions, and evaluations.

Negation is a topic of study in sentiment analysis because it is what Wilson, Wiebe, and Hoffman (2009, page 402) call a **polarity influencer**, an element that can change the polarity of an expression. As they put it, however, "many things besides negation can influence contextual polarity, and even negation is not always straightforward." We discuss different ways of modeling negation in sentiment analysis in Section 8. The study of negative polarity is beyond the scope of this special issue, however.

4. Categorizing and Annotating Modality and Negation

Over the last few years, several corpora of texts from various domains have been annotated at different levels (expression, event, relation, sentence) with information related to modality and negation. Compared to other phenomena like semantic argument structure, dialogue acts, or discourse relations, however, no comprehensive annotation standard has been defined for modality and negation. In this section, we describe the categorization schemes that have been proposed and the corpora that have been annotated.

In the framework of the OntoSem project (Nirenburg and Raskin 2004) a corpus has been annotated with modality categories and an analyzer has been developed that takes as input unrestricted raw text and carries out several levels of linguistic analysis, including modality at the semantic level (Nirenburg and McShane 2008). The output of the semantic analysis is represented as formal text-meaning representations. Modality information is encoded as part of the semantic module in the lexical entries of the modality cues. Four modality attributes are encoded: MODALITY TYPE,

VALUE, SCOPE, and ATTRIBUTED-TO. The MODALITY TYPES are: `polarity`, whether a proposition is positive or negated; `volition`, the extent to which someone wants or does not want the event/state to occur; `obligation`, the extent to which someone considers the event/state to be necessary; `belief`, the extent to which someone believes the content of the proposition; `potential`, the extent to which someone believes that the event/state is possible; `permission`, the extent to which someone believes that the event/state is permitted; and `evaluative`, the extent to which someone believes the event/state is a good thing. The SCALAR VALUE ranges from zero to one. The SCOPE attribute is the predicate that is affected by the modality and the ATTRIBUTED-TO attribute indicates to whom the modality is assigned, the default value being the speaker. In Example (9), *should* is identified as a modality cue and characterized with the type `obligative`, value 0.8, scope *camouflage*, and is attributed to the speaker.

(9) Entrance to the tower should be totally camouflaged

The publicly available MPQA Opinion Corpus[1] (Wiebe, Wilson, and Cardie 2005) contains 10,657 sentences in 535 documents of English newswire annotated with information about **private states** at the word and phrase level. For every expression of private state a **private state frame** is defined indicating the SOURCE of the private state, whose private state is being expressed; the TARGET, what the private state is about; and properties like INTENSITY, SIGNIFICANCE, and TYPE OF ATTITUDE. Three types of private state expressions are considered for the annotation: **explicit mentions** like *fears* in Example (10a), **speech events** like *said* in Example (10b), and **expressive subjective elements**, like *full of absurdities* in Example (10b). Apart from representing private states in private state frames, Wiebe, Wilson, and Cardie also define objective speech event frames that represent "material that is attributed to some source, but is presented as an objective fact" (page 171). Having two types of frames allows a distinction between opinion-oriented material (10a, 10b) and factual material (10c).

(10) a. "The U.S. *fears* a spill-over," said Xirao-Nima.

     b. "The report is *full of absurdities*," Xirao-Nima *said*.

     c. Sergeant O'Leary said the incident took place at 2:00 pm.

Rubin, Liddy, and Kando (2005) define a model for categorizing certainty. The model distinguishes four dimensions: LEVEL, which encodes the degree of certainty; PERSPECTIVE, which encodes whose certainty is involved; FOCUS, the object of certainty; and TIME, which encodes at what time the certainty is expressed. Each dimension is further subdivided into categories, resulting in 72 possible dimension–category combinations. The four certainty LEVELS are `absolute` (Example (11a)), `high` (Example (11b)), `moderate` (Example (11c)), and `low` (Example (11d)). PERSPECTIVE separates the writer's point of view and the reported point of view. FOCUS is divided into `abstract` and `factual` information. TIME can be `past`, `present`, or `future`. The model is used to annotate certainty markers in 32 articles from *The New York Times* along these dimensions. Rubin

---

1 The MPQA corpus is available from `http://www.cs.pitt.edu/mpqa/mpqa_corpus.html`. Last accessed on 8 December 2011.

et al. find that editorials have a higher frequency of modality markers per sentence than news stories.

(11) a. An enduring lesson of the Reagan years, **of course**, is that **it really does take smoke and mirrors** to produce tax cuts, spending initiatives and a balanced budget at the same time.

b. ... but **clearly** an opportunity is at hand for the rest of the world to pressure both sides to devise a lasting peace based on democratic values and respect for human rights.

c. That fear **now seems** exaggerated, but it was not entirely fanciful.

d. So far the presidential candidates are more interested in talking about what a surplus **might buy** than in the painful choices that lie ahead.

The model is adapted in Rubin (2006, 2007) by adding a category `uncertainty` for certainty LEVEL, changing the FOCUS categories into `facts and events` and `opinions`, `emotions`, or `judgements`, and adding the `irrelevant` category for TIME. Inter-annotator agreement measures are reported for 20 articles of the 80 annotated articles randomly selected from *The New York Times* (Rubin 2006). For the task of deciding whether a statement was modalized by an explicit certainty marker or not, an agreement of 0.33 κ_{cohen} is reached. The agreement measures per dimension were 0.15 for level, 0.13 for focus, 0.44 for perspective, and 0.41 for time.

The Automatic Content Extraction 2008 corpus (Linguistic Data Consortium 2008) for relation detection and recognition collects English and Arabic texts from a variety of resources including radio and TV broadcast news, talk shows, newswire articles, Internet news groups, Web logs, and conversational telephone speech. Relations are ordered pairs of entities and are annotated with modality and tense attributes. The two modality attributes are `asserted` and `other`. `Asserted` relations pertain to situations in the real world, whereas `other` relations pertain to situations in "some other world defined by counterfactual constraints elsewhere in the context." If the entities constituting the arguments of a relation are hypothetical, then the relation can still be understood as asserted. In Example (12), the ORG-Aff.Membership relation between terrorists and Al-Qaeda is annotated as `asserted` and the Physical.Located relation between terrorists and Baghdad is annotated as `other`. The attributes for TENSE are `past`, `future`, `present`, and `unspecified`.

(12) We are afraid Al-Qaeda terrorists will be in Baghdad.

The Penn Discourse TreeBank (Prasad et al. 2008) is a corpus annotated with information related to discourse structure. Discourse connectives are considered to be the anchors of discourse relations and to act as predicates taking two **abstract objects**. Abstract objects can be assertions, beliefs, facts, or eventualities. Discourse connectives and their arguments are assigned attribution-related features (Prasad et al. 2006) such as SOURCE (`writer`, `other`, `arbitrary`), TYPE (reflecting the nature of the relation between the agent and the abstract object), SCOPAL POLARITY of attribution, and DETERMINACY (indicating the presence of contexts canceling the entailment of attribution). The text spans signaling the attribution are also marked. Prasad et al. (2006) report that 34% of the discourse relations have some non-writer agent. SCOPAL POLARITY is annotated to identify cases when verbs of attribution (*say, think*, ...) are negated syntactically

(*didn't say*) or lexically (*denied*). An argument of a connective is marked `Neg` for SCOPAL POLARITY when the interpretation of the connective requires the surface negation to take semantic scope over the lower argument. As stated by Prasad et al., in Example (13), the *but* clause entails an interpretation such as "I think it's not a main consideration," for which the negation must take narrow scope over the embedded clause rather than the higher clause.

> (13) "Having the dividend increases is a supportive element in the market outlook, but I don't think it's a main consideration," he says.

TimeML (Pustejovsky et al. 2005) is a specification language for events and temporal expressions in natural language that has been applied to the annotation of corpora like TimeBank (Pustejovsky et al. 2006). As described in Saurí, Verhagen, and Pustejovsky (2006), TimeML encodes different types of modality at the lexical and syntactic level with different tags. At the lexical level, **Situation Selecting Predicates** (SSPs) are encoded by means of the attribute CLASS within the *EVENT* tag, which allows to encode the difference between SSPs that are actions (Example (14a)[2]) and SSPs that are states (Example (14b)). SSPs of perception (Example (14c)) and reporting (Example (14d)) are encoded with more specific values due to their role in providing evidentiality. Information about modal auxiliaries and negative polarity, which are also lexically expressed, is encoded in the attributes MODALITY and POLARITY. Modality at the syntactic level is encoded as an attribute of the tag *SLINK* (Subordination Link), which can have several values: `factive`, `counterfactive`, `evidential`, `negative evidential`, `modal`, and `conditional`.

> (14) a. Companies such as Microsoft or a combined worldcom MCI are **trying** to <u>monopolize</u> Internet access.
>
> b. Analysts also **suspect** suppliers have <u>fallen</u> victim to their own success.
>
> c. Some neighbors told Birmingham police they **saw** a man <u>running</u>.
>
> d. No <u>injuries</u> were **reported** over the weekend.

FactBank (Saurí and Pustejovsky 2009) is a corpus of events annotated with factuality information, which adds to the TimeBank corpus an additional level of semantic information. Events are annotated with a discrete set of factuality values using a battery of criteria that allow annotators to differentiate among these values. It consists of 208 documents that contain 9,488 annotated events. The categorization model is based on Horn's (1989) analysis of epistemic modality in terms of scalar predication. For epistemic modality Horn proposes the scale ⟨certain, {probable/likely}, possible⟩. For the negative counterpart he proposes the scale ⟨uncertain, {unlikely/improbable}, impossible⟩. Saurí and Pustejovsky map this system into the traditional Square of Opposition (Parsons 2008), which originated with Aristotle. The resulting degrees of factuality defined in FactBank are the following: `fact`, `counterfact`, `probable`, `not probable`, `possible`, `not certain`, `certain but unknown output`, and `unknown or uncommitted`. An example of the `certain but unknown output` value is shown in Example (15) for the event COME, and examples of the `unknown or uncommitted` value

---

2 The event affected by the SSP is underlined.

for the same event are found in Example (16). Discriminatory co-predication tests are provided for the annotators to determine the factuality of events. The interannotator agreement reported for assigning factuality values is κ_{cohen} 0.81.

(15) John knows whether Mary came.

(16) a. John does not know whether Mary came.

b. John does not know that Mary came.

c. John knows that Paul said that Mary came.

A corpus of 50,108 event mentions in blogs and Web posts in Japanese has been annotated with information about extended modality (Matsuyoshi et al. 2010). The annotation scheme of extended modality is based on four desiderata: information should be assigned to the event mention; the modality system has to be language independent; polarity should be divided into two classes: POLARITY ON THE ACTUALITY of the event and SUBJECTIVE POLARITY from the perspective of the source's evaluation; and the annotation labels should not be too fine-grained. In Example (17) the polarity on actuality is negative for the events *STUDY* and *PASS* because they did not occur, but the subjective polarity for the *PASS* event is positive. Extended modality is characterized along seven components: SOURCE, indicating who expresses an attitude towards the event; TIME, future or non future; CONDITIONAL, whether a target event mention is a proposition with a condition; PRIMARY MODALITY TYPE, determining the fundamental meaning of the event mention (assertion, volition, wish, imperative, permission, interrogative); ACTUALITY, degree of certainty; EVALUATION, subjective polarity, which can be positive, negative, or neutral; and FOCUS, what aspect of the event is the focus of negation, inference, or interrogation. Reported inter-annotator agreement for two annotators on 300 event mentions ranges from 0.69 to 0.76 κ_{cohen} depending on the category.

(17) If I had studied mathematics harder, I could have passed the examination.

A publicly available modality lexicon[3] has been developed by Baker et al. (2010) in order to automatically annotate a corpus with modality information. This lexicon contains modal cues related to factivity. The lexicon entries consist of five components: the cue sequence of words, part-of-speech (PoS) for each word, a modality type, a head word, and one or more subcategorization codes. Three components are identified in sentences that contain a modality cue: the TRIGGER is the word or sequence of words that expresses modality; the TARGET is the event, state, or relation that the modality scopes over; and the HOLDER is the experiencer or cognizer of the modality. This scheme distinguishes eight modalities: **requirement** (does H require P?), **permissive** (does H allow P?), **success** (does H succeed in P?), **effort** (does H try to do P?), **intention** (does H intend P?), **ability** (can H do P?), **want** (does H want P?), and **belief** (with what strength does H believe P?). The annotation guidelines to annotate the modalities are defined in Baker et al. (2009).

---

3 Web site of the modality lexicon: `http://www.umiacs.umd.edu/~bonnie/ModalityLexicon.txt`. Last accessed on 8 December 2011.

The scope of negation has been annotated on a corpus of Conan Doyle stories (Morante, Schrauwen, and Daelemans 2011)[4] (*The Hound of the Baskervilles* and *The Adventure of Wisteria Lodge*), which have also been annotated with coreference and semantic roles for the SemEval Task *Linking Events and Their Participants in Discourse* (Ruppenhofer et al. 2010). As for negation, the corpus is annotated with negation cues and their scope in a way similar to the BioScope corpus (Vincze et al. 2008) described subsequently, and in addition negated events are also marked, if they occur in factual statements. Blanco and Moldovan (2011) take a different approach by annotating the **focus**, "that part of the scope that is most prominently or explicitly negated," in the 3,993 verbal negations signaled with MNEG in the PropBank corpus. According to the authors, the annotation of the focus allows the derivation of the implicit positive meaning of negated statements. For example, in Example (18) the focus of the negation is on *until 2008*, and the implicit positive meaning is 'They released the UFO files in 2008.'

(18) They didn't release the UFO files until 2008.

The corpora and categorization schemes described here reflect research focusing on general-domain texts. With the growth of research on biomedical text mining, annotation of modality phenomena in biomedical texts has become central. Scientific language makes use of speculation and hedging to express lack of definite belief. Light, Qiu, and Srinivasan (2004) are pioneers in analyzing the use of speculative language in scientific texts. They study the expression of levels of belief in MEDLINE abstracts by means of hypotheses, tentative conclusions, hedges, and speculations, and annotate a corpus of abstracts in order to check whether the distinctions between high speculative, low speculative, and definite sentences could be made reliably. Their findings suggest that the speculative versus definite distinction is reliable while the distinction between low and high speculative is not.

The annotation work by Wilbur, Rzhetsky, and Shatkay (2006) is motivated by the need to identify and characterize parts of scientific documents where reliable information can be found. They define five dimensions to characterize scientific sentences: FOCUS (scientific versus general), POLARITY (positive versus negative statement), LEVEL OF CERTAINTY in the range 0–3, STRENGTH of evidence, and DIRECTION/TREND (increase or decrease in certain measurement).

A corpus[5] of six articles from the functional genomics literature has been annotated at the sentence level for speculation (Medlock and Briscoe 2007). Sentences are annotated as being speculative or not. Of the 1,157 sentences, 380 were found to be speculative. An inter-annotator agreement of 0.93 κ_{cohen} is reported.

BioInfer (Pyysalo et al. 2007) is a corpus of 1,100 sentences from abstracts of biomedical research articles annotated with protein, gene, and RNA relationships. The annotation scheme captures information about the absence of a relation. Statements expressing absence of a relation such as *not affected by* or *independent of* are annotated using a predicate NOT, as in this example: *not*:NOT(*affect*:AFFECT(*deletion of SIR3, silencing*)).

The Genia Event corpus (Kim, Ohta, and Tsujii 2008) contains 9,372 sentences where biological events are annotated with negation and uncertainty. In the case of negation,

---

4 Web site of the Conan Doyle corpus: `http://www.clips.ua.ac.be/BiographTA/corpora.html`. Last accessed on 8 December 2011.
5 The Medlock and Briscoe corpus is available from `http://www.benmedlock.co.uk/hedgeclassif.html`. Last accessed on 8 December 2011.

events are marked with the label `exists` or `non-exists`. In the case of uncertainty, events are labeled into three categories: `certain`, which is chosen by default; `probable`, if the event existence cannot be stated with certainty; and `doubtful`, if the event is under investigation or forms part of a hypothesis. Linguistic cues are not annotated.

The BioScope corpus (Vincze et al. 2008) is a freely available resource[6] that gathers medical and biological texts. It consists of three parts: clinical free-texts (radiology reports), full-text biological articles, and biological article abstracts from the GENIA corpus (Collier et al. 1999). In total it contains 20,000 sentences. Instances of negative and speculative language are annotated with information about the linguistic cues that express them and their scope. Negation is understood as the implication of the non-existence of something as in Example (19a). Speculative statements express the possible existence of something as in Example (19b). The scope of a keyword is determined by syntax and it is extended to the largest syntactic unit to the right of the cue, including all the complements and adjuncts of verbs and auxiliaries. The inter-annotator agreement rate for scopes is defined as the F-measure of one annotation, treating the second one as the gold standard. It ranges from 62.50 for speculation in full articles to 92.46 for negation in abstracts. All agreement measures are lower for speculation than for negation. The BioScope corpus has been provided as a training corpus for the biological track of the 2010 edition of the CoNLL Shared Task on *Learning to Detect Hedges and their Scope in Natural Language Text* (Farkas et al. 2010b). The additional test files provided in the Shared Task are annotated in the same way.

(19) a. Mildly hyperinflated lungs [**without** focal opacity].

b. This result [**suggests** that the valency of Bi in the material is smaller than +3].

Because the Genia Event and BioScope corpus share 958 abstracts, it is possible to compare their annotations, as it is done by Vincze et al. (2010). Their study shows that the scopes of BioScope are not directly useful to detect the certainty status of the events in Genia, and that the BioScope annotation is more easily adaptable to non-biomedical applications. A description of negation cues and their scope in biomedical texts, based on the cues that occur in the BioScope corpus, can be found in Morante (2010), where information is provided relative to the ambiguity of the negation cue and to the type of scope, as well as examples. The description shows that the scope depends mostly on the PoS of the cue and on the syntactic features of the clause.

The NaCTeM team has annotated events in biomedical texts with meta-knowledge that includes polarity and modality (Thompson et al. 2008). The modality categorization scheme covers epistemic modality and speculation and contains information about the following dimensions: KNOWLEDGE TYPE, LEVEL OF CERTAINTY, and POINT OF VIEW. Four types of knowledge are defined, three of which are based on Palmer's (1986) classification of epistemic modality: `speculative`, `deductive`, `sensory`, and `experimental results or findings`. The levels of certainty are four: `absolute`, `high`, `moderate`, and `low`. The possible values for POINT OF VIEW are `writer` and `other`. An updated version of the meta-knowledge annotation scheme is presented by Nawaz, Thompson, and Ananiadou (2010). The scheme consists of six dimensions: KNOWLEDGE TYPE, `certainty level`, `source`, `lexical polarity`, `manner`, and LOGICAL TYPE. Three

---

6 The BioScope corpus is available from `http://www.inf.u-szeged.hu/rgai/bioscope`. Last accessed on 8 December 2011.

levels of certainty are defined: `low confidence` or considerable speculation, `high confidence` or slight speculation, and `no expression` of uncertainty or speculation. Information about negation is encoded in the LEXICAL POLARITY dimension, which identifies negated events. Negation is defined here as "the absence or non-existence of an entity or a process."

For languages other than English there are much fewer resources. A corpus of 6,740 sentences from the Stockholm Electronic Patient Record Corpus (Dalianis and Velupillai 2010) has been annotated with certain and uncertain expressions as well as speculative and negation cues, with the purpose of creating a resource for the development of automatic detection of speculative language in Swedish clinical text. The categories used are: `certain`, `uncertain`, and `undefined` at sentence level, and `negation`, `speculative words`, and `undefined speculative words` at token level. Inter-annotator agreement for `certain` sentences and negation are high, but for the rest of the classes results are lower.

5. Detection of Speculative Sentences

Initial work on processing speculation focuses on classifying sentences as speculative or definite (non-speculative), depending on whether they contain speculation cues.

Light, Qiu, and Srinivasan (2004) explore the ability of a Support Vector Machine (SVM) classifier to perform this task on a corpus of biomedical abstracts using a stemming representation. The results of the system are compared to a majority decision baseline and to a substring matching baseline produced by classifying as speculative sentences which contain the following strings: *suggest*, *potential*, *likely*, *may*, *at least*, *in part*, *possible*, *potential*, *further investigation*, *unlikely*, *putative*, *insights*, *point toward*, *promise*, and *propose*. The precision results are higher for the SVM classifier (84% compared with 55% for the substring matching method), but the recall results are higher for the substring matching method (79% compared with 39% for the SVM classifier).

Medlock and Briscoe (2007) model hedge classification as a weakly supervised machine learning task performed on articles from the functional genomics literature. They develop a probabilistic learner to acquire training data, which returns a labeled data set from which a probabilistic classifier is trained. The training corpus consists of 300,000 randomly selected sentences; the manually annotated test corpus consists of six full articles.[7] Their classifier obtains 0.76 BEP (Break Even Point), outperforming baseline results obtained with a substring matching technique. Error analysis shows that the system has problems distinguishing between a speculative assertion and one relating to a pattern of observed non-universal behavior, like Example (20), which is wrongly classified as speculative.

(20) Each component consists of a set of subcomponents that can be localized within a larger distributed neural system.

Medlock (2008) presents an extension of this work by experimenting with more features (PoS, stems, and bigrams). Experiments show that although the PoS representation

---

7 The Drosophila melanogaster corpus is available at `http://www.benmedlock.co.uk/hedgeclassif.html`. Last accessed on 8 December 2011.

does not yield significant improvement over the results in Medlock and Briscoe (2007), the system achieves a weakly significant improvement with a stemming representation. The best results are obtained with a combination of stems and adjacent stem bigrams representation (0.82 BEP).

Following Medlock and Briscoe (2007), Szarvas (2008) develops a Maximum Entropy classifier that incorporates bigrams and trigrams in the feature representation and performs a reranking based feature selection procedure that allows a reduction of the number of keyword candidates from 2,407 to 253. The system is trained on the data set of Medlock and Briscoe and evaluated on four newly annotated biomedical full articles[8] and on radiology reports. The best results of the system are achieved by performing automatic and manual feature selection consecutively and by adding external dictionaries. The final results on biomedical articles are 85.29 BEP and 85.08 F_1 score. The results for the external corpus of radiology reports are lower, at 82.07 F_1 score.

A different type of system is presented by Kilicoglu and Bergler (2008), who apply a linguistically motivated approach to the same classification task by using knowledge from existing lexical resources and incorporating syntactic patterns, including **unhedgers**, lexical cues, and patterns that strongly suggest non-speculation. Additionally, hedge cues are weighted by automatically assigning an information gain measure to them and by assigning weights semi-automatically based on their types and centrality to hedging. The hypothesis behind this approach is that "a more linguistically oriented approach can enhance recognition of speculative language." The results are evaluated on the Drosophila data set from Medlock and Briscoe (2007) and the four annotated *BMC Bioinformatics* articles from Szarvas (2008). The best results on the Drosophila data set are obtained with the semi-automatic weighting scheme, which achieves a competitive BEP of 0.85. The best results on the *BMC Bioinformatics* articles are obtained also with semi-automatic weighting yielding a BEP of 0.82 improving over previous results. According to Kilicoglu and Bergler, the best results of the semi-automatic weighting scheme are due to the fact that the scheme relies on the particular semantic properties of the hedging indicators. The relatively stable results of the semi-automatic weighting scheme across data sets could indicate that this scheme is more generalizable than one based on machine learning techniques. The false negatives are due to missing syntactic patterns and to certain derivational forms of epistemic words (*suggest–suggestive*) that are not identified. False positives are due to word sense ambiguity of hedging cues like *could* and *appear*, and to weak hedging cues like epistemic deductive verbs (*conclude, estimate*), some adverbs (*essentially, usually*), and nominalizations (*implication, assumption*).

A different task is introduced by Shatkay et al. (2008). The task consists of classifying sentence **fragments** from biomedical texts along five dimensions, two of which are CERTAINTY (four levels) and POLARITY (negated or not). Fragments are individual statements in the sentences as exemplified in Example (21). For certainty level, the feature vector represents single words, bigrams, and trigrams; for polarity detection, it represents single words and syntactic phrases. They perform a binary classification per class using SVMs. Results on polarity classification are 1.0 F-measure for the positive class and 0.95 for the negative class, and results on level of certainty vary

---

8 The four annotated *BMC Bioinformatics* articles are available at `http://www.inf.u-szeged.hu/∼szarvas/homepage/hedge.html`. Last accessed on 8 December 2011.

from 0.99 F-measure for level 3, which is the majority class, to 0.46 F-measure for level 2.

(21) $\langle$fragment 1We demonstrate that ICG-001 binds specifically to CBP$\rangle$

 $\langle$fragment 2 but not the related transcriptional coactivator p3000$\rangle$

Ganter and Strube (2009) introduce a new domain of analysis. They develop a system for automatic detection of Wikipedia sentences that contain **weasel words**, as in Example (22). Weasel words are "words and phrases aimed at creating an impression that something specific and meaningful has been said, when in fact only a vague or ambiguous claim has been communicated."[9] As Ganter and Strube indicate, weasel words are closely related to hedges and private states. Wikipedia editors are advised to avoid weasel words because they "help to obscure the meaning of biased expressions and are therefore dishonest."[10]

(22) a. Others argue {{weasel-inline}} that the news media are simply catering to public demand.

 b. ... therefore America is viewed by some {{weasel-inline}} technology planners as falling further behind Europe.

Ganter and Strube experiment with two classifiers, one based on words preceding the weasel and another one based on syntactic patterns. The similar results (around 0.70 BEP) of the two classifiers show that word frequency and distance to the weasel tag provide sufficient information. The classifier that uses syntactic patterns outperforms the classifier based on words on data manually re-annotated by the authors, however, suggesting that the syntactic patterns detect weasel words that have not yet been tagged.

Classification of uncertain sentences was consolidated as a task with the 2010 edition of the CoNLL Shared Task on *Learning to Detect Hedges and their Scope in Natural Language Text* (Farkas et al. 2010b), where Task 1 consisted in detecting uncertain sentences. Systems were required to perform a binary classification task on two types of data: biological abstracts and full articles, and paragraphs from Wikipedia. As Farkas et al. describe, the approaches to solving the task follow two major directions: Some systems handle the task as a classical sentence disambiguation problem and apply a bag-of-words approach, and other systems focus on identifying speculation cues, so that sentences containing cues would be classified as uncertain. In this second group some systems apply a token-based classification approach and others use sequential labeling. The typical feature set for Task 1 includes the wordform, lemma or stem, PoS and chunk codes; and some systems incorporate features from the dependency and/or constituent parse tree of the sentences. The evaluation of Task 1 is performed at the sentence level using the F_1 score of the uncertain class. The scores for precision are higher than for recall, and systems are ranked in different positions for each of the data sets, which suggests that the systems are optimized for one of the data types. The top-ranked systems for biological data follow a sequence labeling approach, whereas the

---

9 Definition of weasel words in Wikipedia: http://en.wikipedia.org/wiki/Weasel_word. Last accessed on 8 December 2011.
10 Wikipedia instructions about weasel words are available at http://simple.wikipedia.org/wiki/Wikipedia:Avoid_weasel_words. Last accessed on 8 December 2011.

top-ranked systems for Wikipedia data follow a bag-of-words approach. None of the top-ranked systems uses features derived from syntactic parsing. The best system for Wikipedia data (Georgescul 2010) implements an SVM and obtains an F_1 score of 60.2, whereas the best system for biological data (Tang et al. 2010) incorporates conditional random fields (CRF) and obtains an F_1 score of 86.4.

As a follow-up of the CoNLL Shared Task, Velldal (2011) proposes to handle the hedge detection task as a simple disambiguation problem, restricted to the words that have previously been observed as hedge cues. This reduces the number of examples that need to be considered and the relevant feature space. Velldal develops a large-margin SVM classifier based on simple sequence-oriented n-gram features collected for PoS-tags, lemmas, and surface forms. This system produces better results (86.64 F_1) than the best system of the CoNLL Shared Task (Tang et al. 2010).

From the research presented in this section it seems that classifying sentences as to whether they are speculative or not can be performed by using knowledge-poor machine learning approaches as well as by linguistically motivated methods. It would be interesting to determine whether a combination of both approaches would yield better results. In the machine learning approaches the features used to solve this task are mainly shallow features such as words, bigrams, and trigrams. Syntax features do not seem to add new information, although a linguistically informed method based on syntactic patterns can produce similar results to machine-learning approaches based on shallow features. Hedge cues are ambiguous and domain dependent, reducing the portability of hedge classifiers. It has also been shown that it is feasible to build a hedge classifier in an unsupervised manner.

6. Event-Level Detection of Modality and Negation

Although modality and negation detection at the sentence level can be useful for certain purposes, it is often the case that not all the information contained in a sentence is affected by the presence of modality and negation cues. Modality and negation cues are operators that have a scope and only the part of the sentence within the scope will be affected by them. For example, the sentence in Example (23a)[11] would be classified as speculative in a sentence-level classification task, despite the fact that the cue *unlikely* scopes only over the clause headed by the event PRODUCE. In Example (23b) the negation cue scopes over the subject of *led*, assigning negative polarity to the event COPE_WITH, but not to the rest of the events.

(23) a. He is now an enthusiastic proponent of austerity and reform but this
         has lost him voters and [was **unlikely** to produce sufficient growth, or
         jobs, to win him new ones by next spring].

      b. Its [**in**ability to cope with file-sharing] led to the collapse of
         recorded-music sales and the growing dependence on live music.

Research focusing on determining the scope of cues has revolved around two types of tasks: finding the events and concepts that are negated or speculated, and resolving the full scope of cues. Sections 6.1 and 6.2 describe them in detail.

---

11 The two examples are sentences from articles in *The Economist*.

6.1 Finding Speculated and Negated Events and Entities

Research on finding negated concepts originated in the medical domain motivated by the need to index, extract, and encode clinical information that can be useful for patient care, education, and biomedical studies. In order to automatically process information contained in clinical reports it is of great importance to determine whether symptoms, signs, treatments, outcomes, or any other clinical relevant factors are present or not. As Elkin et al. (2005) state, "erroneous assignment of negation can lead to missing allergies and other important health data that can negatively impact patient safety." Chapman et al. (2001a) point out that accurate indexing of reports requires differentiating **pertinent negatives** from positive conditions. Pertinent negatives are "findings and diseases explicitly or implicitly described as absent in a patient."

The first systems developed to find negated concepts in clinical reports are rule-based and use lexical information. NegExpander (Aronow, Fangfang, and Croft 1999) is a module of a health record classification system. It adds a negation prefix to the negated tokens in order to differentiate between a concept and its negated variant. Negfinder (Mutalik, Deshpande, and Nadkarni 2001) finds negated patterns in dictated medical documents. It is a pipeline system that works in three steps: concept finding to identify UMLS concepts; input transformation to replace every instance of a concept with a coded representation; and a lexing/parsing step to identify negations, negation patterns, and negation terminators. In this system negation is defined as "words implying the total absence of a concept or thing in the current situation." Some phenomena are identified as difficulties for the system: The fact that negation cues can be single words or complex verb phrases like *could not be currently identified*; verbs that when preceded by *not* negate their subject, as in *X is not seen*; and the fact that a single negation cue can scope over several concepts (*A, B, and C are absent*) or over some but not all of them (*there is no A, B and C, but D seemed normal*). Elkin et al. (2005) describe a rule-based system that assigns a level of certainty to concepts in electronic health records. Negation assignment is performed by the automated negation assignment grammar as part of the rule based system that decides whether a concept has been positively, negatively, or uncertainly asserted.

Chapman et al. (2001a, 2001b) developed NegEx,[12] a regular expression based algorithm for determining whether a finding or disease mentioned within narrative medical reports is present or absent. The system uses information about negation phrases that are divided in two groups: pseudo-negation phrases that seem to indicate negation, but instead identify double negatives (*not ruled out*), and phrases that are used as negations when they occur before or after Unified Medical Language System terms. The precision of the system is 84% and the recall 78%. Among the system's weaknesses, the authors report detecting the scope of *not* and *no*. In the three examples in (24a–c) the system would find that *infection* is negated. In Example (24d) *edema* would be found as negated, and in Example (24e) *cva* also. NegEx has been also adapted to process Swedish clinical text (Skeppstedt 2010).

(24) a. This is **not** the source of the infection.

   b. We did **not** treat the infection.

   c. We did **not** detect an infection.

---

12 Web site of NegEx: `http://code.google.com/p/negex/`. Last accessed on 8 December 2011.

    d. **No** cyanosis and positive edema.

    e. **No** history of previous cva.

ConText (Harkema et al. 2009) is an extension of NegEx. This system uses also regular expressions and contextual information in order to determine whether clinical conditions mentioned in clinical reports are negated, hypothetical, historical, or experienced by someone other than the patient. As for negation, a term is negated if it falls within the scope of a negation cue. In this approach, the scope of a cue extends to the right of the cue and ends in a termination term or at the end of the sentence. The system is evaluated on six different types of reports obtaining an average precision of 94% and average recall of 92%. Harkema et al. find that negation cues have the same interpretation across report types.

The systems described here cannot determine correctly the scope of negation cues when the concept is separated by multiple words from the cue. This motivated Huang and Lowe (2007) to build a system based on syntax information. Negated phrases are located within a parse tree by combining regular expression matching and a grammatical approach. To construct the negation grammar, the authors manually identify sentences with negations in 30 radiology reports and mark up negation cues, negated phrases, and negation patterns. The system achieves a precision of 98.6% and a recall of 92.6%. The limitations of this system are related to the comprehensiveness of a manually derived grammar and to the performance of the parser.

Apart from rule-based systems, machine learning techniques have also been applied to find negated and speculated concepts. Goldin and Chapman (2003) experiment with Naïve Bayes and decision trees to determine whether a medical observation is negated by the word *not* in a corpus of hospital reports. The F-measure of both classifiers is similar, 89% and 90%, but Naïve Bayes gets a higher precision and the decision tree a higher recall. Averbuch et al. (2004) develop an Information Gain algorithm for learning negative context patterns in discharge summaries and measure the effect of context identification on the performance of medical information retrieval. 4,129 documents are annotated with appearances of certain terms, which are annotated as positive or negative, as in Example (25).

(25) a. The patient presented with episodes of <u>nausea</u> and vomiting associated
      with epigastric pain for the past 2 weeks. POSITIVE

      b. The patient was able to tolerate food without <u>nausea</u> or vomiting.
      NEGATIVE

Their algorithm scores 97.47 F_1. It selects certain items as indicators of negative context (*any, changes in, changes, denies, had no, negative for, of systems, was no, without*), but it does not select *no* and *not*. As Averbuch et al. (2004, page 284) put it, "Apparently, the mere presence of the word "no" or "not" is not sufficient to indicate negation." The authors point out five sources of errors: coordinate clauses with *but*, as in Example (26a) where *weight loss* is predicted as negative; future reference, as in Example (26b), where the symptoms were predicted as positive; negation indicating existence, as in Example (26c), where *nausea* is predicted as negative; positive adjectives, as in Example (26d), where *appetite* and *weight loss* are predicted as negative; and wrong sentence boundaries.

(26) a. There were **no** acute changes, but she did have a 50 pound weight loss.

b. The patient was given clear instructions to call for any worsening pain, fever, chills, bleeding.

c. The patient could **not** tolerate the nausea and vomiting associated with Carboplatininal Pain.

d. There were **no** fevers, headache, or dizziness at home and **no** diffuse abdominal pain, fair appetite with significant weight loss.

Rokach, Romano, and Maimon (2008) present a pattern-based algorithm for identifying context in free-text medical narratives. The algorithm automatically learns patterns similar to the manually written patterns for negation detection using two algorithms: longest common sequence and Teiresias (Rigoutsos and Floratos 1998), an algorithm designed to discover motifs in biological sequences. A non-ranker filter feature selection algorithm is applied to select the informative patterns (35 out of 2,225). In the classification phase three classifiers are combined sequentially, each learning different types of patterns. Experimental results show that the sequential combination of decision tree classifiers obtains 95.9 F-measure, outperforming the results of single hidden Mankov models and CRF classifiers based on several versions of a bag-of-words representation.

Goryachev et al. (2006) compare the performance of four different methods of negation detection, two regular expression based methods that are adaptations of NegEx and NegExpander, and two classification-based methods, Naïve Bayes and SVM, trained on 1,745 discharge reports. They find that the regular expression-based methods show better agreement with humans and better accuracy than the classification methods. Goryachev et al. indicate that the reason why the classifiers do not perform as well as NegEx and NegExpander may be related to the fact that the classifiers are trained on discharge summaries and tested on outpatient notes.

Another comparison of approaches to assertion classification is made by Uzuner, Zhang, and Sibanda (2009), who develop a statistical assertion classifier, StAC, to classify medical problems in patient records into four categories: `positive`, `negative`, `uncertain`, and `alter-association`[13] assertions. The StAC approach makes use of lexical and syntactic context in conjunction with SVM. It is evaluated on discharge summaries and on radiology reports. The comparison with an extended version of the NegEx algorithm (ENegEx), adapted to capture alter-association in addition to positive, negative, and uncertain assertions, shows a better performance of the statistical classifier for all categories, even when it is trained and tested on different corpora. Results also show that the StAC classifier can solve the task by using the words that occur in a four-word window around the target problem and that it performs well across corpora.

This work focuses mostly on negation in clinical documents, but processing negation and speculation also plays a role in extracting relations and events from the abundant literature on molecular biology. Finding negative cases is useful for filtering out false positives in relation extraction, as support for automatic database curation, or for refining pathways.

Sanchez-Graillet and Poesio (2007) develop a heuristics-based system that extracts negated protein–protein interactions using a full dependency parser from articles about chemistry. The system uses cue words and information from the syntax tree to find potential constructions that express negation. If a negation construction is found, the

---

13 Alter-association assertions state that the problem is not associated with the patient.

system extracts the arguments of the predicate that is negated based on the dependency tree. The maximum F_1 score that the system achieves is 62.96%, whereas the upper-bound of the system with gold-standard protein recognition is 76.68% F_1 score.

The *BioNLP'09 Shared Task on Event Extraction* (Kim et al. 2009) addressed bio-molecular event extraction. It consisted of three subtasks each aiming at different levels of specificity, one of which was dedicated to finding whether the recognized biological events are negated or speculated. Six teams submitted systems with results varying from 2.64 to 23.13 F-measure for negation and 8.95 to 25.27 for speculation. To partic-ipate in this subtask the systems had to first perform Task 1 in order to detect events, which explains the low results. The best scores were obtained by a system that applies syntax-based heuristics (Kilicoglu and Bergler 2009). Once events are identified, the system analyzes the dependency path between the event trigger and speculation or negation cues in order to determine whether the event is within the scope of the cues.

Sarafraz and Nenadic (2010a) further explore the potential of machine learning techniques to detect negated events in the BioNLP'09 Shared Task data. They train an SVM with a model that represents lexical, semantic, and syntax features. The system works with gold-standard event detection and results are obtained by performing 10-fold cross-validation experiments. Evaluation is performed only on gene regula-tion events, which means that the results are not comparable with the Shared Task results. The best results are obtained when all features are combined, achieving a 53.85 F_1 score. Error analysis shows that contrastive patterns like that in Example (27) with the cue *unlike* are recurrent as a source of errors. Sarafraz and Nenadic (2010b) have also compared a machine learning approach with a rule-based approach based on command relations, finding that the machine learning approach produces better re-sults. Optimal results are obtained when individual classifiers are trained for each event class.

> (27) **Unlike** TNFR1, LMP1 can interact directly with receptor-interacting
> protein (RIP) and stably associates with RIP in EBV-transformed
> lymphoblastoid cell lines.

Modality and negation processing at the event level has also been performed on texts from a domain outside the biomedical domain. Here we describe systems that process the factuality of events and a modality tagger.

EvITA and SlinkET (Saurí, Verhagen, and Pustejovsky 2006a, 2006b) are two sys-tems for automatically identifying and tagging events in text and assigning to them contextual modality features. EvITA assigns modality and polarity values to events using pattern-matching techniques over chunks. SlinkET is a rule-based system that identifies contexts of subordination that involve some types of modality, referred to as SLINKs in TimeML (Pustejovsky et al. 2005), and assigns one of the following types to them: `factive`, `counterfactive`, `evidential`, `negative evidential` or `modal`. The reported performance for SlinkET is 92% precision and 56% recall (Saurí, Verhagen, and Pustejovsky 2006a). DeFacto (Saurí 2008) is a factuality profiler. As Saurí puts it, the algorithm assumes a conceptual model where factuality is a property that speakers (sources) attribute to events. Two relevant aspects of the algorithm are that it processes the interaction of different factuality markers scoping over the same event and that it identifies the relevant sources of the event. The system is described in detail in the article by Saurí and Pustejovsky included in this special issue.

Baker et al. (2010) take a different approach. Instead of focusing on an event in order to find its factuality, they focus on modality cues in order to find the predicate

that is within their scope (**target**). They describe two modality taggers that identify modality cues and modality targets, a string-based tagger and a structure-based tagger, and compare their performances. The string-based tagger takes as input text tagged with PoS and marks as modality cues words or phrases that match exactly cues from a modality lexicon. More information about the modality taggers and their application in machine translation can be found in the article by Baker et al. included in this special issue.

Finally, Diab et al. (2009) model belief categorization as a sequence labeling task, which allows them to treat cue detection and scope recognition in a unified fashion. Diab et al. distinguish three belief categories. For **committed belief** the writer indicates clearly that he or she believes a proposition. In the case of **non-committed belief** the writer identifies the proposition as something in which he or she could believe but about which the belief is not strong. This category is further subdivided into **weak belief**, which is often indicated by modals, such as *may*, and **reported speech**. The final category, **not applicable**, refers to cases which typically do not have a belief value associated with them, for example because the proposition does not have a truth value. This category covers questions and wishes. Diab et al. manually annotated a data set consisting of 10,000 words with these categories and then used it to train and test an automatic system for belief identification. The system makes use of a variety of lexical, contextual, and syntactic features. Diab et al. found that relatively simple features such as the tokens in a window around the target word and the PoS tags lead to the best performance, possibly due to the fact that some of the higher level features, such as the verb type, are noisy.

6.2 Full Scope Resolution

The scope resolution task consists of determining at a sentence level which tokens are affected by modality and negation cues. Thanks to the existence of the BioScope corpus several full scope resolvers have been developed. The task was first modeled as a classification problem with the purpose of finding the scope of negation cues in biomedical texts (Morante, Liekens, and Daelemans 2008). It was further developed for modality and negation cues by recent work on the same corpus (Morante and Daelemans 2009a, 2009b; Özgür and Radev 2009), and it was consolidated with the edition of the 2010 CoNLL Shared Task on *Learning to Detect Hedges and their Scope in Natural Language Text* (Farkas et al. 2010a).

Morante, Liekens, and Daelemans (2008) approach the scope resolution task as a classification task. Their conception of the task is inspired by Ramshaw and Marcus's (1995) representation of text chunking as a tagging problem and by the standard CoNLL representation format (Buchholz and Marsi 2006). By setting up the task in this way they show that the task can be modeled as a sequence labeling problem, and by conforming to the existing CoNLL standards they show that scope resolution could be integrated in a joint learning setting with dependency parsing and semantic role labeling. Their system is a memory-based scope finder that tackles the task in two phases: cue identification and scope resolution, which are modeled as consecutive token level classification tasks. Morante and Daelemans (2009b) present another scope resolution system that uses a different architecture, can deal with multiword negation cues, and is tested on the three subcorpora of the BioScope corpus. For resolving the scope, three classifiers (*k*NN, SVM, CRF++) predict whether a token is the first token in the scope sequence, the last, or neither. A fourth classifier is a metalearner that uses the predictions of the

three classifiers to predict the scope classes. The system is evaluated on three corpora using as measure the **percentage of fully correct scopes** (PCS), which is 66.07 for the corpus of abstracts on which the classifiers are trained, 41.00 for the full articles and 70.75 for the clinical reports. They show that the system is portable to different corpora, although performance fluctuates.

Full scope resolution of negation cues has been performed as a support task to determine the polarity of sentiments. In this context, negation is conceived as a contextual valence shifter (Kennedy and Inkpen 2006). If a sentiment is found within the scope of a negation cue, its polarity should be reversed. Several proposals define the scope of a negation cue in terms of a certain number of words to the right of the cue (Pang, Lee, and Vaithyanathan 2002; Hu and Liu 2004), but this solution is not accurate enough. This is why research has been performed on integrating scope resolver into sentiment analysis systems (Jia, Yu, and Meng 2009; Councill, McDonald, and Velikovich 2010).

Jia, Yu, and Meng (2009) describe a rule-based system that uses information from a parse tree. The algorithm first detects a candidate scope and then prunes the words within the candidate scope that do not belong to the scope. The candidate scope of a negation term t is formed by the descendant leaf nodes of the least common-ancestor of the node representing t and the node representing the word t' immediately to the right of t, that are found to the right of t'. Heuristic rules are applied in order to determine the boundaries of the candidate scope. The rules involve the use of **delimiters** (elements that mark the end of the scope), and **conditional delimiters** (elements that mark the end of the scope under certain conditions). Additionally, situations are defined in which a negation cue does not have a scope: phrases like *not only, not just, not to mention, no wonder*, negative rhetorical questions, and restricted comparative sentences. Jia, Yu, and Meng report that incorporating their scope resolution algorithm into two systems that determine the polarity of sentiment words in reviews and in the TREC blogosphere collection produces better accuracy results than incorporating other algorithms that are described in the literature.

Councill, McDonald, and Velikovich (2010) present a system in some aspects similar to the system described by Morante and Daelemans (2009b). The main differences with Morante et al.'s system are that in the first phase, the cues are detected by means of a dictionary of 35 cues instead of being machine learned; in the second phase only a CRF classifier is used, and this classifier incorporates features from dependency syntax. The system is trained and evaluated on the abstracts and clinical reports of the BioScope corpus and on a corpus of product reviews. The PCS reported for the BioScope corpus is 53.7 and 39.8 for the Product Reviews corpus. Cross training results are also reported showing that the system obtains better results for the Product Reviews corpus when trained on BioScope, which, according to the authors, would indicate that the scope boundaries are more difficult to predict in the Product Reviews corpus. Councill et al. also report that the scores of their sentiment analysis system with negation incorporated improve by 29.5% and 11.4% for positive and negative sentiment, respectively. For negative sentiment precision improves 46.8% and recall 6.6%.

It is worth mentioning that the systems trained on the BioScope corpus cannot deal with intersentential, implicit, and affixal negation. Further research could focus on these aspects of negation. Apart from scope resolvers for negation, several full scope resolvers have been developed for modality.

Morante and Daelemans (2009a) test whether the scope resolver for negation (Morante and Daelemans 2009b) is portable to resolve the scope of hedge cues, showing that the same scope resolution approach can be applied to both negation and hedging.

In the scope resolution phase, the system achieves 65.55% PCS in the abstracts corpus, which is very similar to the result obtained by the negation resolver (66.07% PCS). The system is also evaluated on the three types of text of the BioScope corpus. The difference in performance for abstracts and full articles follows the same trends as in the negation system, whereas the drop in performance for the clinical subcorpus is higher, which indicates that there is more variation of modality cues across corpora than there is of negation cues.

The modality scope resolver described by Özgur and Radev (2009) solves the task in two phases also, but differently from Morante and Daelemans (2009a); in the second phase the scope boundaries are found with a rule-based module that uses information from the syntax tree. This system is evaluated on the abstracts and full articles of the BioScope corpus. The scope resolution is evaluated in terms of accuracy, achieving 79.89% in abstracts and 61.13% in full articles.

Task 2 of the 2010 edition of the CoNLL Shared Task (Farkas et al. 2010b) consisted of resolving the scope of hedge cues on biomedical texts. A scope-level F_1 measure was used as the main evaluation metric where true positives were scopes which exactly matched the gold-standard cues and gold-standard scope boundaries assigned to the cue word. The best system (Morante, van Asch, and Daelemans 2010) achieved a F_1 score of 57.3. As Farkas et al. (2010b) describe, each Task 2 system was built upon a Task 1 system, attempting to recognize the scopes for the predicted cue phrases. Most systems regarded multiple cues in a sentence to be independent from each other and formed different classification instances from them. The scope resolution for a certain cue was typically carried out by a token based classification. Systems differ in the number of class labels used as a target and in the machine learning approaches applied. Most systems, following Morante and Daelemans (2009a), used three class labels: `first`, `last`, and `none`, and two systems used four classes by adding `inside`, whereas three systems followed a binary classification approach. Most systems included a postprocessing mechanism to produce continuous scopes, according to the BioScope annotation. Sequence labeling and token-based classification machine learning approaches were applied, and information from the dependency path between the cue and the token in question was generally encoded in the feature space.

The system that scored the best results for Task 2 (Morante, van Asch, and Daelemans 2010) follows the same approach as Morante and Daelemans (2009a), although it introduces substantial differences: This system uses only one classifier to solve Task 2, whereas the system described in Morante and Daelemans (2009a) used three classifiers and a metalearner; this system uses features from both shallow and dependency syntax, instead of only shallow syntax features; and it incorporates in the feature representation information from a lexicon of hedge cues generated from the training data.

As a follow-up of the CoNLL Shared Task, Øvrelid, Velldal, and Oepen (2010) investigate the contribution of syntax to scope resolution. They apply a hybrid, two-stage approach to the scope resolution task. In the first stage, a Maximum Entropy classifier, combining surface-oriented and syntax features, identifies cue words, and multiword cues are identified in a postprocessing step. In the second stage a small set of hand-crafted rules operating over dependency representations are applied to resolve the scope. This system is evaluated following exactly the same settings as the CoNLL Shared Task. The results do not improve over the best shared task results but show that handcrafted syntax-based rules achieve a very competitive performance. Øvrelid et al. report that the errors of their system are mostly of two classes: (i) failing to recognize phrase and clause boundaries, as in Example (28a), and (ii) not dealing successfully with

relatively superficial properties of the text as in Example (28b). The scope boundaries produced by the system are marked with '‖'.

(28) a. ... [the reverse complement ‖mR of m will be **considered** to be ...‖].

b. This ‖[**might** affect the results] if there is a systematic bias on the composition of a protein interaction set‖.

Finally, Zhu et al. (2010) approach the scope learning problem via simplified shallow semantic parsing. The cue is regarded as the predicate and its scope is mapped into several constituents as the arguments of the cue. The system resolves the scope of negation and modality cues in the standard two phase approach. For cue identification they apply an SVM that uses features from the surrounding words and from the structure of the syntax tree. The scope resolution task is different than in previous systems. The task is addressed in three consecutive phases: (1) argument pruning, consisting of collecting as argument candidates any constituent in the parse tree whose parent covers the given cue except the cue node itself and its ancestral constituents; (2) argument identification where a binary classifier is applied to determine the argument candidates as either valid arguments or non-arguments; and (3) postprocessing to guarantee that the scope is a continuous sequence of arguments. The system is trained on the abstracts part of the BioScope corpus and tested on the three parts of the BioScope corpus. Evaluating the system following the CoNLL Shared Task setting would shed more light on the advantages of the semantic parsing approach as compared to other approaches.

From the systems and results described in this section, we can conclude that although there has been substantial research on the scope resolution task, there is still room for improvement. The performance of scope resolvers is still far from having reached the level of well established tasks like semantic role labeling or parsing. Probably, better results can be obtained by a combination of more experimental work with algorithms and a deeper analysis of the task from a linguistic perspective so that the representation models can be improved. The article by Velldal et al. in this special issue provides new insights into the task.

7. Processing Contradiction and Contrast

The concept of negation is closely related to the discourse-level concepts of "contradiction" and "contrast," which typically require an explicit or implicit negation.

Contradiction is a relation that holds between two documents with contradictory content. Detecting contradiction is important for tasks which extract information from multi-document collections, such as question-answering and multi-document summarization. Since 2007 contradiction detection has also been included as a subtask in the Textual Entailment Challenge (Giampiccolo et al. 2007), spurring an increased interest in the development of systems which can automatically detect contradictions. The two contradictory sentence pairs in Examples (29) and (30) (both from Harabagiu et al. [2006]) illustrate the relation between contradiction and negation. In Example (29) the contradiction is signaled by the explicit negation marker *never*, whereas in Example (30) the negation is implicit and signaled by the use of *call off* in the second sentence (which is an antonym of *begin* in the first sentence).

(29) a. Joachim Johansson held off a dramatic fightback from defending champion Andy Roddick, to reach the semi-finals of the US Open on Thursday night.

b. Defending champion Andy Roddick **never** took on Joachim Johansson.

(30) a. In California, one hundred twenty Central Americans, due to be deported, **began** a hunger strike when their deportation was delayed.

b. A hunger strike was **called off**.

Although contradiction typically occurs across documents, **contrast** is a discourse relation within documents. At least some types of contrast involve negation, notably those that involve a denial of expectation. The negation can be explicit as in Example (31a), implicit (31b), or entailed (31c) (see Umbach [2004]).

(31) a. John cleaned his room, but he didn't wash the dishes.

b. John cleaned his room, but he skipped the washing up.

c. John cleaned up the room, but Bill did the dishes.

Given this interrelation between negation and contradiction on the one hand and negation and contrast on the other, it it not surprising that negation detection has been studied in the context of discourse relation classification and contradiction detection. Most studies in this area use fairly standard—i.e., sentence-based—methods for negation detection. Once the negation has been detected it is then used as a feature for the higher-level tasks of contradiction or contrast detection.

For instance, Harabagiu, Hickl, and Lacatusu (2006) discuss a system which first detects negated expressions and then finds contradictions on the basis of the detected negations. To detect explicit negation Harabagiu et al. use a lexicon of explicit cues. To determine the scope they use a set of heuristics, which varies depending on whether the negated object is an event, an entity, or a state. For events, the negation is assumed to scope over the whole predicate–argument structure. For entities and for states realized by nominalizations the negation is assumed to scope over the whole NP. Implicit negations are detected by searching for antonymy chains in WordNet. de Marneffe, Rafferty, and Manning (2008) also make use of negation detection to discover contradictions. They do so rather implicitly, however, by using a number of features which check for explicit negation, polarity, and antonymy. Ritter et al. (2008) present a contradiction detection system that uses the TEXTRUNNER system (Banko et al. 2007) to extract relations of the form *R(x,y)* (e.g., *was_born_in(Mozart,Salzburg)*). They then inspect potential contradictions (i.e., relations which overlap in one variable but not in the other) and filter out non-contradictions by looking, for example, for synonyms and meronyms.

In the context of contrast detection in discourse processing, negation detection is rarely used as an explicit step. An exception is Kim et al. (2006), who are concerned with discovering contrastive information about protein interaction in biomedical texts. They only deal with explicitly marked negation which occurs in the context of a contrast relation marked by a contrast signaling connective such as *but*. Unlike Kim et al., Pitler, Louis, and Nenkova (2009) are concerned with detecting **implicit** discourse relations— namely, relations which are not explicitly signalled by a connective such as *but*. To detect such relations, they define a number of features, including polarity features. Hence they make implicit use of negation information but do not aim to detect it as a separate subtask.

8. Positive and Negative Opinions

Much work in the NLP community has been carried out in the area of identifying positive and negative opinions, also known as **opinion mining**, **sentiment analysis**, or **subjectivity analysis**.[14] Sentiment analysis touches on the topic of this special issue as both negation and modality cues can help determine the opinion of an opinion holder on a subject. Negation in particular has received attention in the sentiment analysis community as negation can affect the polarity of an expression. Negation and polarity are two different concepts, however (see Section 3.1). The relation between negation and polarity is also not always entirely straightforward. For example, whereas negation can change the polarity of an expression from positive to negative (e.g., *good* vs. *not good* in Examples (32a) vs. (32b)) it can also shift negative polarity to neutral or even positive polarity (32c).

  (32) a. This is a **good** camera.

     b. This is **not** a **good** camera.

     c. This is **by no means** a **bad** camera.

In this section, we discuss some approaches that make explicit use of negation in the context of sentiment analysis. For a recent general overview of work on sentiment analysis, we refer the reader to Pang and Lee (2008).

Wiegand et al. (2010) present a survey of the role of negation in sentiment analysis. They indicate that it is necessary to perform fine-grained linguistic analysis in order to extract features for machine learning or rule-based opinion analysis systems. The features allow the incorporation of information about linguistic phenomena such as negation (Wiegand et al. 2010, page 60). Early approaches made use of negation in a bag-of-words model by prefixing a word x with a negation marker if a negation word was detected immediately preceding x (Pang, Lee, and Vaithyanathan 2002). Thus x and NOT_x were treated as two completely separate features. Although this model is relatively simple to compute and leads to an improvement over a bag-of-words model without negation, Pang, Lee, and Vaithyanathan (2002) found that the effect of adding negation was relatively small, possibly because the introduction of additional features corresponding to negated words increases the feature space and thereby also data sparseness. Later work introduced more sophisticated use of negation, for example, by explicitly modeling negation expressions as polarity shifters, which change the polarity of an expression (Kennedy and Inkpen 2006; Polanyi and Zaenen 2006), or by introducing specific negation features (Wilson, Wiebe, and Hoffman 2005; Wilson, Wiebe, and Hwa 2006; Wilson 2008). It was found that these more sophisticated models typically lead to a significant improvement over a simple bag-of-words model with negation prefixes. This improvement can to a large extent be directly attributed to the better modeling of negation (Wilson, Wiebe, and Hoffman 2009). Whereas modeling negation in opinion mining frequently involves determining the polarity of opinions (Hu and Liu 2004; Kim and Hovy 2004; Wilson, Wiebe, and Hoffman 2005; Wilson 2008), some researchers have also used negation models to

---

14 The three terms are used sometimes interchangeably and sometimes reserved for somewhat different contexts. We follow here the definitions of Pang and Lee (2008) who use "opinion mining" and "sentiment analysis" as largely synonymous terms and "subjectivity analysis" as a cover term for both.

determine the strength of opinions (Popescu and Etzioni 2005; Wilson, Wiebe, and Hwa 2006). Choi and Cardie (2010) found that performing both tasks jointly can lead to a significant improvement over a pipeline model in which the two tasks are performed separately. Councill, McDonald, and Velikovich (2010) also show that explicit modeling of negation has a positive effect on polarity detection.

9. Overview of the Articles in this Special Issue

For this special issue we invited articles on all aspects of the computational modeling and processing of modality and negation. Given that this area is both theoretically complex—with several competing linguistic theories having been put forward for various aspects of negation and modality—and computationally challenging, we particularly encouraged submissions with a substantial analysis component, either in the form of a data or task analysis or in the form of a detailed error analysis. We received 25 submissions overall, reflecting a significant interest in these phenomena in the computational linguistics community. After a rigorous review process, we selected five articles, covering various aspects of the topic. Three of the articles (Saurí and Pustejovsky; de Marneffe et al.; and Szarvas et al.) deal with one specific aspect of modality, namely, certainty (in the widest sense) from both a theoretical and a computational perspective. The remaining two articles (Velldal et al. and Baker et al.) deal with both negation and modality detection in a more application-focused setting. The following paragraphs provide a detailed overview of the articles.

In the first article, Saurí and Pustejovsky introduce their model of factuality. They distinguish the dimensions of polarity and certainty and use a four-point scale for the latter. They also explicitly model different sources and embedding of factuality across several levels. They then present a linguistically motivated, symbolic system, DeFacto, for computing factuality and attributing it to the correct sources. The model operates on dependency parses and exploits a number of lexical cues together with hard-coded rules to process factuality within a sentence in a top–down fashion.

Whereas Saurí and Pustejowsky focus on lexical and intra-sentential aspects of factuality, the article by de Marneffe et al. looks specifically at the pragmatic component of factuality (called **veridicality** in their article). They argue that although individual lexemes might be associated with discrete veridicality categories out of context, specific usages are better viewed as evoking probability distributions over veridicality categories, where world knowledge and discourse context can shift the probabilities in one or the other direction. To support this hypothesis, de Marneffe et al. carried out an annotation study with linguistically naive subjects, which provides evidence for considerable variation between subjects, especially with respect to neighboring veridicality categories. In a second step, de Marneffe et al. show how this type of pragmatic veridicality can be modeled in a supervised machine learning setting.

In the following article, Szarvas et al. provide a cross-domain and cross-genre view of (un-)certainty. They propose a novel categorization scheme for uncertainty that unifies existing schemes, which, they argue, are to some extent domain- and genre-dependent. They provide a detailed analysis of different linguistic manifestations of uncertainty in several types of text and then propose a method for adapting uncertainty detection systems to novel domains. They show that instead of simply boosting the available training data from the target domain with randomly selected data from the source domain, it is often more beneficial to select those instances from the source domain that contain uncertainty cues that are also observed in the target domain. In

this scenario, the additional data from the source domain is exploited to fine-tune the disambiguation of target domain cues rather than to learn novel cues.

Moving from certainty to negation and speculation in a more general sense, Velldal et al. show how deep and shallow approaches can be combined for cue detection and scope resolution. They assume a closed class of speculation and negation cues and cast cue detection as a disambiguation rather than a classification task, using supervised machine learning based on *n*-gram features.

In a second step, they tackle scope resolution, for which they propose two models. The first implements a number of syntax-driven rules over dependency structures, and the second model is data-driven and ranks candidate scopes on the basis of constituent trees.

The final article, by Baker et al., also addresses modality and negation processing, but within a particular application scenario, namely, machine translation. The authors propose a novel annotation scheme for modality and negation and two rule-based taggers for identifying cues and scopes. The first tagger employs string matching in combination with a semi-automatically developed cue lexicon; the second goes beyond the surface string and utilizes heuristics based on syntax. In the machine translation process, syntax trees in the source language are then automatically enriched with modality and negation information before being translated.

10. Final Remarks

In this article, we have given an overview of the treatment of negation and modality in computational linguistics. Although much work has been done in recent years and many models for dealing with various aspects of these two phenomena have been proposed, it is clear that much still remains to be done.

The first challenge is a theoretical one and pertains to the categorization and annotation of negation and, especially, modality. Currently, many annotation schemes exist in parallel (see Section 4). As a consequence, the existing annotated corpora are all relatively small. Significant progress in this area depends on the availability of annotated resources, however, both for training and testing automated systems and for (corpus) linguistic studies that can support the development of linguistically informed systems. Ideally, any larger scale resource creation project should be preceded by a discussion in the computational linguistics community about which aspects of negation and modality should be annotated and how this should be done (see, e.g., Nirenburg and McShane [2008]). To some extent this is already happening and the public release of annotated resources such as the MPQA (Wiebe, Wilson, and Cardie 2005) or the BioScope (Vincze et al. 2008) corpus, as well as the organization of shared tasks (Farkas et al. 2010a), are steps in the right direction. Related to this challenge is the question of which aspects of extra-propositional meaning need to be modeled for which applications. Outside sentiment analysis, relatively little research has been carried out in this area so far.

A second challenge involves the adequate modeling of modality and negation. For example, although we can detect extra-propositional content, few researchers so far have investigated how *interactions* between extra-propositional meaning aspects can be adequately modeled. Also, most approaches have addressed the detection of negation at a sentence or predicate level. Discourse-level interdependencies between different aspects of extra-propositional content have been largely ignored. To address this challenge, we believe that more research into linguistically motivated approaches is necessary.

Finally, most research so far has been carried out on English and on selected domains and genres (biomedical, reviews, newswire). It would be interesting to also look at different languages and devise methods for cross-lingual bootstrapping. It would also be good to broaden the set of domains and genres (including fiction, scientific texts, weblogs, etc.) since extra-propositional meaning is particularly susceptible to domain and genre effects.

Acknowledgments

References

Aikhenvald, Alexandra Y. 2004. *Evidentiality*. Oxford University Press, New York.

Aronow, David B., Feng Fangfang, and W. Bruce Croft. 1999. Ad hoc classification of radiology reports. *JAMIA*, 6(5):393–411.

Averbuch, Mordechai, Tom H. Karson, Benjamin Ben-Ami, Oded Maimon, and Lior Rokach. 2004. Context-sensitive medical information retrieval. In *Proceedings of the 11th World Congress on Medical Informatics (MEDINFO-2004)*, pages 1–8, San Francisco, CA.

Baker, Kathrin, Michael Bloodgood, Mona Diab, Bonnie Dorr, Ed Hovy, Lori Levin, Marjorie McShane, Teruko Mitamura, Sergei Nirenburg, Christine Piatko, Owen Rambow, and Gramm Richardson. 2009. SIMT SCALE 2009 modality annotation guidelines. Technical report 4, Human Language Technology Center of Excellence, Baltimore, MD.

Baker, Kathrin, Michael Bloodgood, Bonnie Dorr, Nathaniel W. Filardo, Lori Levin, and Christine Piatko. 2010. A modality lexicon and its use in automatic tagging. In *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC'10)*, pages 1402–1407, Valetta.

Banfield, Ann. 1982. *Unspeakable Sentences*. Routledge and Kegan Paul, Boston, MA.

Banko, Michele, Michael Cafarella, Stephen Soderland, Matt Broadhead, and Oren Etzioni. 2007. Open information extraction from the Web. In *Proceedings of International Joint Conferences on Artificial Intelligence 2007*, pages 2670–2676, Hyderabad.

Blanco, Eduardo and Dan Moldovan. 2011. Semantic representation of negation using focus detection. In *Proceedings of 49th Annual Meeting of the Association for Computational Linguistics*, pages 19–24, Portland, OR.

Boas, Franz, 1938. Language. In Franz Boas, *General Anthropology*. D.C. Heath and Company, Boston, MA, pages 124–145.

Buchholz, Sabine and Erwin Marsi. 2006. CoNLL-X shared task on multilingual dependency parsing. In *Proceedings of the X CoNLL Shared Task*, pages 149–164, New York.

Chafe, Wallace. 1986. Evidentiality in English conversation and academic writing. In W. Chafe and J. Nichols, editors. *Evidentiality: The Linguistic Coding of Epistemology*. Ablex, Norwood, NJ.

Chapman, Wendy W., Will Bridewell, Paul Hanbury, Gregory F. Cooper, and Bruce G. Buchanan. 2001a. A simple algorithm for identifying negated findings and diseases in discharge summaries. *J Biomed Inform*, 34:301–310.

Chapman, Wendy W., Paul Hanbury, Gregory F. Cooper, and Bruce G. Buchanan. 2001b. Evaluation of negation phrases in narrative clinical reports. In *Proceedings of the American Medical Informatics Association Symposium. 2001*, pages 105–109, Washington, DC.

Choi, Yejin and Claire Cardie. 2010. Hierarchical sequential learning for extracting opinions and their attributes. In *Proceedings of the ACL 2010 Conference Short Papers*, pages 269–274, Uppsala.

Collier, Nigel, Hyun S. Park, Norihiro Ogata, Yuka Tateisi, Chikashi Nobata, Tomoko Ohta, Tateshi Sekimizu, Hisao Imai, Katsutoshi Ibushi, and Jun'ichi Tsujii. 1999. The GENIA project: Corpus-based knowledge acquisition and information extraction from genome research papers. In *Proceedings of EACL-99*, pages 271–272, Bergen.

Councill, Isaac, Ryan McDonald, and Leonid Velikovich. 2010. What's great and what's not: Learning to classify the scope of negation for improved sentiment analysis. In *Proceedings of the Workshop on Negation*

and Speculation in Natural Language Processing, pages 51–59, Uppsala.

Dalianis, Hercules and Sumithra Velupillai. 2010. How certain are clinical assessments? Annotating Swedish clinical text for (un)certainties, speculations and negations. In *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC'10)*, Valletta.

de Haan, Frederik. 1995. *The Interaction of Modality and Negation: A Typological Study*. Garland Publishing, Inc., New York.

de Haan, Frederik. 1999. Evidentiality and epistemic modality: Setting the boundaries. *Journal of Linguistics*, 18:83–102.

de Marneffe, Marie-Catherine, Bill Maccartney, Trond Grenager, Daniel Cer, Anna Rafferty, and Christopher D. Manning. 2006. Learning to distinguish valid textual entailments. In *Proceedings of the Second PASCAL Challenges Workshop on Recognising Textual Entailment*, pages 74–79, Venice.

de Marneffe, Marie-Catherine, Anna N. Rafferty, and Christopher D. Manning. 2008. Finding contradictions in text. In *Proceedings of ACL 2008*, pages 1039–1047, Columbus, OH.

Di Marco, Chrysanne, Frederick Kroon, and Robert Mercer. 2006. Using hedges to classify citations in scientific articles. In W. Bruce Croft, James Shanahan, Yan Qu, and Janyce Wiebe, editors, *Computing Attitude and Affect in Text: Theory and Applications*, volume 20 of *The Information Retrieval Series*. Springer, Amsterdam, pages 247–263.

Diab, Mona T., Lori Levin, Teruko Mitamura, Owen Rambow, Vinodkumar Prabhakaran, and Weiwei Guo. 2009. Committed belief annotation and tagging. In *ACL-IJNLP 09: Proceedings of the Third Linguistic Annotation Workshop*, pages 68–73, Singapore.

Elkin, Peter L., Steven H. Brown, Brent A. Bauer, Casey S. Husser, William Carruth, Larry R. Bergstrom, and Dietlind L. Wahner-Roedler. 2005. A controlled trial of automated classification of negation from clinical notes. *BMC Medical Informatics and Decision Making*, 5(13). doi: 10.1186/1472-6947-5-13.

Farkas, Richárd, Veronika Vincze, György Móra, János Csirik, and György Szarvas. 2010a. The CoNLL 2010 shared task: Learning to detect hedges and their scope in natural language text. In *Proceedings of*

the CoNLL2010 Shared Task, pages 1–12, Uppsala.

Farkas, Richárd, Veronika Vincze, György Szarvas, György Móra, and János Csirik, editors. 2010b. *Proceedings of the Fourteenth Conference on Computational Natural Language Learning*, Uppsala.

Friedman, Carol, Philip Alderson, John H. M. Austin, James J. Cimino, and Stephen B. Johnson. 1994. A general natural–language text processor for clinical radiology. *JAMIA*, 1(2):161–174.

Ganter, Viola and Michael Strube. 2009. Finding hedges by chasing weasels: Hedge detection using wikipedia tags and shallow linguistic features. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, pages 173–176, Suntec.

Garson, James. 2009. Modal logic. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Standord University, CA. Available at http://plato.stanford.edu/ archives/win2009/entries/ logic-modal/.

Georgescul, Maria. 2010. A hedgehop over a max-margin framework using hedge cues. In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning*, pages 26–31, Uppsala.

Giampiccolo, Danilo, Bernardo Magnini, Ido Dagan, and Bill Dolan. 2007. The third pascal recognizing textual entailment challenge. In *Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing*, pages 1–9, Prague.

Goldin, Ilya M. and Wendy W. Chapman. 2003. Learning to detect negation with 'Not' in medical texts. In *Proceedings of ACM-SIGIR 2003*, Toronto.

Goryachev, Sergey, Margarita Sordo, Qing T. Zeng, and Long Ngo. 2006. Implementation and evaluation of four different methods of negation detection. Technical report DSG. Harvard Medical School, Boston, MA.

Grabar, Natalia and Thierry Hamon. 2009. Exploitation of speculation markers to identify the structure of biomedical scientifc writing. In *AMIA 2009 Symposium Proceedings*, pages 203–207, San Francisco, CA.

Harabagiu, Sanda, Andrew Hickl, and Finley Lacatusu. 2006. Negation, contrast and contradiction in text processing. In *Proceedings of the 21st International Conference on Artificial Intelligence*, pages 755–762, Las Vegas, NV.

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

Hickl, Andrew and Jeremy Bensley. 2007. A discourse commitment-based framework for recognizing textual entailment. In *Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing*, pages 171–176, Stroudsburg, PA.

Hoeksema, Jack. 2000. Negative polarity items: triggering, scope, and c-command. In L. Horn and Y. Kato, editors, *Negation and Polarity*. Oxford University Press, Oxford, pages 115–146.

Horn, Laurence R. 1989. *A Natural History of Negation*. Chicago University Press, Chicago, IL.

Hu, Minqing and Bing Liu. 2004. Mining and summarizing customer reviews. In *KDD '04: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 168–177, New York.

Huang, Yang and Henry J. Lowe. 2007. A novel hybrid approach to automated negation detection in clinical radiology reports. *Journal of the American Medical Informatics Association*, 14(3):304–311.

Hyland, Ken. 1998. *Hedging in Scientific Research Articles*. John Benjamins B.V., Amsterdam.

Ifantidou, Elly. 2001. *Evidentials and Relevance*. John Benjamins, Amsterdam.

Israel, Michael. 2004. The pragmatics of polarity. In L. Horn and G. Ward, editors, *The Handbook of Pragmatics*. Blackwell, Oxford, pages 701–723.

Jespersen, Otto. 1924. *The Philosophy of Grammar*. Allen and Unwin, London.

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

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

Kilicoglu, Halil and Sabine Bergler. 2008. Recognizing speculative language in biomedical research articles: a linguistically motivated perspective. *BMC Bioinformatics*, 9(Suppl 11):S10.

Kilicoglu, Halil and Sabine Bergler. 2009. Syntactic dependency based heuristics for biological event extraction. In *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing: Shared Task*, BioNLP '09, pages 119–127, Stroudsburg, PA.

Kim, Jin-Dong, Tomoko Ohta, Sampo Pyysalo, Yoshinobu Kano, and Jun'ichi Tsujii. 2009. Overview of BioNLP shared task on event extraction. In *Proceedings of the BioNLP 2009 Workshop Companion Volume for Shared Task 2009*, pages 1–9, Boulder, CO.

Kim, Jin-Dong, Tomoko Ohta, and Jun'ichi Tsujii. 2008. Corpus annotation for mining biomedical events from literature. *BMC Bioinformatics*, 9(10). doi: 10.1186/1471-2105-9-10.

Kim, Jung-Jae, Zhuo Zhang, Jong C. Park, and See-Kiong Ng. 2006. BioContrasts: Extracting and exploiting protein–protein contrastive relations from biomedical literature. *Bioinformatics*, 22:597–605.

Kim, Soo-Min and Ed Hovy. 2004. Determining the sentiment of opinions. In *COLING '04: Proceedings of the 20th International Conference on Computational Linguistics*, page 1367, Morristown, NJ.

Kratzer, Angelika. 1981. The notional category of modality. In H. J. Eikmeyer and H. Rieser, editors, *Words, Worlds, and Contexts. New Approaches in Word Semantics*. De Gruyter, Berlin, pages 38–74.

Kratzer, Angelika. 1991. Modality. In A. von Stechow and D. Wunderlich, editors, *Semantics: An International Handbook of Contemporary Research*. De Gruyter, Berlin, pages 639–650.

Kripke, Saul. 1963. Semantic considerations on modal logic. *Acta Philosophica Fennica*, 16:83–94.

Lakoff, George. 1972. Hedges: a study in meaning criteria and the logic of fuzzy concepts. *Chicago Linguistics Society Papers*, 8:183–228.

Lawler, John. 2010. Negation and negative polarity. In P. C. Hogan, editor, *Cambridge Encyclopedia of the Language Sciences*. Cambridge University Press, Cambridge, UK, pages 554–555.

Light, Mark, Xin Y. Qiu, and Padmini Srinivasan. 2004. The language of bioscience: Facts, speculations, and statements in between. In *Proceedings of BioLINK 2004*, pages 17–24, Boston, MA.

Linguistic Data Consortium. 2008. ACE (Automatic Content Extraction) English annotation guidelines for relations.

Technical Report Version 6.2 2008.04.28. LDC, Philadelphia, PA.

Lyons, John. 1977. *Semantics*. Cambridge University Press, Cambridge, UK.

Matsuyoshi, Suguru, Megumi Eguchi, Chitose Sao, Koji Murakami, Kentaro Inui, and Yuji Matsumoto. 2010. Annotating event mentions in text with modality, focus, and source information. In *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC'10)*, pages 1456–1463, Valletta.

Medlock, Ben. 2008. Exploring hedge identification in biomedical literature. *JBI*, 41:636–654.

Medlock, Ben and Ted Briscoe. 2007. Weakly supervised learning for hedge classification in scientific literature. In *Proceedings of ACL 2007*, pages 992–999, Prague.

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

Morante, Roser and Walter Daelemans. 2009a. Learning the scope of hedge cues in biomedical texts. In *Proceedings of BioNLP 2009*, pages 28–36, Boulder, CO.

Morante, Roser and Walter Daelemans. 2009b. A metalearning approach to processing the scope of negation. In *Proceedings of CoNLL 2009*, pages 28–36, Boulder, CO.

Morante, Roser, Anthony Liekens, and Walter Daelemans. 2008. Learning the scope of negation in biomedical texts. In *Proceedings of EMNLP 2008*, pages 715–724, Honolulu, HI.

Morante, Roser, Sarah Schrauwen, and Walter Daelemans. 2011. Annotation of negation cues and their scope. Guidelines v1.0. Technical Report 3, CLiPS, Antwerp, Belgium.

Morante, Roser, Vincent van Asch, and Walter Daelemans. 2010. Memory-based resolution of in-sentence scopes of hedge cues. In *Proceedings of CoNLL*, pages 40–47, Uppsala.

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

Nawaz, Raheel, Paul Thompson, and Sophia Ananiadou. 2010. Evaluating a meta-knowledge annotation scheme for bio-events. In *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*, pages 69–77, Uppsala.

Nirenburg, Sergei and Marjorie McShane. 2008. Annotating modality. OntoSem final project report. University of Maryland, Baltimore County.

Nirenburg, Sergei and Victor Raskin. 2004. *Ontological Semantics*. MIT Press, Cambridge, MA.

Øvrelid, Lilja, Erik Velldal, and Stephan Oepen. 2010. Syntactic scope resolution in uncertainty analysis. In *Proceedings of the 23rd International Conference on Computational Linguistics*, *COLING '10*, pages 1379–1387, Stroudsburg, PA.

Özgür, Arzucan and Dragomir R. Radev. 2009. Detecting speculations and their scopes in scientific text. In *Proceedings of EMNLP 2009*, pages 1398–1407, Singapore.

Palmer, Frank R. 1986. *Mood and Modality*. Cambridge University Press, Cambridge, UK.

Pang, Bo and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1–2):1–135.

Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing*, pages 79–86. Philadelphia, PA.

Parsons, Terence. 2008. The traditional square of opposition. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Stanford University. Available at `http://plato.stanford.edu/archives/fall2008/entries/square/`.

Payne, Thomas E. 1997. *Describing Morphosyntax*. Cambridge University Press, Cambridge, UK.

Pitler, Emily, Annie Louis, and Ani Nenkova. 2009. Automatic sense prediction for implicit discourse relations in text. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pages 683–691, Suntec.

Polanyi, Livia and Annie Zaenen. 2006. Contextual valence shifters. In W. Bruce Croft, James Shanahan, Yan Qu, and Janyce Wiebe, editors, *Computing Attitude and Affect in Text: Theory and Applications*, volume 20 of *The Information Retrieval Series*. Springer, Netherlands, pages 1–10.

Popescu, Ana-Maria and Oren Etzioni. 2005. Extracting product features and opinions from reviews. In *HLT '05: Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 339–346, Morristown, NJ.

Portner, Paul. 2009. *Modality*. Oxford University Press, Oxford, UK.

Prabhakaran, Vinodkumar, Owen Rambow, and Mona Diab. 2010. Automatic committed belief tagging. In *Proceedings of COLING 2010*, pages 1014–1022, Beijing.

Prasad, Rashmi, Nikhil Dinesh, Alan Lee, Aravind Joshi, and Bonnie Webber. 2006. Annotating attribution in the Penn Discourse TreeBank. In *SST '06: Proceedings of the Workshop on Sentiment and Subjectivity in Text*, pages 31–38, Morristown, NJ.

Prasad, Rashmi, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi, and Bonnie Webber. 2008. The Penn Discourse TreeBank 2.0. In *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, pages 2961–2968, Marrakech.

Pustejovsky, James, Robert Knippen, Jessica Littman, and Roser Saurí. 2005. Temporal and event information in natural language text. *Language Resources and Evaluation*, 39(2–3):123–164.

Pustejovsky, James, Mark Verhagen, Roser Saurí, Jessica Littman, Robert Gaizauskas, Graham Katz, Inderjeet Mani, Robert Knippen, and Andrea Setzer. 2006. Timebank 1.2. Linguistic Data Consortium, Philadelphia, PA.

Pyysalo, Sampo, Filip Ginter, Juho Heimonen, Jari Björne, Jorma Boberg, Jouni Järvinen, and Tapio Salakoski. 2007. BioInfer: A corpus for information extraction in the biomedical domain. *BMC Bioinformatics*, 8(50). doi: 10.1186/1471-2105-8-50.

Ramshaw, Lance and Mitch Marcus. 1995. Text chunking using transformation-based learning. In *Proceedings of ACL Third Workshop on Very Large Corpora*, pages 82–94, Cambridge, MA.

Rigoutsos, Isidore and Aros Floratos. 1998. Combinatorial pattern discovery in biological sequences: The TEIRESIAS algorithm. *Bioinformatics*, 14(1):55–67.

Ritter, Alan, Stephen Soderland, Doug Downey, and Oren Etzioni. 2008. It's a contradiction—no, it's not: A case study using functional relations. In *Proceedings of EMNLP 2008*, pages 11–20, Honolulu, HI.

Rokach, Lior, Roni Romano, and Oded Maimon. 2008. Negation recognition in medical narrative reports. *Information Retrieval Online*, 11(6):499–538.

Rubin, Victoria L. 2006. *Identifying Certainty in Texts*. Ph.D. thesis, Syracuse University, Syracuse, NY.

Rubin, Victoria L. 2007. Stating with certainty or stating with doubt: intercoder reliability results for manual annotation of epistemically modalized statements. In *NAACL '07: Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Companion Volume, Short Papers*, pages 141–144, Morristown, NJ.

Rubin, Victoria L., Elizabeth Liddy, and Noriko Kando. 2005. Certainty identification in texts: Categorization model and manual tagging results. In *Computing Attitude and Affect in Text: Theory and Applications*, volume 20 of *Information Retrieval Series*. Springer-Verlag, New York, pages 61–76.

Ruppenhofer, Joseph, Caroline Sporleder, Roser Morante, Colin Baker, and Martha Palmer. 2010. Semeval-2010 task 10: Linking events and their participants in discourse. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pages 45–50, Uppsala.

Salkie, Raphael, Pierre Busuttil, and Johan van der Auwera. 2009. Introduction. *Modality in English. Theory and Description*, Mouton de Gruyter, Berlin, pages 1–8.

Sanchez-Graillet, Olivia and Massimo Poesio. 2007. Negation of protein–protein interactions: analysis and extraction. *Bioinformatics*, 23(13):424–432.

Sarafraz, Farzaneh and Goran Nenadic. 2010a. Identification of negated regulation events in the literature: Exploring the feature space. In *Proceedings of the Symposium for Semantic Mining in Biomedicine (SMBM 2010)*, pages 137–141, Hinxton.

Sarafraz, Farzaneh and Goran Nenadic. 2010b. Using SVMs with the command relation features to identify negated events in biomedical literature. In *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*, pages 78–85, Uppsala.

Saurí, Roser. 2008. *A Factuality Profiler for Eventualities in Text*. Ph.D. thesis, Brandeis University, Waltham, MA.

Saurí, Roser and James Pustejovsky. 2009. FactBank: A corpus annotated with event factuality. *Language Resources and Evaluation*, 43(3):227–268.

Saurí, Roser, Mark Verhagen, and James
  Pustejovsky. 2006a. Annotating and
  recognizing event modality in text.
  In *Proceedings of FLAIRS 2006*,
  pages 333–339, Melbourne.

Saurí, Roser, Mark Verhagen, and James
  Pustejovsky. 2006b. SlinkET: A partial
  modality parser for events. In *Proceedings
  of LREC 2006*, pages 1332–1337, Genoa.

Seifert, Stephan and Werner Welte. 1987.
  *A Basic Bibliography of Negation in Natural
  Language*. Günter Narr, Tübingen.

Shatkay, Hagit, Fengxia Pan, Andrey
  Rzhetsky, and W. John Wilbur. 2008.
  Multi-dimensional classification of
  biomedical texts: Toward automated
  practical provision of high-utility
  text to diverse users. *Bioinformatics*,
  24(18):2086–2093.

Skeppstedt, Maria. 2010. Negation detection
  in Swedish clinical text. In *Proceedings
  of the NAACL HLT 2010 Second Louhi
  Workshop on Text and Data Mining of Health
  Documents*, pages 15–21, Los Angeles, CA.

Snow, Rion, Lucy Vanderwende, and Arul
  Menezes. 2006. Effectively using syntax for
  recognizing false entailment. In *Proceedings
  of the Main Conference on Human Language
  Technology Conference of the North American
  Chapter of the Association of Computational
  Linguistics*, pages 33–40, Morristown, NJ.

Su, Qi, Chu-Ren Huang, and Helen Kai-yun
  Chen. 2010. Evidentiality for text
  trustworthiness detection. In *Proceedings of
  the 2010 Workshop on NLP and Linguistics:
  Finding the Common Ground*, pages 10–17,
  Uppsala.

Szarvas, György. 2008. Hedge classification
  in biomedical texts with a weakly
  supervised selection of keywords. In
  Proceedings of ACL 2008, pages 281–289,
  Columbus, OH.

Tang, Buzhou, Xiaolong Wang, Xuan Wang,
  Bo Yuan, and Shixi Fan. 2010. A cascade
  method for detecting hedges and their
  scope in natural language text. In
  *Proceedings of the Fourteenth Conference
  on Computational Natural Language
  Learning*, pages 13–17, Uppsala.

Thompson, Paul, Gilua Venturi, John
  McNaught, Simonetta Montemagni, and
  Sophia Ananiadou. 2008. Categorising
  modality in biomedical texts. In *Proceedings
  of the LREC 2008 Workshop on Building and
  Evaluating Resources for Biomedical Text
  Mining 2008*, pages 27–34, Marrakech.

Tottie, Gunnel. 1991. *Negation in English
  Speech and Writing: A Study in Variation*.
  Academic Press, New York.

Umbach, Carla. 2004. On the notion of
  contrast in information structure and
  discourse structure. *Journal of Semantics*,
  21:155–175.

Uzuner, Özlem, Xiaoran Zhang, and
  Tawanda Sibanda. 2009. Machine learning
  and rule-based approaches to assertion
  classification. *JAMIA*, 16(1):109–115.

van der Auwera, Johan and Vladimir A.
  Plungian. 1998. Modality's semantic map.
  Linguistic Typology, 2:79–124.

van der Wouden, Ton. 1997. *Negative
  Contexts: Collocation, Polarity, and Multiple
  Negation*. Routledge, London.

Velldal, Erik. 2011. Predicting speculation:
  A simple disambiguation approach
  to hedge detection in biomedical
  literature. *Journal of Biomedical Semantics*,
  2(Suppl 5):S7.

Vincze, Veronika, György Szarvas, Richard
  Farkas, György Móra, and János Csirik.
  2008. The BioScope corpus: biomedical
  texts annotated for uncertainty, negation
  and their scopes. *BMC Bioinformatics*,
  9(Suppl 11):S9.

Vincze, Veronika, György Szarvas, György
  Móra, Tomoko Ohta, and Richard Farkas.
  2010. Linguistic scope-based and
  biological event-based speculation and
  negation annotations in the genia event
  and BioScope corpora. In *Proceedings
  of the Symposium for Semantic Mining in
  Biomedicine (SMBM 2010)*, pages 84–92,
  Hinxton.

von Fintel, Kai. 2006. Modality and
  language. In D. M. Borchert, editor,
  Encyclopedia of Philosophy. MacMillan
  Reference, Detroit, MI, second edition.
  Available at `http://mit.edu/fintel/
  fintel-2006-modality.pdf`.

von Wright, Georg H. 1951. *An Essay in
  Modal Logic*. North Holland, Amsterdam.

Wiebe, Janyce. 1994. Tracking point of view
  in narrative. *Computational Linguistics*,
  20(2):233–287.

Wiebe, Janyce, Theresa Wilson, Rebecca
  Bruce, Matthew Bell, and Melanie Martin.
  2004. Learning subjective language.
  Computational Linguistics, 30(3):277–308.

Wiebe, Janyce, Theresa Wilson, and Claire
  Cardie. 2005. Annotating expressions of
  opinions and emotions in language.
  Language Resources and Evaluation,
  38:165–210.

Wiegand, Michael, Alexandra Balahur,
  Benjamin Roth, Dietrich Klakow, and
  Andrés Montoyo. 2010. A survey on the
  role of negation in sentiment analysis. In
  Proceedings of the Workshop on Negation and

Speculation in Natural Language Processing, pages 60–68, Uppsala.

Wilbur, W. John, Andrey Rzhetsky, and Hagit Shatkay. 2006. New directions in biomedical text annotations: Definitions, guidelines and corpus construction. *BMC Bioinformatics*, 7:356.

Wilson, Theresa. 2008. *Fine-grained Subjectivity and Sentiment Analysis: Recognizing the Intensity, Polarity, and Attitudes of Private States*. Ph.D. thesis, University of Pittsburgh, Pittsburgh, PA.

Wilson, Theresa, Paul Hoffmann, Swapna Somasundaran, Jason Kessler, Janyce Wiebe, Yejin Choi, Claire Cardie, Ellen Riloff, and Siddharth Patwardhan. 2005. OpinionFinder: A system for subjectivity analysis. In *Proceedings of HLT/EMNLP on Interactive Demonstrations*, pages 34–35, Morristown, NJ.

Wilson, Theresa, Janyce Wiebe, and Paul Hoffman. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of HLT-EMNLP*, pages 347–354, Vancouver.

Wilson, Theresa, Janyce Wiebe, and Paul Hoffman. 2009. Recognizing contextual polarity: An exploration of features for phrase-level analysis. *Computational Linguistics*, 35(3):399–433.

Wilson, Theresa, Janyce Wiebe, and Rebecca Hwa. 2006. Recognizing strong and weak opinion clauses. *Computational Intelligence*, 22(2):73–99.

Zhu, Qiaoming, Junhui Li, Hongling Wang, and Guodong Zhou. 2010. A unified framework for scope learning via simplified shallow semantic parsing. In *Proceedings of EMNLP 2010*, pages 714–724, Cambridge, MA.