

Annotating Negation and Speculation: the Case of the Review Domain

Natalia Konstantinova and Sheila C. M. de Sousa

Research Group in Computational Linguistics

University of Wolverhampton

Stafford Street, Wolverhampton, WV1 1SB, UK

{n.konstantinova, sheila.castilhomonteirodesousa}@wlv.ac.uk

Abstract

The paper presents the annotation of negation and speculation which is important for many NLP applications. Unlike previous research focusing on the medical domain, we investigate the review domain and attempt to annotate the SFU Review Corpus. In order to guarantee consistent annotation, we develop specific guidelines. Given the lack of research into annotation in the review domain, we explore the possibility of adapting the existing BioScope guidelines for the domain of interest. In order to reveal cases that need additional investigation, initially a small part of the corpus was annotated and this information was used for developing the guidelines. The paper describes the general principles our guidelines are based on and discusses differences with those in BioScope. It discusses the cases which were difficult to annotate. We include some insight into future work in order to improve the annotation process as well.

1 Introduction

Identification of negation and speculation is a very important problem for a wide range of NLP applications, including but not limited to information extraction, text mining, opinion mining and textual entailment. For all of these tasks it is crucial to know when a part of the text should get e.g. the opposite meaning (in the case of negation) or should be treated as subjective and non-factual (in the case of speculation).

Speculation and negation are important aspects of language. Speculation is related to the broader concept of “modality” which has been extensively studied both in linguistics and philosophy (Saurí, 2008). Various classifications of modality can be found in literature (Morante and Daelemans, 2009). Related terms like “hedging”, “evidentiality”, “uncertainty”, and “factuality” are also used when talking about different aspects of modality. Saurí et al. (2006) state that modality “expresses the speaker’s degree of commitment to the events being referred to in a text”.

Negation is part of the broader concept of “polarity”, which indicates whether a statement is presented as positive or negative (Saurí, 2008). In simple propositional logic, negation is an operator that reverses the truth value of a proposition (Miestamo, 2007).

In defining speculation and negation we follow the definitions introduced by Vincze (2010): “speculation is understood as the possible existence of a thing is claimed – neither its existence nor its non-existence is known for sure”, so there is not enough evidence in the text to say whether information is true or not. Whereas “negation is seen as the implication of nonexistence of something”.

These two phenomena are interrelated (de Haan, 1997) and have similar characteristics in the text: they both have scope, so affect part of the text which is denoted by the presence of negation or speculation cue words.

The problem of treatment of negation and speculation is quite recent, but it is becoming more popular (more details can be found in Section 2). A large scale corpus is needed for training statistical algorithms to identify of these aspects of the language. However most of the work is done for the biomedical domain and general domain texts have not received much attention (Morante et al., 2011). To our knowledge there is no big corpus from the review domain annotated with negation and speculation. This motivated our work of annotation of the SFU Review Corpus (Taboada et al., 2006) which is widely used in the domain of sentiment analysis and opinion mining. Identification of speculation in reviews can help by providing a measure of the reliability of the opinion contained and can be used for opinion mining (e.g. as suggested in (Wilson et al., 2005)). Also there is no doubt that negation is important for this task as well, because the phrase “this movie is good” has completely different polarity from “this movie is not good”, even though they both contain the positive word “good”.

It was decided to use the currently available guidelines for the BioScope corpus (Vincze et al., 2008) and attempt to adapt them to the review domain.

The structure of the paper is the following: Sec-

tion 2 outlines related research, Section 3 describes the corpus used for the annotation and the annotation tool. Section 4 discusses the way the BioScope guidelines should be adapted to the review domain in order to take into account the peculiarities of another domain. The paper finishes with the conclusions and discussion of the directions of the future work (Section 5).

2 Related Work

The topic of negation and speculation became popular only recently, there are not a lot of works tackling this problem. The workshop organised at ACL 2010 (NeSp-NLP 2010)¹ was the key event to bring together researchers working on this problem. Also CoNLL-2010 Shared Task Learning to detect hedges and their scope in natural language text² contributed a lot to the development of this research topic.

Annotation of these phenomena was done at different levels ranging from words (Hassan and Radev, 2010) to whole events (Saurí, 2008). Just recently the idea of annotating keywords and scope was introduced by (Vincze, 2010; Kim et al., 2008).

There are several already annotated corpora: the GENIA Event corpus (Kim et al., 2008), which contains annotation of biological events with negation and two types of uncertainty. Medlock and Briscoe (2007) based their system on a corpus consisting of six papers from genomics literature, which were annotated for speculation. Settles et al. (2008) constructed a corpus where sentences were classified as either speculative or definite, however, no keywords were marked in the corpus.

The research community is trying to explore other domains and not only biomedical texts, so the CoNLL-2010 Shared Task on Hedge Detection (Farkas et al., 2010) included not only biomedical texts, but also Wikipedia articles, which were annotated for weasel words (“a word is considered to be a weasel word if it creates an impression that something important has been said, but what is really communicated is vague, misleading, evasive or ambiguous”).

As can be noticed most of the work was done for the biomedical domain and there are only now some attempts to annotate general texts like in (Councill et al., 2010). Morante et al. (2011) also discuss the need for corpora which cover other domains. The authors point out that existing guidelines should be adapted to new domains and mention that they are currently annotating texts by Conan Doyle.

We are aware of only one corpus in the review domain described in (Councill et al., 2010), however it

was annotated only for negation, but not speculation. Also this corpus is not big and contains only 2111 sentences in total, out of which 679 sentences contain negation.

There are several guidelines available for this task: guidelines for annotation of speculation in the biomedical domain can be found in (Light et al., 2004; Medlock, 2006) (however no cues are annotated there); partial guidelines for annotation of speculation and its keywords are presented in (Farkas et al., 2010). As mentioned earlier (Councill et al., 2010) provide some guidelines for annotation of negation. However the most detailed guidelines for both negation and speculation can be found for the BioScope corpus and are freely available online³.

3 Annotation Process

The aim of this research was to further study the problem of negation and speculation and to adapt the BioScope guidelines for the annotation of texts from the review domain. A small part of the corpus was initially annotated to provide a comparison of the domains and reveal cases that need to be treated differently in the review domain. The following sections will provide more information about the corpus and the annotation tool used for the task.

3.1 Corpus Description

The SFU Review corpus (Taboada et al., 2006) was chosen for our annotation of negation and speculation. As mentioned earlier, the choice of the corpus was motivated by the lack of annotated corpora for the review domain and also by the need for identification of these phenomena in this domain. This corpus consists of 400 reviews from the website Epinions.com. All the texts are split into several sections such as movies, music, books, hotels etc. Each text gets a label based on whether it is a positive or negative review. All the texts differ in size and are written by different people (more information about the size of the corpus can be found in Table 1).

The BioScope corpus (Vincze et al., 2008) consists of three different types of texts, which is done to ensure the heterogeneity of language used in the biomedical domain. It includes abstracts of the GENIA corpus, 9 full scientific articles and clinical free-texts (more information is provided in Table 2).

As can be seen from Tables 1 and 2 the amount of sentences in the SFU Review corpus is 16,705 and therefore the corpus is of comparable size with BioScope, which consists of more than 20,000 annotated sentences altogether (Vincze et al., 2008).

In the first stage of our work reported here we annotated 20% of the SFU Review corpus using the

¹Proceedings of the workshop can be found at: <http://aclweb.org/anthology-new/W/W10/#3100>

²Website: <http://www.inf.u-szeged.hu/rgai/coNll2010st/>

³Website: <http://www.inf.u-szeged.hu/rgai/bioscope>

Domain	#Sentences
Books	1,596
Cars	2,960
Computers	2,972
Cookware	1,473
Hotels	2,129
Movies	1,722
Music	2,817
Phones	1,036
Total	16,705

Table 1: Statistics of the SFU Review corpus

Subcorpora	#Documents	#Sentences
GENIA Abstracts	1,273	11,872
Full papers	9	2,624
Clinical free-texts	1,954	6,383
Total	3,236	20,879

Table 2: Statistics of three BioScope subcorpora

BioScope guidelines. 10 texts were taken from each of 8 domains described in Table 1 to ensure different kinds of texts are studied. This initial step of annotation was used to understand what cases cannot be covered by the BioScope guidelines and how these guidelines should be adapted to the review domain (more detailed discussion of this is presented in Section 4).

The next section will present the annotation tool which was used for our task.

3.2 Annotation Tool

To speed up annotation and ensure its consistency the annotation tool PALinkA (Orăsan, 2003) was used. It is a language- and task-independent tool which allows you to define your own link types. Users can benefit from its intuitive graphical interface which does not require complicated training and is easy to use. The output of this program is a valid xml document, which makes the following processing easier. And the users do not need any technical education, the tool itself prevents them from introducing mistakes into the xml file structure.

The tool allowed us to select keywords and annotate them as negation or speculation. Afterwards the scope was marked in the text and then linked to the cue it belonged to. Graphical interface does not show xml tags in the texts, but uses colours to denote the keywords and scope, which makes annotation representative and easy to analyse and correct if needed. When complex keywords, such as “either...or”, “neither...nor” were annotated, there was a possibility to link the scope to both keywords. The

use of this annotation tool made us introduce some additional changes to the annotation guidelines described in the next section.

4 Adaptation of Guidelines

Consistent and detailed guidelines are needed when annotating a corpus in order to avoid mistakes and to ensure consistency of the annotation. We attempted to adapt the existing BioScope guidelines in order to fit the needs of the review domain. The BioScope guidelines consist of two parts: speculation and negation. Each part provides information about the marking schemes, the keywords used and the scopes to be annotated. The authors attempted to provide an extensive description of all different cases and also give examples illustrating their rules.

To illustrate examples of the annotation process we use the keywords in bold and their types in subscript; we use () to indicate the scope of speculative keywords; and [] to indicate the scope of negative keywords.

4.1 Main Principles

The BioScope guidelines are based on four main principles (Vincze, 2010):

- Each keyword has a scope.
- The scope must include its keyword.
- Min-max strategy.
 - The minimal unit expressing hedge/negation is marked as the keyword.
 - The scope is extended to the maximal syntactic unit.
- No intersecting scopes are allowed.

There are several principles we also try to follow in order to make annotation consistent:

Min-max strategy: We follow the min-max strategy suggested before in (Vincze, 2010; Farkas et al., 2010). When annotating cues, we try to choose the minimal unit which expresses negation or speculation. In this situation special attention should be paid to distinguishing complex cues and sequences of several keywords. However when annotating scope we try to annotate the maximum words affected by the phenomenon:

*They ended up hitting me in the nuts, which, to say the least, was **probably**_{spec}(better than what the director of this film did to the memory of Dr.Seuss).*

Negation scope: Similar to the BioScope guidelines for the negation scope, only the words that are modified by the negation cue are included in the scope:

*It **isn't**_{neg} [scary], but it is enthralling.*

Elliptic sentences: For elliptic sentences the keyword is marked and the scope is neglected:

The Bioscope guidelines provide an example of such a case:

*This decrease was seen in patients who responded to the therapy as well as in those who did **not**_{neg}.*

When annotating the SFU Review corpus we follow the strategy suggested in the BioScope guidelines:

*I later discovered that my 11 year old understood all of them. I wish he **hadn't**_{neg}.*

Complex keywords: We also follow the principles of the Bioscope guidelines when annotating complex keywords. When speculation or negation is expressed through a phrase rather than a single word and these words cannot express speculation separately, they are annotated as complex keywords:

*I **have a feeling**_{spec} (that many readers would have given up before the end due to boredom, frustration or the maddening feeling of ' What the hell is Patterson thinking when he wrote this?').*

In this case, *have a feeling* could be substituted by (I) *think* which clearly expresses uncertainty. However the words *have, a, feeling, that* cannot express uncertainty on their own.

4.2 Differences with BioScope

Some differences between the BioScope guidelines and ours are presented in this Section.

Keywords: Unlike the Bioscope corpus, where the cue words are annotated as part of the scope, for the SFU corpus we decided not to include the cue words in the scope.

The choice of the annotation tool was one of the reasons why the keywords were not included in the scope. When using PALinkA the annotation is done more easily and more intuitively if instead of including the keywords in the scope, we link the scope to the keyword it belongs to, while making it possible to have embedded scopes for different keywords. Therefore the resulting xml file is easier to read as one could have the same scope linked to different keyword IDs.

Scope: When the annotator is unsure of the scope of a keyword only the keyword should be annotated.

Type of keyword: When the annotator is unsure what type the keyword should be assigned to (whether it expresses negation or speculation), nothing should be annotated.

For these last cases we set up an 'undecided' category. Those cases will additionally be discussed and annotated at the next stage.

Coordination: The Bioscope guidelines suggest extending the scope for speculation and negation

keywords to all members of the coordination. However in the case of the review domain as the keywords were not included in the scope, the scopes were annotated separately and then linked to the keywords:

*As far as I remember , vacation with accommodation in (Rio), (Golden Nugget),(Excalibur) **or**_{spec} (Las Vegas Hilton) were available for cheaper rates than what I paid for Riviera.*

Embedded scopes: Although keywords are not included in their own scope, a keyword can be included in the scope of other keywords and situations of embedded scopes are possible:

*I'm **not sure**_{spec} (**if**_{spec} (he **should**_{spec} ((be angrier at his widow for giving studios the rights to his stories), **or**_{spec} (to the studios for stabbing his widow in the back when she trusted them))))).*

There were also cases when the combination of different types of keywords (ie. negation and speculation ones) resulted in the embedded scopes:

*It **isn't**_{neg} [(vulgar) **or**_{spec} (sexual)]*

It should be noted that while the scope for the keyword **or**_{spec} should include (vulgar) and (sexual), the scope for the keyword **isn't**_{neg} should include [vulgar or sexual]. It is explained by the fact that *isn't* modifies both coordinations, and should be understood as 'it isn't vulgar and it isn't sexual either'.

No scope: Unlike the BioScope guidelines which mention only the cases of negation keywords without scope, situations where speculation keywords had no scope were encountered as well in the review domain:

*This movie didn't have anything to do with a children's movie as it **should**_{spec}.*

4.3 Problematic Cases

While annotating the review domain using the BioScope guidelines we had to face some problematic cases of annotation that had to be discussed additionally.

Differences of the domains: First of all, we had to consider the differences between both domains (biomedical and review) to be able to adapt the guidelines properly. While the BioScope corpus consists of professional biomedical writings and thus a reliable source of texts, in the review domain we are likely to find ungrammatical sentences and misspellings. In the review domain it is not uncommon to find words such as 'ain't', 'whatcha', etc. Also the vocabulary of the domains is different and therefore different words can be considered as cues of negation or speculation. We had to take these peculiarities into account both when developing the guidelines and annotating the corpus.

Titles: As we are dealing with review texts, a great number of them include titles of the books or songs or even quotations from them which authors were referring to. Therefore it was not unusual to find sentences which contain the name of a song/book such as:

Ludacris spits fluidly on “it wasn’t us”

and

When ya came in the party and you saw the crowd shoulda read the sign, ‘no suckas allowed’

We believe that even when these sentences contain a cue word they should not be annotated because they do not express the writer’s uncertainty or negation.

Keyword sequences: The presence of the sequences of the keywords created additional difficulties for the annotation. We feel that the nature of the review domain texts introduces a greater possibility of encountering such cases than in the biomedical domain. Therefore special care should be taken when distinguishing several keywords that go one after another. Although some examples of two or more keywords in a sequence could be also considered as complex keywords they should be annotated separately if they can express hedge on their own:

I didn’t_{neg} [think_{spec} (it would_{spec} (be possible_{spec} (for anyone to rip the heart out of a Dr. Seuss book)))].

In this example the keywords *didn’t* and *think* may seem complex keywords but they should be annotated as separate keywords since *didn’t* negates *think* which is the leading cue of the whole idea of speculation.

Not sure: Also it was noted that the case of the keyword *not sure* can be difficult for annotation as its scope should include all the elements it modifies, for instance, it should include all the elements on the right in the following example:

not sure_{spec} (if he should be angrier at his widow for giving studios the rights to his stories, or to the studios for stabbing his widow in the back when she trusted them).

Great number of keywords: Close attention should be paid to sentences with a great number of keywords, which can lead the annotator to make mistakes. One of these difficult cases is presented below as an illustration:

This creative re-engineering draws (the viewer)¹ or_{1spec} (reader)¹ into a parallel universe where age-old lessons can_{spec} ((be taught)² or_{2spec} (re-taught)²) without_{neg} [(the obstructions created in the minds)^{3,4,5}, or_{3spec} (interferences)^{3,4,5}, or_{4spec} (misconceptions)^{3,4,5} if_{spec} (you prefer), or_{5spec} even (pre-concepts)^{3,4,5}] that may_{spec} (probably_{spec}

(lead to misunderstandings)).

While for the keywords **or_{1spec}** and **or_{2spec}** the scopes are easily identified, for the **or_{3,4,5spec}** the scopes are tricky since they should include all the members modified by the keyword *not* even if these members are syntactically distant from the keywords.

Passive voice: The case of the passive voice turned out to be a difficult one and generated a lot of discussions. As Morante et al. (2011) noticed there are some inconsistencies in the way the BioScope guidelines describe this problem. Therefore additional discussions and more studies are needed to decide how to mark the scope in sentences containing the passive voice. Therefore at the initial stage of annotation it was decided to mark these cases with a special label ‘undecided’. However in the final version of the guidelines we are planning to describe the ways to treat the passive voice and also correct the annotation accordingly.

As can be noted, the examples of the difficult cases of the annotation presented in this Section reveal once again the need for more detailed and specific guidelines for the review domain.

5 Conclusions and Future Work

A lot of work in the field of negation and speculation was done for the biomedical domain, but there is a need for studies in other domains. In this work we attempted to study the review domain and the ways the BioScope guidelines can be adapted to this domain. The research showed the need for detailed guidelines, however we understand that they cannot account for all possible cases in the corpus and therefore difficult cases should be discussed by several annotators.

We made an initial attempt of annotation of the SFU Review Corpus and annotated 20% of the corpus, this information was used for studying the differences of the review and biomedical domains and developing the guidelines for the review domain. We provided analysis of the ways the BioScope guidelines can be adopted to the review domain and what cases should be additionally discussed. We are planning to use the created guidelines to annotate the whole SFU Review Corpus. Also several annotators will be involved in this process and that will allow us to calculate the inter-annotator agreement. Based on this information we will refine the guidelines if needed and correct the annotation. Once this is done, we are planning to make both corpus and guidelines publicly available. We hope that this corpus will be helpful for further development of negation and speculation detection.

We are also planning to analyse the differences of speculation and negation cues in different domains and get more insight into the differences of the

review domain and that of biomedical texts.

Acknowledgements

We would like to thank the reviewers for their valuable comments, which helped us a lot in improving the paper. We are also grateful to Prof. Maite Taboada and Prof. Ruslan Mitkov for their support of our research. We wish to thank our colleagues Alison Carminke, Noa P. Cruz Díaz, Dr. Constantin Orăsan and Wilker Aziz for their help with various aspects of our work.

References

- Isaac Councill, 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, Sweden, July. University of Antwerp.
- Ferdinand de Haan. 1997. *The interaction of modality and negation: a typological study*. Garland Publishing, New York, USA.
- Richárd Farkas, Veronika Vincze, György Móra, János Csirik, and György Szarvas. 2010. The CoNLL-2010 shared task: learning to detect hedges and their scope in natural language text. In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning — Shared Task, CoNLL '10: Shared Task*, pages 1–12, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Ahmed Hassan and Dragomir R. Radev. 2010. Identifying text polarity using random walks. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 395–403, Uppsala, Sweden, July. Association for Computational Linguistics.
- Jin-Dong Kim, Tomoko Ohta, and Jun'ichi Tsujii. 2008. Corpus annotation for mining biomedical events from literature. *BMC Bioinformatics*, 9.
- Marc Light, Xin Ying Qiu, and Padmini Srinivasan. 2004. The language of bioscience: Facts, speculations, and statements in between. In Lynette Hirschman and James Pustejovsky, editors, *HLT-NAACL 2004 Workshop: BioLINK 2004, Linking Biological Literature, Ontologies and Databases*, pages 17–24, Boston, Massachusetts, USA, May 6. Association for Computational Linguistics.
- Ben Medlock and Ted Briscoe. 2007. Weakly supervised learning for hedge classification in scientific literature. In *Proceedings of the ACL*, pages 992–999, Prague, Czech Republic, June.
- Ben Medlock. 2006. Guidelines for speculative sentence annotation.
- Matti Miestamo. 2007. Negation an overview of typological research. *Language and Linguistics Compass*, 1(5):552–570, September.
- Roser Morante and Walter Daelemans. 2009. Learning the scope of hedge cues in biomedical texts. In *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing, BioNLP '09*, pages 28–36, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Roser Morante, Sarah Schrauwen, and Walter Daelemans. 2011. Corpus-based approaches to processing the scope of negation cues: an evaluation of the state of the art. In *Proceedings of the Ninth International Conference on Computational Semantics, IWCS '11*, pages 350–354, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Constantin Orăsan. 2003. PALinkA: a highly customizable tool for discourse annotation. In *Proceedings of the 4th SIGdial Workshop on Discourse and Dialog*, pages 39 – 43, Sapporo, Japan, July, 5 -6.
- Roser Saurí, Marc Verhagen, and James Pustejovsky. 2006. Annotating and recognizing event modality in text. In *In The 19th International FLAIRS Conference, FLAIRS 2006*.
- Roser Saurí. 2008. *A factuality profiler for eventualities in text*. Ph.D. thesis, Waltham, MA, USA.
- Burr Settles, Mark Craven, and Lewis Friedland. 2008. Active learning with real annotation costs. In *Proceedings of the NIPS Workshop on Cost-Sensitive Learning*, pages 1–10.
- Maite Taboada, Caroline Anthony, and Kimberly Voll. 2006. Methods for creating semantic orientation dictionaries. In *Proceedings of 5th International Conference on Language Resources and Evaluation (LREC)*, pages 427–432, Genoa, Italy, May.
- Veronika Vincze, György Szarvas, Richárd 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+.
- Veronika Vincze. 2010. Speculation and negation annotation in natural language texts: what the case of BioScope might (not) reveal. In *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*, pages 28–31, Uppsala, Sweden, July. University of Antwerp.
- Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, HLT '05*, pages 347–354, Stroudsburg, PA, USA. Association for Computational Linguistics.