T6: Processing modality and negation

Roser Morante, CLiPS - University of Antwerp, Belgium

ABSTRACT

1.1 Tutorial content

Modality and negation are ubiquitous phenomena in language. Generally speaking, modality is a grammatical category that allows to express aspects related to the speaker's attitude towards her statements in terms of degree of certainty, reliability, and subjectivity. In this tutorial modality is understood in a broad sense, which involves related concepts like subjectivity, hedging, evidentiality, uncertainty, committed belief, and factuality. Negation is a grammatical category that allows to change the truth value of a proposition. Modality and negation are treated together because they are interrelated phenomena and are protypically expressed by linguistic devices that share some formal characteristics. For example, modality and negation cues function as operators that scope over certain parts of the sentence.

From a natural language processing perspective, a very relevant aspect of modality and negation is that they encode extra-propositional aspects of meaning. While traditionally most research has focused on propositional aspects of meaning, the interest in processing extra-propositonal aspects has grown in recent years, as a natural consequence of the consolidation of areas that focus on the computational treatment of propositional aspects. Given a sentence, researchers aim at going beyond determining 'who/what does what to whom/what where and when', which would be the goal of a typical semantic role labeling or event extraction task, and are interested in finding also features such as the source, certainty level, epistemological type, truth value, and subjective aspects of the statements contained in a text. Additionally, researchers are also interested in analysing discourse level phenomena such as finding contradictions and textual entailments or modelling how the status of events changes throughout a text. Modality and negation play a main role in these phenomena.

That there is growing interest in these topics among the NLP community is reflected by a number of recent publications, the edition of the workshop 'Negation and Speculation in Natural Language Processing (NeSp-NLP 2010)', as well as the popularity of the CoNLL 2010 shared task on 'Learning to detect hedges and their scope in natural language tex't and the future publication of a special issue of the journal Computational Linguistics. Research on modality and negation has also been stimulated by the release

of a number of data sets annotated with various types of information related to these phenomena.

This tutorial is divided in five modules. In Module 1, I will introduce modality and negation as devices that express extra-propositional aspects of meaning, I will define related concepts and I will show why it is interesting and complex to process them. In Module 2, I will present different categorisation schemes and annotation efforts, as well as an overview of existing resources. In Module 3, I will describe how several related tasks have been modelled and solved. I will present in detail the rule-based and machine learning approaches that have been used to solve the tasks. In Module 4, I will focus on applications that have incorporated the treatment of modality and negation, and on research that analyses the impact of processing these phenomena. The applications range from sentiment analysis to biomedical text mining. Finally, in Module 5, I will summarize achievements and point out open problems.

1.2 Relevance for the ACL community

Processing modality and negation is relevant for the ACL community because of several reasons. First, the treatment of modality and negation is very relevant for all NLP applications that involve text understanding. This includes applications that need to discriminate between factual and non-factual information (uncertain facts, opinions, attitudes, emotions, and beliefs), like information extraction, opinion mining, sentiment analysis, (biomedical) text mining, or question answering, as well as other applications that process the meaning of texts, like recognizing textual entailment, paraphrasing, or summarization. Incorporating information about modality and negation has been shown to be useful for a number of applications, such as biomedical text processing (Friedman et al., 1994; Di Marco and Mercer, 2005; Mutalik et al., 2001; Chapman et al., 2001), opinion mining and sentiment analysis (Wilson et al., 2005a), recognizing textual entailment (Marneffe et al., 2006; Snow et al., 2006), and automatic style checking (Ganter and Strube, 2009). Hence this topic is of general importance to the NLP community as a whole, as evidenced by the fact that a number of researchers and groups are currently working on this phenomena.

Second, this topic has received a noticeable boost from several recent events: the workshop Negation and Speculation in Natural Language Processing (NeSp-NLP 2010), which I co-organized in Uppsala just before ACL 2010; the CoNLL Shared Task 2010 on Learning to detect hedges and their scope in natural language text, which attracted 51 submissions from 23 teams; and the publication of the Special Issue on Modality and Negation by the journal Computational Linguistics, that will appear at the end of 2011. This SI has received a considerable number of submissions, which shows that the community is active in treating these phenomena. Research on modality and

negation is also supported by the fact that a number of data sets annotated with various aspects of modality and negation information have been made available, such as the MPQA Opinion Corpus (Wiebe et al., 2005), Rubinâ€™s (2006; 2007) certainty corpus, the ACE 2008 corpus (Linguistic Data Consortium, 2008), and the FactBank corpus (2009), the BioScope corpus (Vincze et al., 2008).

Given that there is clearly substantial interest in this topic from the ACL community and given that the research on this area is evolving quickly, I believe that the proposed tutorial will help attendees to keep up to date with recent advances in the field and discover new directions for future research.

OUTLINE

1. **Introduction: modality and negation as extra-propositional aspects of meaning**
    1. Defining modality
    2. Concepts related to modality: hedging, evidentiality, uncertainty, factuality, subjectivity, non-committed belief
    3. Defining negation
    4. Negation versus negative polarity
    5. Why is it interesting to process modality and negation?
    6. Complex aspects of processing modality and negation
2. **Categorising and annotating negation and modality**
    1. Annotation schemes
    2. Existing resources
    3. Future directions
3. **Tasks related to processing modality and negation**
    1. Cue detection
    2. Detecting speculated sentences
    3. Scope resolution
    4. Finding negated/speculated events
    5. Modality tagging
    6. Belief categorisation
    7. Processing contradiction and contrast
    8. Incorporating negation/modality in parsers
4. **Modality and negation in applications**
    1. Sentiment analysis
    2. Recognizing textual entailment
    3. Machine translation
    4. Text mining
5. **Open problems and discussion**

BIO

Roser Morante
Computational Linguistics and Psycholinguistics (CLiPS) Research Center
University of Antwerp
Email: roser.morante --AT-- ua dot ac dot be
http://www.clips.ua.ac.be/~roser/

Dr. Roser Morante is a senior researcher at CLiPS, a research center associated with the Linguistics Department of the Faculty of Arts at the University of Antwerp, Belgium. She obtained her PhD in Computational Linguistics at the University of Tilburg, The Netherlands, where she also worked as a postdoctoral researcher. She is currently working on the Biograph project led by Professor Walter Daelemans, where she applies text mining techniques to extract biomedical relations from scientific texts. In the project she has worked extensively on both modality and negation. She proposed the first model of the scope finding task as a machine learning classification task and has developed systems for finding the scope of negation and hedge cues. The system that her team submitted to the CoNLL Shared Task 2010 scored first in Task 2 on finding the scope of hedge cues. She has co-oganized the Workshop on Negation and Speculation in Natural Language Processing (NeSp-NLP 2010) and she is currently a Guest Editor of the Special Issue on Modality and Negation for the journal Computational Linguistics. She has also been involved in the organization of the Workshop Advances on Bio Text Mining 2010, the SemEval 2010 shared task on Linking Events and Their Participants in Discourse, and the evaluation exercise Processing modality and negation for machine reading, a pilot task of the Question Answering for Machine Reading Evaluation at CLEF 2011.

My research interests revolve around applying supervised machine learning techniques to semantics oriented natural language processing tasks at sentence and document level, like semantic role labeling, relation extraction, coreference resolution, paraphrasing, or textual entailment. In the past I have worked on machine translation, dialogue systems and semantic role labeling. At this moment I am working mostly on biomedical text mining: processing negation and modality, extraction of biomedical events, extraction of protein-protein interactions, semantic role labeling, and coreference resolution.