

# **Negation and Speculation in Natural Language Processing (NeSp-NLP 2010)**

**Proceedings of the Workshop**

10 July 2010  
Uppsala, Sweden

Roser Morante and Caroline Sporleder (eds.)

Proceedings of the Workshop on  
Negation and Speculation in Natural Language Processing  
(NeSp-NLP 2010)

July 2010

ISBN: 9789057282669

EAN: 9789057282669

University of Antwerp

Prinsstraat 13

B-2000 Antwerp

Belgium

Tel: +32(0)3 265 41 11

Fax: +32(0)3 265 44 20

<http://www.ua.ac.be>

These proceedings were compiled with the ACLPUB package.

**Organizers:**

Roser Morante - University of Antwerp (Belgium)  
Caroline Sporleder - Saarland University (Germany)

**Program Committee:**

Timothy Baldwin - University Melbourne (Australia)  
Xavier Carreras - Technical University of Catalonia (Spain)  
Wendy W. Chapman - University of Pittsburgh (USA)  
Kevin B. Cohen - University of Colorado (USA)  
Walter Daelemans - University of Antwerp (Belgium)  
Guy De Pauw - University of Antwerp (Belgium)  
Bonnie Dorr - University of Maryland (USA)  
Roxana Girju - University of Illinois at Urbana-Champaign (USA)  
Iris Hendrickx - University of Lisbon (Portugal)  
Veronique Hoste - University College Ghent (Belgium)  
Halil Kilicoglu - Concordia University (Canada)  
Martin Krallinger - CNIO (Spain)  
Lori Levin - Carnegie Mellon University (USA)  
Maria Liakata - Aberystwyth University, European Bioinformatics Institute (UK)  
Lluís Màrquez - Technical University of Catalonia (Spain)  
Erwin Marsi - Tilburg University (The Netherlands)  
Arzucan Özgür - University of Michigan (USA)  
Manfred Pinkal - Saarland University (Germany)  
Sampo Pyysalo - University of Tokyo (Japan)  
Owen Rambow - Columbia University (USA)  
Josef Ruppenhofer - Saarland University (Germany)  
Roser Saurí - Barcelona Media Innovation Center (Spain)  
Khalil Sima'an - University of Amsterdam (The Netherlands)  
Mihai Surdeanu - Stanford University (USA)  
Antal van den Bosch - Tilburg University (The Netherlands)  
Michael Wiegand - Saarland University (Germany)

**Invited Presentations:**

Ed Hovy - ISI, University of Southern California (USA)  
Martin Krallinger - CNIO (Spain)  
Maria Liakata - Aberystwyth University, European Bioinformatics Institute (UK)  
Raheel Nawaz, Paul Thompson, Sophia Ananiadou - NaCTeM (UK )  
Veronika Vincze - BioScope group, University of Szeged (Hungary)

**Institutions that organize the workshop:**



**ACL SIGs that endorse the workshop:**



**Website:**

<http://www.clips.ua.ac.be/NeSpNLP2010/>

**Sponsors:**

GOA project Biograph from the University of Antwerp  
Cluster of Excellence, MMCI, Saarland University

## Introduction

These proceedings contain the papers and invited talks presented at the Workshop on Negation and Speculation in Natural Language Processing (NeSp-NLP 2010) that was held on the 10th of July, 2010 in Uppsala, Sweden. The program consisted of five invited talks, seven presentations of long papers and two of short papers.

When we thought of organising this workshop, we aimed at bringing together researchers working on negation and speculation from any area related to computational language learning and processing. Specific goals were to describe the lexical aspects of negation and speculation, to define how the semantics of these phenomena can be modelled for computational purposes, to explore techniques aimed at learning the factuality of an statement, and to analyse how the treatment of these phenomena affects the efficiency of Natural Language Processing (NLP) applications.

Negation and speculation are two linguistic phenomena involved in deep understanding of text. They are resources used to express the factuality of statements, which indicates to which extent a statement is or is not a fact. Negation turns an affirmative statement into negative (it rains/it does not rain). Speculation is used to express levels of certainty (it might rain/apparently, it will rain/ it is likely to rain/it is not clear whether it will rain/we suspect that it will rain). We knew that negation and speculation (or modality) have been extensively studied from a theoretical perspective. Furthermore, we also believed that there was enough interest on these topics among the NLP community and that there was enough research going on, so as to organise a topical workshop, the first of its kind as far as we know.

We cannot be exhaustive here about all the NLP related work on these topics that has been published before the workshop. We apologise for mentioning only some references as examples of research that is being carried out, which motivated our decision of organising a workshop. As recent references, the BioScope corpus has been annotated with negation and speculation cues and their scope (Vincze et al. 2009); events in the FactBank corpus (Saurí and Pustejovsky 2009) have been annotated with factuality information; the CoNLL Shared Task 2010 (Farkas et al. 2010) focused on *Learning to detect hedges and their scope in natural language text*. The biomedical text mining community has produced tools to process negation, like Context (Harkema et al. 2009), and negation has also received attention from researchers working on sentiment analysis (Wilson et al. 2009 and work cited in Wiegand et al. 2010).

We proposed the following topics in the call for papers of the workshop:

- Lexical aspects of negation and speculation
- Linguistic resources with information about negation and speculation: corpora, dictionaries, lexical databases
- Descriptive analysis of negation and speculation cues
- Negation and speculation across domains and genres
- Negation and speculation in biomedical texts and biomedical text mining
- Handling negation and speculation in NLP: dialogue systems, sentiment analysis, text mining, textual entailment, information extraction, machine translation, paraphrasing

- Learning the scope of negation and speculation cues
- Interaction of negation and speculation for evaluating the factuality of an statement
- Corpora annotation: guidelines, bootstrapping techniques, quality assessment
- Modelling factuality for computational purposes
- Algorithms to learn negation and speculation
- Structured prediction of negation and speculation
- Joint learning of negation and speculation
- Inference of factual knowledge

Although we did not receive submissions addressing all the proposed topics, the fact that we received submissions addressing some of them makes us consider that the main goal of the workshop was achieved, and that there is a growing interest in processing negation and speculation within several NLP subareas. From the nine accepted papers, six report research on biomedical texts, four of which are related to either manual or automatic annotation of corpora, one to automatically identifying negated biomedical events, and one to evaluating whether identifying negation and speculation helps in classifying medical reports. Two papers deal with negation in sentiment analysis, one focuses on automatically learning the scope and another surveys the role of negation in sentiment analysis. One paper reports research on the relation between positive and negative pairs in textual entailment.

Four of the five invited presentations are from the biomedical domain. Maria Liakata presents an annotation scheme for annotating full papers with zones of conceptualisation levels to identify the core components that constitute a scientific investigation. Veronika Vincze presents the difficulties encountered during annotation process of the BioScope corpus. Martin Krallinger elaborates on the importance of negations and experimental qualifiers to extract information from biomedical literature, and Raheel Nawaz, Paul Thompson, and Sophia Ananiadou discuss the evaluation of a meta-knowledge annotation scheme for bio-events. Finally, Ed Hovy, invites us to consider Distributional Semantics as a model for richer and more semantics-oriented statistics-based NLP. He presents a specific model of Distributional Semantics, and explores the possibilities for accommodating the phenomena of negation and modality.

We would like to thank the authors of the papers for their interesting contributions, the members of the program committee for their insightful reviews, and the presenters of invited talks for accepting the invitation to give a talk at the workshop and share their work. We are grateful to Walter Daelemans for encouraging us to organise the workshop. The workshop would not have been possible without their help. We appreciate very much the knowledge, time, and effort that they invested in the workshop. We are also thankful to the University of Antwerp and Saarland University for their institutional support and to the SIGs that endorsed the workshop. We sincerely hope that in the future the NLP community will benefit from the findings made by researchers working on negation, speculation and other phenomena involved in determining the factuality of an event.

Roser Morante and Caroline Sporleder  
July 2010

## References

- R. Farkas, V. Vincze, G. Móra, G. Szarvas, and J. Csirik (2010) The CoNLL 2010 shared task: Learning to detect hedges and their scope in natural language text. In *Proceedings of the CoNLL 2010 Shared Task*.
- J. Harkema, J.N. Dowling, T. Thornblade, and W. W. Chapman (2009) ConText: An algorithm for determining negation, experiencer, and temporal status from clinical reports. *Journal of Biomedical Informatics* 42(5), 839-851.
- R. Saurí and J. Pustejovsky (2009) FactBank: A Corpus Annotated with Event Factuality. *Language Resources and Evaluation*, 43(3):227-268.
- V. Vincze, G. Szarvas, R. Farkas, G. Móra, and J. Csirik (2008) The BioScope corpus: biomedical texts annotated for uncertainty, negation and their scopes. *BMC Bioinformatics*, 9(Suppl 11):S9.
- T. Wilson, J. Wiebe, and P. Hoffman (2009) Recognizing contextual polarity: An exploration for phrase-level analysis. *Computational Linguistics* 35:3.
- M. Wiegand, A. Balahur, B. Roth, D. Klakow, and A. Montoyo (2010) A Survey on the role of negation in sentiment analysis. In this volume.





## Program NeSp-NLP 2010

08.45 - 9:00	Opening
9:00 - 9:35	<i>Zones of conceptualisation in scientific papers: a window to negative and speculative statements</i> Maria Liakata (Invited presentation)
9:35 - 10:00	<i>Creating and evaluating a consensus for negated and speculative words in a Swedish clinical corpus</i> Hercules Dalianis and Maria Skeppstedt
10:00 - 10:25	<i>Towards a better understanding of uncertainties and speculations in Swedish clinical text Analysis of an initial annotation trial</i> Sumithra Velupillai
10:25 - 10:40	<i>Does negation really matter?</i> Ira Goldstein and Özlem Uzuner
10:40 - 11:10	Coffee break
11:10 - 11:45	<i>Speculation and negation annotation in natural language texts: what the case of BioScope might (not) reveal</i> Veronika Vincze (Invited presentation)
11:45 - 12:10	<i>Automatic annotation of speculation in biomedical texts: new perspectives and large-scale evaluation</i> Julien Desclés, Olfa Makkaoui, and Taouise Hacène
12:10 - 12:25	<i>Levels of certainty in knowledge-intensive corpora: an initial annotation study</i> Aron Henriksson and Sumithra Velupillai
12:25 - 12:50	<i>Importance of negations and experimental qualifiers in biomedical literature</i> Martin Krallinger (Invited presentation)
12:50 - 13.50	Lunch
13:50 - 14:50	<i>Negation and modality in distributional semantics</i> Ed Hovy (Invited presentation)
14:50 - 15:15	<i>What's great and what's not: Learning to classify the scope of negation for improved sentiment analysis</i> Isaac Council, Ryan McDonald, and Leonid Velikovich
15:15 - 15:40	<i>A survey on the role of negation in sentiment analysis</i> Michael Wiegand, Alexandra Balahur, Benjamin Roth, Dietrich Klakow, and Andrés Montoyo
15.40 - 16.10	Coffee break
16:10 - 16:45	<i>Evaluating a meta-knowledge annotation scheme for bio-events</i> Raheel Nawaz, Paul Thompson, and Sophia Ananiadou (Invited presentation)
16:45 - 17:10	<i>Using SVMs with the command relation features to identify negated events in biomedical literature</i> Farzaneh Sarafraz and Goran Nenadic
17:10 - 17:35	<i>Contradiction-focused qualitative evaluation of textual entailment</i> Bernardo Magnini and Elena Cabrio
17:35 - 18.00	<i>Discussion</i>



## Table of Contents

<i>Zones of conceptualisation in scientific papers: a window to negative and speculative statements</i> Maria Liakata .....	1
<i>Creating and evaluating a consensus for negated and speculative words in a Swedish clinical corpus</i> Hercules Dalianis and Maria Skeppstedt .....	5
<i>Towards a better understanding of uncertainties and speculations in Swedish clinical text Analysis of an initial annotation trial</i> Sumithra Velupillai .....	14
<i>Does negation really matter?</i> Ira Goldstein and Özlem Uzuner .....	23
<i>Speculation and negation annotation in natural language texts: what the case of BioScope might (not) reveal</i> Veronika Vincze .....	28
<i>Automatic annotation of speculation in biomedical texts: new perspectives and large-scale evaluation</i> Julien Desclés, Olfa Makkaoui and Taouise Hacène .....	32
<i>Levels of certainty in knowledge-intensive corpora: an initial annotation study</i> Aron Henriksson and Sumithra Velupillai .....	41
<i>Importance of negations and experimental qualifiers in biomedical literature</i> Martin Krallinger .....	46
<i>Negation and modality in distributional semantics</i> Ed Hovy .....	50
<i>What's great and what's not: learning to classify the scope of negation for improved sentiment analysis</i> Isaac Council, Ryan McDonald and Leonid Velikovich .....	51
<i>A survey on the role of negation in sentiment analysis</i> Michael Wiegand, Alexandra Balahur, Benjamin Roth, Dietrich Klakow and Andrés Montoyo .....	60
<i>Evaluating a meta-knowledge annotation scheme for bio-events</i> Raheel Nawaz, Paul Thompson and Sophia Ananiadou .....	69
<i>Using SVMs with the Command Relation features to identify negated events in biomedical literature</i> Farzaneh Sarafraz and Goran Nenadic .....	78
<i>Contradiction-focused qualitative evaluation of textual entailment</i> Bernardo Magnini and Elena Cabrio .....	86
<i>Discussion items</i> .....	95

