

# Exploring Surface-level Heuristics for Negation and Speculation Discovery in Clinical Texts

**Emilia Apostolova**

DePaul University  
Chicago, IL USA

emilia.aposto@gmail.com

**Noriko Tomuro**

DePaul University  
Chicago, IL USA

tomuro@cs.depaul.edu

## Abstract

We investigate the automatic identification of negated and speculative statements in biomedical texts, focusing on the clinical domain. Our goal is to evaluate the performance of simple, Regex-based algorithms that have the advantage of low computational cost, simple implementation, and do not rely on the accurate computation of deep linguistic features of idiosyncratic clinical texts. The performance of the NegEx algorithm with an additional set of Regex-based rules reveals promising results (evaluated on the BioScope corpus). Current and future work focuses on a bootstrapping algorithm for the discovery of new rules from unannotated clinical texts.

## 1 Motivation

Finding negated and speculative (hedging) statements is an important subtask for biomedical Information Extraction (IE) systems. The task of hedge detection is of particular importance in the sub-genre of clinical texts which tend to avoid unqualified negations or assertions.

Negation/Speculation discovery is typically broken down into two subtasks - discovering the negation/speculation cue (a phrase or a syntactic pattern) and establishing its scope. While a number of cue and scope discovery algorithms have been developed, high performing systems typically rely on machine learning and more involved feature creation. Deep linguistic feature creation could pose problems, as the idiosyncrasies of clinical texts often confuse off-the-shelf NLP feature generation tools (e.g. relying on proper punctuation and grammaticality). In addition, computationally expensive algorithms could pose problems for high-volume IE systems.

In contrast, simple Regex-based algorithms have demonstrated larger practical significance as

they offer reasonable performance at a low development and computational cost. NegEx<sup>1</sup> (Chapman et al., 2001), a simple rule-based algorithm developed for the discovery of negation of findings and diseases in discharge summaries, has been implemented in a number of BioNLP systems, including Metamap<sup>2</sup>, CaTIES<sup>3</sup>, and Mayo Clinic's Clinical IE System (Savova et al., 2008). In NegEx, a list of phrases split into subsets are used to identify cues and their corresponding scopes (token widows preceding or following the cues).

## 2 Method

Negation/Speculation in general English could be expressed by almost any combination of morphologic, syntactic, semantic, and discourse-level means. However, the scientific 'dryness' of the biomedical genre and clinical texts in particular, limits language variability and simplifies the task. We evaluated the performance of the NegEx algorithm on the BioScope corpus (Szarvas et al., 2008). BioScope corpus statistics are shown in Tables 1 and 2.

Corpus Type	Sentences	Documents	Mean Document Size
Radiology Reports	7520	1954	3.85
Biological Full Papers	3352	9	372.44
Biological Paper Abstracts	14565	1273	11.44

Table 1: Statistics of the BioScope corpus. Document sizes represent number of sentences.

Corpus Type	Negation Cues	Speculation Cues	Negation	Speculation
Rad Reports	872	1137	6.6%	13.4%
Full Papers	378	682	13.76%	22.29%
Paper Abstracts	1757	2694	13.45%	17.69%

Table 2: The percentage of speculative sentences (last column) is larger than the percentage of negated sentences.

We first evaluated the performance of an unmodified version of the NegEx algorithm on the task of cue detection (Table 3). Without any tuning or modifications, NegEx performed well on identifying negation cues across all documents, achiev-

<sup>1</sup> <http://code.google.com/p/negex/>

<sup>2</sup> ©The National Library of Medicine

<sup>3</sup> <http://caties.cabig.upmc.edu/Wiki.jsp?page=Home>

ing an F-score of 90% on the clinical texts. For the task of identifying speculation cues, we simply used the NegEx Conditional Possibility Phrase list (35 speculative cue phrases). The overall performance of this simplistic approach revealed poor results.

	TP	FP	FN	Precision	Recall	F-score
Negation						
Rad Reports	836	131	36	86.45	95.87	90.92
Full Papers	307	74	71	80.58	81.22	80.9
Paper Abstracts	1390	211	367	86.82	79.11	82.79
Speculation						
Rad Reports	62	1	1075	98.41	5.45	10.33
Full Papers	1	0	681	100.0	0.15	0.3
Paper Abstracts	0	5	2694	0.0	0.0	0

Table 3: NegEx performance on identifying Negation and Speculation Cues (non-exact boundary). (TP=true positive, FP=false positive, FN=false negative)

As shown in Figure 1, speculation cues exhibit wider variability and a rule matching only 35 phrases proved inefficient. To enrich the list of speculation cues, we used hedging cues from the FlySlip corpus of speculative sentences<sup>4</sup>. Without any synonym expansion or fine-tuning, the performance of speculation cue detection improved significantly as shown in Table 4, achieving an F-score of 86% on the clinical dataset<sup>5</sup>.

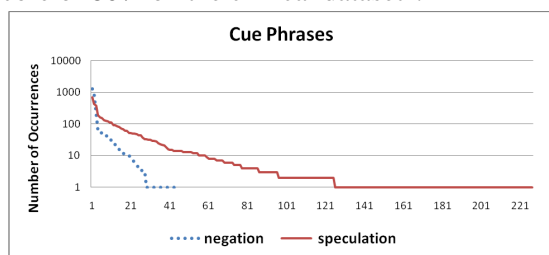


Figure 1: The number of occurrences (Y axis) of the 228 unique speculation cues and the 45 unique negation cues of the BioScope corpus (X axis).

Corpus	TP	FP	FN	Precision	Recall	F-score
Rad Reports	903	52	234	94.55	79.42	86.33
Full Papers	439	553	243	44.25	64.37	52.45
Paper Abstracts	1741	1811	953	49.01	64.63	55.75

Table 4: NegEx performance on identifying speculation cues (non-exact boundary) with the addition of the FlySlip hedging cues.

We next measured the performance of NegEx on scope detection. Newly introduced speculation cues from the FlySlip corpus were automatically classified into preceding or following their scope based the position of of their annotated ‘topic’. Table 5 shows the results of scope identification.

### 3 Discussion

Our results show that a simple, surface-level algorithm could be sufficient for the task of negation

<sup>4</sup><http://www.wiki.cl.cam.ac.uk/rowiki/NaturalLanguage/FlySlip/Flyslip-resources>

<sup>5</sup>To avoid fine-tuning cues on the corpus we did not set aside a training subset of the BioScope corpus for speculation cue enhancements and instead used an independent hedging corpus (FlySlip).

	TP	FP	FN	Precision	Recall	F-score
Negation						
Rad Reports	4003	267	140	94.12	97.61	95.18
Full Papers	2129	1835	525	54.45	80.12	64.01
Paper Abstracts	10049	6023	1728	63.04	85.13	72.31
Speculation						
Rad Reports	2817	1459	2471	65.87	53.27	58.90
Full Papers	3313	2372	2958	58.27	52.83	55.41
Paper Abstracts	17219	6329	9477	73.12	64.50	68.54

Table 5: NegEx performance on identifying scopes of correctly identified cues. Precision and recall are computed based on the number of correctly identified scope tokens excluding punctuation (i.e. number of tokens within cue scopes). Best results were achieved with no scope window size (i.e. using sentence boundaries).

and hedge detection in clinical texts. Using the NegEx algorithm and the FlySlip hedging corpus, without any modifications or additions, we were able to achieve an impressive F-score of 90.92% and 86.33% for negation and speculation cue discovery respectively<sup>6</sup>. We are currently expanding the set of speculation cues using an unannotated dataset of clinical texts and a bootstrapping algorithm (Medlock, 2008). The algorithm is based on the intuition that speculative cues tend to co-occur and this redundancy could be explored to probabilistically discover new cues from high-confidence existing ones. We are also exploring the discovery of degree of speculativeness (e.g. *very unlikely* vs *very likely*).

While NegEx performed well on the task of identifying negation scope (F-score 95.18), further work is needed on the discovery of speculation scopes (F-score 58.90). As hedging cues require a more fine-tuned set of rules, in future work we will evaluate linguistically motivated approaches (Kilicoglu and Bergler, 2008) for the creation of a set of surface-level speculation scope rules.

### References

- W.W. Chapman, W. Bridewell, P. Hanbury, G.F. Cooper, and B.G. Buchanan. 2001. A simple algorithm for identifying negated findings and diseases in discharge summaries. *Journal of biomedical informatics*, 34(5):301–310.
- H. Kilicoglu and S. Bergler. 2008. Recognizing speculative language in biomedical research articles: a linguistically motivated perspective. *BMC bioinformatics*, 9(Suppl 11):S10.
- B. Medlock. 2008. Exploring hedge identification in biomedical literature. *Journal of Biomedical Informatics*, 41(4):636–654.
- G.K. Savova, K. Kipper-Schuler, J.D. Buntrock, and C.G. Chute. 2008. UIMA-based Clinical Information Extraction System. In *Proc. UIMA for NLP Workshop. LREC*.
- G. Szarvas, V. Vincze, R. Farkas, and J. Csirik. 2008. The BioScope corpus: annotation for negation, uncertainty and their scope in biomedical texts. In *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing*, pages 38–45. Association for Computational Linguistics.

<sup>6</sup>The enhanced speculation cue phrase lists and a UIMA-based NegEx implementation are available upon request.