INFERRING THE SCOPE OF SPECULATION USING DEPENDENCY ANALYSIS

Miguel Ballesteros, Virginia Francisco, Alberto Díaz, Jesús Herrera and Pablo Gervás
Departamento de Ingeniera 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*

Keywords: Speculation, Dependency Parsing, Information Extraction, Biomedical Texts.

Abstract: In the last few years speculation detection systems for biomedical texts have been developed successfully, most of them using machine–learning approaches. In this paper we present a system that finds the scope of speculation in English sentences, by means of dependency syntactic analysis. It infers which words are affected by speculation by browsing dependency syntactic structures. Thus, firstly an algorithm detects hedge cues[a]. Secondly the scope of these hedge cues is computed. We tested the system with the Bioscope corpus, annotated with speculation and obtaining competitive results compared with the state of the art systems.

[a]The cue is defined as the lexical marker that expresses speculation, like *might* or *may*.

1 INTRODUCTION

Every text contains information that includes uncertainty, deniability or speculation. Interest in speculation has grown in recent years in the context of research on information extraction and text mining. Moreover, lots of information in texts consisted on non–factual information that informs about probability, casuality or uncertainty. For example in *he may be wrong but he thinks you would be wise to go*, *may* expresses contingency and condition and *would* expresses uncertainty.

Nowadays speculation detection is an emergent task and it has been one of the most recent advances in natural language processing research. Detecting uncertain and hedged assertions is essential in most text mining tasks where, in general, the aim is to derive factual knowledge from textual data. Moreover, the presence of speculation can yield also to obtain non factual information from texts.

In this paper we present a speculation scope finding system for English using dependency analysis. The aim of this paper is to show that dependency analysis is useful to detect speculation and the words within the scope. But finding speculative sentences is not the goal of our system, our aim is to infer the scope of speculation. Therefore, our proposal detect hedge cues and annotate the scope in sentences. It is general and the rules used can be applied to different corpora.

In Section 2 we show the background of the present work. Section 3 describes the scope of speculation in the Bioscope corpus. Section 4 describes our system. In Section 5 we show the results and, finally, in Section 6 we give our conclusions and future work.

2 PREVIOUS WORK

In this section the state–of–the–art on related approaches about speculation detection and dependency parsing is briefly described.

2.1 Speculation Detection and the Scope of Speculation Problem

This Section presents systems that infer the scope of speculation and predict which words of a sentence are inside or outside the scope.

The recent CoNLL' 2010 shared task was devoted to systems that infer the scope of speculation (Farkas et al., 2010), using data–sets from the Wikipedia and a part of Biological publications of the Bioscope corpus and other scientific publications. This shared task is a starting point to evaluate systems that infer the scope of speculation. One of the main conclusions of the shared task is that dependency syntactic parsing is useful to achieve higher results when it is used instead of other technologies. For instance, Velldal et al (Velldal et al., 2010) adopted some heuristic rules from dependency parsed trees to infer the scope of speculation.

Morante and Daelemans published a machine learning approach for the biomedical domain (Morante and Daelemans, 2009). The system was evaluated with the Bioscope corpus and their results were 61.51% recall and 83.37% precision.

Zhu et al. applied their framework to speculation using shallow semantic parsing (Zhu et al., 2010). They evaluated it with the Bioscope corpus using hedge cues. Their results, without using golden cues (which means that their system does not need to find where is the cue and which one is it) were 62.54% recall and 76.55% precision.

The work done by Özgur and Radev, show an interesting approach employing some heuristic rules from constituency parse tree perspective on speculation scope identification (Özgür and Radev, 2009). They obtain an accuracy of 79.89% and 61.13% using golden cues and the abstracts and full papers subcollections of Bioscope.

Additionally, Aggarwal and Yu used Conditional Random Fields (CRF) to infer the scope of speculation (Agarwal and Yu, 2010). We consider that their results are not comparable with the other systems of the state of the art because they take into account the whole corpus to measure it. Thus, any sentence without any scopes tagged (that is a frequent situation in the Bioscope corpus (see Section 3), counts as a perfect annotated scope.

Finally, it is important to mention that Morante et al also adapted their system to participate in the CoNLL' 2010 task (Morante et al., 2010).

2.2 Dependency Parsing

The basic idea of Dependency Parsing is that syntactic structure consists of lexical items, linked by binary asymmetric relations called dependencies. A dependency structure for a sentence is a labeled directed tree, composed of a set of nodes, labeled with words, and a set of arcs labeled with dependency types (Nivre, 2006).

We selected Minipar (Lin, 1998). to develop our approach, and this decision was because of four main reasons:

- Regarding precision, Minipar is a state–of–the–art dependency parser.
- It is a domain independent dependency parser.
- It does not need any training. It is a rule–based approach, so we do not need to depend on any training corpora specifically developed with sentences from a concrete domain.
- It does not need a lemmatizer or a part–of–speech tagger to pre–annotate the testing sentences and training corpora.

Moreover, we tested Minipar manually with sentences from Bioscope and it worked well. To infer the scope of speculation it is not the domain what we care for this work, it is the hedge syntactic structures and how well are the hedge cues attached in the dependency tree.

3 THE SCOPE OF SPECULATION IN THE BIOSCOPE CORPUS

Bioscope (Szarvas et al., 2008) is an open access corpus, annotated manually with the scope of speculation for the biomedical domain.

The Bioscope corpus contains more than 20,000 sentences, divided in three different collections, as shown in Table 1. The documents inside the Bioscope collections are annotated not only with speculations, but also with negation. Table 1 shows the number of documents, sentences, speculation sentences and hedge cues for each collection and the percentage of scopes to the right and to the left in the Bioscope corpus, considering only sentences with speculation.

Table 1: The statistics of the Bioscope corpus considering only speculations

Collection	Clinical	Papers	Abstracts
Documents	1,954	9	1,273
Sentences	6,383	2,670	11,871
% Hedge Sentences	13.39	19.44	26.43
Hedge Cues	1,189	714	2,769
%Scopes to the right	73.28	76.55	82.45
%Scopes to the left	26.71	23.44	17.54

In Bioscope all the scopes include a cue, but, it is worth emphasizing that when the scope is opened at the cue and continues to the right of the cue (Scopes to the right in Table 2), the scope affected by the cue leaves the subject out. This correspond to sentences in active voice and they are the most frequent case. Additionally, there are some cases in which the scope is opened to the left of the cue (Scopes to the left in Table 2). The most frequent one is the passive voice. As shown in (Szarvas et al., 2008), passive voice is an exception in the way of tagging sentences in Bioscope. In this case the subject is marked within the scope, because if the sentence had been written in active voice, it would be the object of a transitive verb.

4 SPECULATION SCOPE FINDING APPROACH BASED ON DEPENDENCY ANALYSIS

When studying how to develop an algorithm that detects wordforms within the scope of speculation, we found that dependency parsing could be very useful, because it allows to consider which nodes depend on others and leads to detect which nodes are affected by speculation. Our system traverses the dependency tree, searching for hedge cues to determine the correct scope of them over the tree. Our contribution lies in the identification of the scope, which is not explicit in the dependency tree. Therefore, we can consider the nodes that shared the same branch with a hedge signal or which nodes directly depend on a hedge signal. Additionally, if we run through the tree until we find terminals, we can find the wordforms deepest in the tree structure that depend on, or are related to a hedge signal that infers the scope of the cues.

A parsing given by Minipar is the input for the Hedge Wordfoms Detection Algorithm described in Section 4.2, which returns the set of wordforms within the scope of speculation. Then, the Scope Finding Algorithm described in Section 4.3 acts on that set using the passive voice module, returning an annotated sentence.

Following, we describe the Speculation Cue Lexicon used in our system, the algorithm that detects the wordforms within the scope of speculation and, finally, the Scope–Finding Algorithm that finishes the task.

4.1 Speculation Cue Lexicon

To determine the scope of speculation, first of all a set of hedge cues must be established. We classified the hedge cues that are considered in our biomedical system obtained from the Bioscope corpus.

We show an excerpt of the lexicon considered for our system configuration for the Biomedical domain in Table 2. The lexicon only shows the lemmas of each wordform but our system is able to parse not only the lemma, but all kind of verb forms, such as third person, past tense, etc.

4.2 Hedged Wordforms Detection Algorithm

We implemented an algorithm that takes the dependency tree for a sentence returned by Minipar, and returns the hedge cue and a set with the words affected by the cue.

Table 2: Lemmas for the Hedge Cues of the Biomedical lexicon considered in our Biomedical system.

appear	can	could	either
indicate that	indicate	imply	evaluate for
likely	may	might	or
possible	possibly	potential	potentially
propose	putative	rule out	suggest
think	unknown	whether	would

The algorithm runs through the dependency tree of a sentence and does the following steps:

1. It detects all the nodes that are contained in the speculation cue lexicon.
   - If the node is an auxiliary verb and it is affected directly by a cue, the algorithm marks the verb affected by it as a cue. This is because the words that are affected by this cue depends on the verb.
   - If the cue is a different wordform, contained in the lexicon, it is marked as a cue.
2. For the rest of nodes, if a node directly depends on any of the ones previously marked as a cue, the system marks it as affected. Moreover, the detection is propagated from the cue word through the dependency graph until it finds terminals, so wordforms that are not directly related with the cue are detected too.

4.3 Scope Finding Algorithm

This algorithm uses the set of words returned by the Affected Wordforms Detection Algorithm, described in the previous Subsection, and the dependency tree given by Minipar. This second–step algorithm returns sentences annotated with the scopes of cues, inferring where to open a scope and where to close it.

Where to open a scope is related to the voice of the sentence: 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. So, the first step of this algorithm is to determine the voice of the sentence.

Thus, we considered that the Scope Finding Algorithm must be divided into two main processes: first, to detect if the sentence follows a passive voice structure or not, and second, to annotate the sentence with the scope of the cue considered in the lexicon, or scopes if there is more than one (which is a common situation when there is more than one cue).

A sentence is in passive voice if:

- It contains a transitive verb, such as, *show, consider, see, use, detect*, etc.

- It follows this pattern[1]: *modal verb + be + past participle.*

Once our system has decided if the sentence is in passive voice or not, the Scope Finding Algorithm iterates the sentence, token by token, and applies a set of rules about scope opening and closing. Only one rule is applied for each token.

1. Scope opening:

   a. If the token is contained in the set of nodes marked as affected by the Affected Wordforms Detection Algorithm and the scope for the cue involved is not open: the system opens the scope at the token and establishes that the scope for the cue involved is already opened.

   b. If the token is a cue (contained in the lexicon) and the sentence is in passive voice: the system goes backward and opens the scope just before the subject of the sentence. The system opens and closes the cue at the token.

   c. If the token is a cue and the sentence is not in passive voice: the system opens the scope just before the token. The system opens and closes the cue at the token.

2. Scope closing:

   a. If the token is a punctuation symbol, followed by some wordforms that indicate another statement, such as *but*: the system closes the scope just after the token.

   b. If the token is any wordform and all the nodes that are marked as affected by the hedge cue are already included in the scope: the system closes the scope just before the token.

   c. If the token is at the end of sentence: the system closes the scope at the end of the sentence.

3. Other case: if none of the previous rules has been applied the token is added to the annotated sentence.

At this point, the system has computed the scope (scopes) of the cue (cues) for a given sentence, by inferring which nodes pertain to that scope (scopes) from the node (nodes) marked as affected.

Thus, our system is able to parse sentences like the one shown in Figure 1 where we illustrate the text processing of a sentence from the Bioscope corpus.

5 EVALUATION

We tested the whole Hedge Scope Finding System with the three collections of Bioscope: the Scientific Papers Collection, the Abstracts Collection and the Clinical Radiology Reports Collection. In this Section, we discuss the evaluation design, the results obtained by our system and the discussion in which we compared our system with similar approaches.

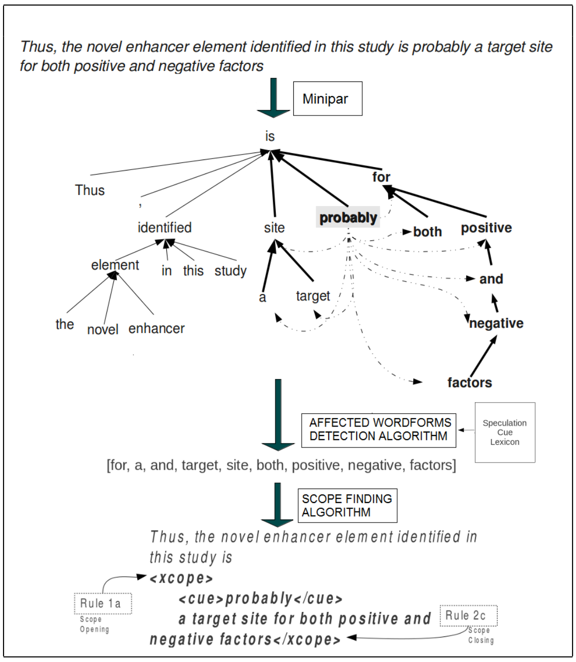

Figure 1: The processing of a sentence by our system. The rule applied to open the scope is *1a*, and the rule applied to close the scope is *2c*. These rules are described in the present Section.

5.1 Evaluation Design

Our first step was to select the sentences containing speculations in the three collections of Bioscope. In this way we only evaluated our system with 13.39% of the clinical sentences, 19.44% of the papers sentences, and 17.70% of the abstracts sentences. These are the percentage of hedge sentences for each collection.

We evaluated using Precision per token, Recall per token and to balance them, we used micro F1. Additionally, we evaluate our system with the percentage of correct scopes (PCS). Most of the systems of the state of the art, such as (Morante and Daelemans, 2009), used this metric as well. We also decided to evaluate it with the percentage of correct hedge cues (PCHC). Both of them are recall measures.

By using all these measures we are considering not only a token–based evaluation but a whole scope classification measure. Also, PCHC gives a measure for the failures of the system when predicting hedge cues, which cause a decrease in PCS.

[1] We only consider modal verbs because it is what we care for the Speculation Scope problem.

5.2 Results and Discussion

In this Section we show the results given by our system. As shown in Table 3 the results vary, depending on the collection used to evaluate. It is worth emphasizing that we did scope identification with automatic cue recognition, so the input of our system is the sentence without any extra information. Therefore, it means that we do not use neither golden cues (which means that the system did not need to find where was the cue and which one was it) nor golden trees (which means that the tree is given and it is certainly correct).

Table 3: Results of our Speculation Scope Finding System, evaluated with the three Collections of Bioscope

Collection	Precision	Recall	F1	PCS	PCHC
Papers	82.78%	73.88%	78.08%	39.43%	80.38%
Abstracts	87.96%	75.35%	81.14%	46.75%	79.50%
Clinical	83.96%	67.15%	74.62%	36.20%	67.19%
Average	86.70%	74.62%	79.54%	43.96%	77.26%

As can be observed in Table 3, one of the main problems is to correctly detect the hedge cue. One of the main reasons is that there are some hedge cues that are not always considered as cues. For instance, the wordform *can* is not commonly used as a hedge cue in the papers collection (just 31.57%) but it is more frequent as a hedge cue in the abstracts collection. Therefore we found that hedge cue classification is a really difficult task, as some cues are not always used as hedge cues. In systems like ours we need to decide which cues are included and which ones not, so mistakes in this decision may result in a loss of accuracy because the scopes of the speculative sentences that contain these non common hedge cues are not correctly annotated.

5.3 Comparison with the State of the Art Systems

In this Section we compared the results of our system with the best state of the art systems (Morante and Daelemans, 2009) and (Zhu et al., 2010). The main comparison is shown in Table 4.

Our system does not need any training, so we did our test with all the corpus. Morante et al. performed 10-fold cross validation experiments with the abstracts collection. For the other two collections, they trained with the abstracts set and they tested with the corresponding collection. We can also show that the results obtained by them for the abstracts collection are very high if we compare our results for the other collections. This is probably because they trained their system with the abstracts collection. The

Table 4: Results of our work, evaluated with the three collections of Bioscope and compared with the systems of Morante et al., Zhu et al.

Collection	System	Precision	Recall	F1	PCS	PCHC
Papers	Our Results	**82.78%**	**73.88%**	**78.08%**	**39.43%**	80.38%
	Morante et al.	67.97%	53.16%	59.66%	35.92%	**92.15%**
	Zhu et al.	56.27%	58.20%	57.22%	–	–
Abstracts	Our Results	**87.96%**	**75.35%**	**81.14%**	46.75%	79.50%
	Morante et al.	85.77%	72.44%	78.54%	**65.55%**	**96.03%**
	Zhu et al.	81.58%	73.34%	77.24%	–	–
Clinical	Our Results	**83.96%**	**67.15%**	**74.62%**	**36.20%**	**67.19%**
	Morante et al.	68.21%	26.49%	38.16%	26.21%	64.44%
	Zhu et al.	70.46%	25.59%	37.55%	–	–
Average	Our Results	**86.70%**	**74.62%**	**79.54%**	43.96%	72.26%
	Morante et al.	83.37%	61.51%	68.71%	**54.68%**	**89.58%**
	Zhu et al.	76.55%	62.54%	67.41%	–	–

ways of annotating hedge scopes in the abstracts collection and the clinical reports collection are really different, which leads to a loss of accuracy in these cases. The Scientific Papers collection is more similar, but there are some infrequent cues in the Abstracts collection that appear in the Scientific Papers collection, like *would*.

As we can observe in the results for the Clinical Reports Collection, the differences are greater than in the other cases for the recall measure. The results of Morante et al. system mean that considering their precision results, their system correctly classifies most of the tokens. But regarding recall, their system detects very few tokens. For a system that includes all the correct tokens except one for each scope, the precision and recall measures would be very high, but the PCS measure would be zero. This means that our system leaves out some of the tokens for each scope out, but most of the tokens are correctly included. We can conclude that their results are completely derived from the fact that they train the models using the abstracts collection. As a result, this factor deeply affects the recall in the Clinical Reports collection because it contains somewhat different hedge cues and, more important, uses them in a different way. Nevertheless, for us, the problem is not as deep as their case, because we used a configurable lexicon of wordforms which is the same for the three collections and includes all the wordforms that appear in the whole corpus.

In Table 5 we show the percentage of correct scopes (PCS) per speculation cue, for hedge cues that occur 20 or more times in one of the subcorpora compared with Morante et al.

The differences in the PCS measure (percentage of correct scopes) show that their system correctly annotates more scopes than ours, but our results in Precision and Recall show that we classified more correct tokens within the scope of speculation. Morante et al. used machine–learning to predict the correct

Table 5: PCS per hedge cue for hedge cues that occur 20 or more times in one of the subcorpora. Comparing the results of our system (Ours) with the results of Morante et al.' system (Mor.). The column annotated as # shows the number of appearances for each case.

	Abstracts			Papers			Clinical		
	#	**Mor.**	**Ours**	#	**Mor.**	**Ours**	#	**Mor.**	**Ours**
appear	143	**58.04**	18.88	39	**28.20**	12.82	–	–	–
can	48	12.15	**45.83**	25	0.00	**24.00**	22	0.00	**27.27**
consistent with	–	–	–	–	–	–	67	0.00	**46.29**
could	67	11.94	**34.33**	28	14.28	**46.43**	36	22.22	**33.33**
either	28	0.00	0.00	–	–	–	–	–	–
evaluate for	–	–	–	–	–	–	86	**3.84**	0.00
imply	21	**90.47**	0.00	–	–	–	–	–	–
indicate	23	73.91	**86.21**	–	–	–	–	–	–
indicate that	276	**89.49**	47.32	–	–	–	–	–	–
likely	59	**59.32**	42.37	36	30.55	**36.11**	63	**66.66**	60.32
may	516	**81.39**	44.96	68	54.41	**55.88**	107	**80.37**	39.25
might	72	73.61	27.78	40	**35.00**	25.00	–	–	–
or	120	0.00	**13.33**	–	–	–	276	0.00	**26.09**
possible	50	**66.00**	34.00	24	**54.16**	29.17	26	80.76	**100.0**
possibly	25	**52.00**	24.00	–	–	–	–	–	–
potential	45	28.88	**40.00**	–	–	–	–	–	–
potentially	21	**52.38**	38.10	–	–	–	–	–	–
propose	38	**63.15**	15.79	–	–	–	–	–	–
putative	39	17.94	**28.20**	–	–	–	–	–	–
rule out	–	–	–	–	–	–	61	0.00	**24.59**
suggest	613	**92.33**	32.62	70	**62.85**	30.0	64	**90.62**	59.38
think	35	31.42	0.00	–	–	–	–	–	–
unknown	26	15.38	0.00	–	–	–	–	–	–
whether	96	**72.91**	23.96	–	–	–	–	–	–
would	–	–	–	21	28.57	28.57	–	–	–

hedge cue, while we only have a lexicon.

6 CONCLUSIONS AND FUTURE WORK

The potential of an accurate speculation scope finding system is undeniable. This papers presents a high performance system able to infer the scope of speculations. From the results of our experiments we can conclude that dependency parsing is a valuable auxiliary technique for speculation detection, at least in the particular case of English. We obtained similar results as the state–of–the–art systems without using machine learning, just using a rule–based approach with the help of an algorithm that runs through dependency syntactic structures.

As a suggestion for future work, we consider that the scope of speculation must not always be annotated as continuous. In Bioscope, the scope of speculation leaves normally the subject out (when the scopes are to the right), nonetheless, we consider that the subject must always be considered as a part of the scope. Thus, we suggest that the scope must be discontinuous in the way of considering other wordforms that in Bioscope are out of the scope, but are directly affected by the speculation cue.

Finally, it is worth to mention that the system can be accessed online at `http://minerva.fdi.ucm.es:8888/ScopeTaggerSpec`.

ACKNOWLEDGEMENTS

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

Agarwal, S. and Yu, H. (2010). Detecting hedge cues and their scope in biomedical text with conditional random fields. *Biomedical Informatics*.

Farkas, R., Vincze, V., Móra, G., Csirik, J., and Szarvas, G. (2010). The conll-2010 shared task: learning to detect hedges and their scope in natural language text. In *CoNLL '10: Shared Task*.

Lin, D. (1998). Dependency-based evaluation of MINIPAR. In *Proc. Workshop on the Evaluation of Parsing Systems*, Granada.

Morante, R. and Daelemans, W. (2009). Learning the scope of hedge cues in biomedical texts. In *BioNLP '09*.

Morante, R., Van Asch, V., and Daelemans, W. (2010). Memory-based resolution of in-sentence scopes of hedge cues. In *CoNLL '10: Shared Task*.

Nivre, J. (2006). *Inductive Dependency Parsing*. Text, Speech and Language Technology. Springer, Dordrecht.

Özgür, A. and Radev, D. R. (2009). Detecting speculations and their scopes in scientific text. In *EMNLP*.

Szarvas, G., Vincze, V., Farkas, R., and Csirik, J. (2008). The bioscope corpus: annotation for negation, uncertainty and their scope in biomedical texts. In *BioNLP*.

Velldal, E., Øvrelid, L., and Oepen, S. (2010). Resolving speculation: Maxent cue classification and dependency-based scope rules. In *CoNLL*.

Zhu, Q., Li, J., Wang, H., and Zhou, G. (2010). A unified framework for scope learning via simplified shallow semantic parsing. In *EMNLP '10*.