Procesamiento del Lenguaje Natural, Revista nº 47 septiembre de 2011, pp 315-316 recibido 29-04-2011 aceptado 23-05-2011

# Inferring the Scope of Negation and Speculation Via Dependency Analysis

## *Obteniendo el Alcance de la Negación y la Especulación a través de Análisis de Dependencias*

**Miguel Ballesteros, Virginia Francisco, Alberto Díaz, Jesús Herrera, Pablo Gervás**
Departamento de Ingeniería del Software e Inteligencia Artificial
Universidad Complutense de Madrid
C/ Profesor José García Santesmases s/n, Madrid, Spain
{miballes, virginia, albertodiaz, jesus.herrera}@fdi.ucm.es, pgervas@sip.ucm.es

**Resumen:** En los últimos años se han desarrollado de manera satifactoria sistemas basados en la detección y clasificación del alcance de la negación y la especulación, la mayoría utilizando aprendizaje automático. En este trabajo presentamos un sistema competitivo que infiere el alcance de la negación y la especulación para el Inglés a partir del análisis sintáctico de las frases sin utilizar aprendizaje automático. Un algoritmo voraz detecta en primer lugar las señales de negación y especulación, y en segundo lugar las palabras que están dentro del alcance de esas señales; luego, se ha desarrollado otro algoritmo para inferir y finalmente anotar el alcance de la negación y la especulación. Hemos evaluado nuestro sistema con el corpus Bioscope.
**Palabras clave:** Análisis sintáctico de dependencias, Procesamiento de negación y modalidad.

**Abstract:** In the last few years negation and speculation scope classification systems have been developed successfully, most of them using machine–learning approaches. In this paper we present a competitive system that finds the scope of negation and speculation in sentences for English by means of syntactic dependency analysis without using machine–learning. A greedy algorithm first detects the cues (lexical markers that express negation and speculation), and second the words within the scope of these cues; finally, a Scope Finding algorithm uses these affected words to infer and to annotate the scope of the negation and the speculation. We tested our system with the Bioscope corpus, annotated with speculation and negation.
**Keywords:** Dependency parsing, Processing negation and modality.

## 1 Introduction

Every text contains information that includes uncertainty, deniability or speculation. In the context on information extraction and text mining, the study of these phenomena has grown in importance in recent years, because it is important to distinguish between speculative/negative statements and factual ones. Negation and speculation are complex phenomena in natural languages and have been an active research topic for decades. In 2010, a Workshop on Negation and Speculation in Natural Language Processing (NeSp-NLP 2010) was held in Uppsala, Sweden, bringing together researchers working on negation and speculation from fields related to computational language learning and processing (Morante and Sporleder, 2010).

Our work is framed in this context and our main goal is to show how to classify the scope of negation and speculation by means of syntactic dependency parsing.

## 2 System Architecture

We developed an algorithm that detects wordforms within the scope of cues based on dependency analysis, using the Minipar parser, and we show that the domain application is somewhat open using a different lexicon of cues. We also implemented a Scope Finding Algorithm that uses the output of the Affected Wordforms Detection Algorithm to annotate sentences with the scope of cues. The system annotates the sentences using the

ISSN 1135-5948 

guidelines of the Bioscope corpus[1] (Vincze et al., 2008).

The Affected Wordforms Detection Algorithm uses the dependency binary asymmetric relations that inform about which nodes depend on others, so we can consider the nodes that shared the same branch with a negation or hedge signal or which nodes directly depend on a negation or hedge signal. Additionally, our algorithm runs through the tree until it finds terminals, it can find the wordforms deepest in the tree structure that depend on, or are related to, a negation or hedge signal that infers the scope of the cues. Our contribution lies in the identification of the scope, which is not explicit in the dependency tree.

Finally, a Scope Finding algorithm was implemented and it uses the set of words returned by the Affected Wordforms Detection Algorithm and the dependency tree given by Minipar. This second algorithm returns sentences annotated with the scope of cues, inferring where the scope must be opened and where it must be closed. The first step of our Scope Finding algorithm is related to the voice of the sentence, because where to open the scope is related to it: if the sentence is in passive voice the scope must be opened to the left of the cue and if the sentence is in active voice, the scope must generally be opened to the right of the cue.

## 3 Evaluation

We tested the whole system with the three collections of Bioscope: the Scientific Papers Collection, the Abstracts Collection and the Clinical Radiology Reports Collection. In Bioscope, every sentence is annotated with information about negation and speculation. The scopes can also be nested. We also compared our system with the best and most reliable system of the state of the art (Morante and Daelemans, 2009b; Morante and Daelemans, 2009a), that are based on machine–learning. Taking all of this into account, our results are really competitive.

In Table 1 we show the data obtain for the Percentage of Correct Scopes (PCS) for the two different tasks and the three collections of Bioscope. Our system did not use neither golden cues nor golden trees, that means that our system had as input the plain text and it returned all the needed information, where to open the scope, where to close it and which wordform is the cue and where it is.

[1] www.inf.u-szeged.hu/rgai/bioscope

| Task | Papers | Abstracts | Clinical |
|---|---|---|---|
| Our System. Negation | **61.43%** | **68.92%** | **89.06%** |
| Morante et al. Negation | 41.00% | 66.07% | 70.75% |
| Our System. Speculation | **39.43%** | 46.75% | **36.20%** |
| Morante et al. Speculation | 35.92% | **65.55%** | 26.21% |

Table 1: Results of our approach, evaluated with the three Collections of Bioscope, looking attentively at PCS (Percentage of Correct Scopes) and compared with Morante's results.

## 4 Demo

We made our system accesible via web, therefore, in the demonstration we will show the execution and the architecture of the two online systems developed (negation approach[2] and speculation approach[3]), using a slot of at least 10 minutes.

## Acknowledgments

This research is funded by the Spanish Ministry of Education and Science (TIN2009-14659-C03-01 Project), Universidad Complutense de Madrid and Banco Santander Central Hispano (GR58/08 Research Group Grant).

## References

Morante, Roser and Walter Daelemans. 2009a. Learning the scope of hedge cues in biomedical texts. In *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing*.

Morante, Roser and Walter Daelemans. 2009b. A metalearning approach to processing the scope of negation. In *CoNLL*.

Morante, Roser and Caroline Sporleder, editors. 2010. *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*, 07/2010.

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

[2] http://minerva.fdi.ucm.es:8888/ScopeTagger/
[3] http://minerva.fdi.ucm.es:8888/ScopeTaggerSpec/