

# Recognizing Speculative Language in Biomedical Research Articles: A Linguistically Motivated Perspective

Halil Kilicoglu and Sabine Bergler

Department of Computer Science and Software Engineering  
Concordia University

Montreal, Quebec, Canada

{h\_kilico, bergler}@cse.concordia.ca

## Abstract

We explore a linguistically motivated approach to the problem of recognizing speculative language (“hedging”) in biomedical research articles. We describe a method, which draws on prior linguistic work as well as existing lexical resources and extends them by introducing syntactic patterns and a simple weighting scheme to estimate the speculation level of the sentences. We show that speculative language can be recognized successfully with such an approach, discuss some shortcomings of the method and point out future research possibilities.

## 1 Introduction

Science involves making hypotheses, experimenting, and reasoning to reach conclusions, which are often tentative and provisional. Scientific writing, particularly in biomedical research articles, reflects this, as it is rich in speculative statements, also known as *hedges*. Most text processing systems ignore hedging and focus on factual language (assertions). Although assertions, sometimes mere co-occurrence of terms, are the focus of most information extraction and text mining applications, identifying hedged text is crucial, because hedging alters, in some cases even reverses, factual statements. For instance, the italicized fragment in example (1) below implies a factual statement while example (2) contains two hedging cues (*indicate* and *might*), which render the factual proposition speculative:

- (1) Each empty cell *indicates that* the corresponding TPase query was not used at the particular stage of PSI-BLAST analysis.
- (2) These experiments indicated that the roX genes might function as nuclear entry sites for the assembly of the MSL proteins on the X chromosome.

These examples not only illustrate the phenomenon of hedging in the biomedical literature, they also highlight some of the difficulties in recognizing hedges. The word *indicate* plays a different role in each example, acting as a hedging cue only in the second.

In recent years, there has been increasing interest in the speculative aspect of biomedical language (Light et al., 2004, Wilbur et al., 2006, Medlock and Briscoe, 2007). In general, these studies focus on issues regarding annotating speculation and approach the problem of recognizing speculation as a text classification problem, using the well-known “bag of words” method (Light et al., 2004, Medlock and Briscoe, 2007) or simple substring matching (Light et al., 2004). While both approaches perform reasonably well, they do not take into account the more complex and strategic ways hedging can occur in biomedical research articles. In example (3), hedging is achieved with a combination of referring to experimental results (*We ... show that ... indicating*) and the prepositional phrase *to our knowledge*:

- (3) We further show that D-mib is specifically required for Ser endocytosis and signaling during wing development *indicating for the first time to our knowledge that* endocytosis regulates Ser signaling.

In this paper, we extend previous work through linguistically motivated techniques. In particular, we pay special attention to syntactic structures. We

address lexical hedges by drawing on a set of lexical hedging cues and expanding and refining it in a semi-automatic manner to acquire a hedging dictionary. To capture more complex strategic hedges, we determine syntactic patterns that commonly act as hedging indicators by analyzing a publicly available hedge classification dataset. Furthermore, recognizing that “not all hedges are created equal”, we use a weighting scheme, which also takes into consideration the strengthening or weakening effect of certain syntactic structures on lexical hedging cues. Our results demonstrate that linguistic knowledge can be used effectively to enhance the understanding of speculative language.

## 2 Related Work

The term *hedging* was first used in linguistic context by Lakoff (1972). He proposed that natural language sentences can be true or false to some extent, contrary to the dominant truth-conditional semantics paradigm of the era. He was mainly concerned with how words and phrases, such as *mainly* and *rather*, make sentences fuzzier or less fuzzy.

Hyland (1998) provides one of the most comprehensive accounts of hedging in scientific articles in the linguistics literature. He views hedges as polypragmatic devices with an array of purposes such as weakening the force of statement, expressing deference to the reader and signaling uncertainty. He proposes a fuzzy model, in which he categorizes scientific hedges by their pragmatic purpose, such as reliability hedges and reader-oriented hedges. He also identifies the principal syntactic realization devices for different types of hedges, including epistemic verbs (verbs indicating the speaker’s mode of knowing), adverbs and modal auxiliaries and presents the most frequently used members of these types based on analysis of a molecular biology article corpus.

Palmer (1986) identifies epistemic modality, which expresses the speaker’s degree of commitment to the truth of proposition and is closely linked to hedging. He identifies three types of epistemic modality: “speculatives” express uncertainty, “deductives” indicate an inference from observable evidence, and “assumptives” indicate inference from what is generally known. He focuses mainly on the use of modal verbs in expressing various types of epistemic modality.

In their investigation of event recognition in news text, Sauri et al. (2006) address event modality at the lexical and syntactic level by means of SLINKs (subordination links), some of which (“modal”, “evidential”) indicate hedges. They use corpus-induced lexical knowledge from TimeBank (Pustejovsky et al. (2003)), standard linguistic predicate classifications, and rely on a finite-state syntactic module to identify subordinated events based on the subcategorization properties of the subordinating event.

DiMarco and Mercer (2004) study the intended communicative purpose (dispute, confirmation, use of materials, tools, etc.) of citations in scientific text and show that hedging is used more frequently in citation contexts.

In the medical field, Friedman et al. (1994) discuss uncertainty in radiology reports and their natural language processing system assigns one of five levels of certainty to extracted findings.

Light et al. (2004) explore issues with annotating speculative language in biomedicine and outline potential applications. They manually annotate a corpus of approximately 2,000 sentences from MEDLINE abstracts. Each sentence is annotated as being definite, low speculative and highly speculative. They experiment with simple substring matching and a SVM classifier, which uses single words as features. They obtain slightly better accuracy with simple substring matching suggesting that more sophisticated linguistic knowledge may play a significant role in identification of speculative language. It is also worth noting that both techniques yield better accuracy over full abstracts than on the last two sentences of abstracts, in which speculative language is found to be more prevalent.

Medlock and Briscoe (2007) extend Light et al.’s (2004) work, taking full-text articles into consideration and applying a weakly supervised learning model, which also uses single words as features, to classify sentences as simply speculative or non-speculative. They manually annotate a test set and employ a probabilistic model for training set acquisition using *suggest* and *likely* as seed words. They use Light et al.’s substring matching as the baseline and improve to a recall/precision break-even point (BEP) of 0.76, using a SVM committee-based model from 0.60 recall/precision BEP of the baseline. They note that their learning models are unsuccessful in identify-

ing assertive statements of knowledge paucity, generally marked syntactically rather than lexically.

Wilbur et al. (2006) suggest that factual information mining is not sufficient and present an annotation scheme, in which they identify five qualitative dimensions that characterize scientific sentences: focus (generic, scientific, methodology), evidence (E0-E3), certainty (0-3), polarity (positive, negative) and trend (+,-). Certainty and evidence dimensions, in particular, are interesting in terms of hedging. They present this annotation scheme as the basis for a corpus that will be used to automatically classify biomedical text.

Discussion of hedging in Hyland (1998) provides the basic linguistic underpinnings of the study presented here. Our goals are similar to those outlined in the work of Light et al. (2004) and Medlock and Briscoe (2007); however, we propose that a more linguistically oriented approach not only could enhance recognizing speculation, but would also bring us closer to characterizing the semantics of speculative language. Some of the work discussed above (in particular, Sauri et al. (2006) and Wilbur et al. (2006)) will be relevant in that regard.

### 3 Methods

To develop an automatic method to identify speculative sentences, we first compiled a set of core lexical surface realizations of hedging drawn from Hyland (1998). Next, we augmented this set by analyzing a corpus of 521 sentences, 213 of which are speculative, and also noted certain syntactic structures used for hedging. Furthermore, we identified lexical cues and syntactic patterns that strongly suggest non-speculative contexts (“unhedgers”). We then expanded and manually refined the set of lexical hedging and “unhedging” cues using WordNet (Fellbaum, 1998) and the UMLS SPECIALIST Lexicon (McCray et al., 1994). Next, we quantified the strength of the hedging cues and patterns through corpus analysis. Finally, to recognize the syntactic patterns, we used the Stanford Lexicalized Parser (Klein and Manning, 2003) and its dependency parse representation (deMarneffe et al., 2006). We use weights assigned to hedging cues to compute an overall hedging score for each sentence.

To evaluate the effectiveness of our method, we used basic information retrieval evaluation metrics: precision, recall, accuracy and  $F_1$  score. In addition, we measure the recall/precision break-even point (BEP), which indicates the point at which precision and recall are equal, to provide a comparison to results previously reported. As baseline, we use the substring matching method, described in Light et al. (2004) in addition to another substring matching method, which uses terms ranked in top 15 in Medlock and Briscoe (2007). To measure the statistical significance of differences between the performances of baseline and our system, we used the binomial sign test.

### 4 Data Set

In our experiments, we use the publicly available hedge classification dataset<sup>1</sup>, reported in Medlock and Briscoe (2007). This dataset consists of a manually annotated test set of 1537 sentences (380 speculative) extracted from six full-text articles on *Drosophila melanogaster* (fruit-fly) and a training set of 13,964 sentences (6423 speculative) automatically induced using a probabilistic acquisition model. A pool of 300,000 sentences randomly selected from an archive of 5579 full-text articles forms the basis for training data acquisition and drives their weakly supervised hedge classification approach.

While this probabilistic model for training data acquisition is suitable for the type of weakly supervised learning approach they describe, we find that it may not be suitable as a fair data sample, since the speculative instances overemphasize certain hedging cues used as seed terms (*suggest*, *likely*). On the other hand, the manually annotated test set is valuable for our purposes. To train our system, we (the first author) manually annotated a separate training set of 521 sentences (213 speculative) from the pool, using the annotation guidelines provided. Despite being admittedly small, the training set seems to provide a good sample, as the distribution of surface realization features (epistemic verbs (32%), adverbs (26%), adjectives (19%), modal verbs (%21)) correspond roughly to that presented in Hyland (1998).

### 5 Core Surface Realizations of Hedging

---

<sup>1</sup> <http://www.benmedlock.co.uk/hedgeclassif.html>

Hyland (1998) provides the most comprehensive account of surface realizations of hedging in scientific articles, categorizing them into two classes: lexical and non-lexical features. Lexical features include modal auxiliaries (*may* and *might* being the strongest indicators), epistemic verbs, adjectives, adverbs and nouns. Some common examples of these feature types are given in Table 1.

Feature Type	Examples
Modal auxiliaries	<i>may, might, could, would, should</i>
Epistemic judgment verbs	<i>suggest, indicate, speculate, believe, assume</i>
Epistemic evidential verbs	<i>appear, seem</i>
Epistemic deductive verbs	<i>conclude, infer, deduce</i>
Epistemic adjectives	<i>likely, probable, possible</i>
Epistemic adverbs	<i>probably, possibly, perhaps, generally</i>
Epistemic nouns	<i>possibility, suggestion</i>

Table 1. Lexical surface features of hedging

Non-lexical hedges usually include reference to limiting experimental conditions, reference to a model or theory or admission to a lack of knowledge. Their surface realizations typically go beyond words and even phrases. An example is given in sentence (4), with hedging cues italicized.

(4) Whereas much attention has focused on elucidating basic mechanisms governing axon development, *relatively little is known* about the genetic programs required for the establishment of dendrite arborization patterns that are hallmarks of distinct neuronal types.

While lexical features can arguably be exploited effectively by machine learning approaches, automatic identification of non-lexical hedges automatically seems to require syntactic and, in some cases, semantic analysis of the text.

Our first step was to expand on the core lexical surface realizations identified by Hyland (1998).

## 6 Expansion of Lexical Hedging Cues

Epistemic verbs, adjectives, adverbs and nouns provide the bulk of the hedging cues. Although epistemic features are commonly referred to and analyzed in the linguistics literature and various

widely used lexicons exist that classify different part-of-speech (e.g., VerbNet (Kipper Schuler, 2005) for verb classes), we are unaware of any such comprehensive classification based on epistemological status of the words. We explore inducing such a lexicon from the core lexical examples identified in Hyland (1998) (a total of 63 hedging cues) and expanding it semi-automatically using two lexicons: WordNet (Fellbaum, 1998) and UMLS SPECIALIST Lexicon (McCray, 1994).

We first extracted synonyms for each epistemic term in our list using WordNet synsets. We then removed those synonyms that did not occur in our pool of sentences, since they are likely to be very uncommon words in scientific articles. Expanding epistemic verbs is somewhat more involved than expanding other epistemic types, as they tend to have more synsets, indicating a greater degree of word sense ambiguity (*assume* has 9 synsets). Based on the observation that an epistemic verb taking a clausal complement marked with *that* is a very strong indication of hedging, we only considered verb senses which subcategorize for a *that* complement. Expansion via WordNet resulted in 66 additional lexical features.

Next, we considered the case of nominalizations. Again, based on corpus analysis, we noted that nominalizations of epistemic verbs and adjectives are a common and effective means of hedging in molecular biology articles. The UMLS SPECIALIST Lexicon provides syntactic information, including nominalizations, for biomedical as well as general English terms. We extracted the nominalizations of words in our expanded dictionary of epistemic verbs and adjectives from UMLS SPECIALIST Lexicon and discarded those that do not occur in our pool of sentences, resulting in an additional 48 terms. Additional 5 lexical hedging cues (e.g., *tend, support*) were identified via manual corpus analysis and further expanded using the methodology described above.

An interesting class of cues are terms expressing strong certainty (“unhedgers”). Used within the scope of negation, these terms suggest hedging, while in the absence of negation they strongly suggest a non-speculative context. Examples of these include verbs indicating certainty, such as *know, demonstrate, prove and show*, and adjectives, such as *clear*. These features were also added to the dictionary and used together with other surface

cues to recognize speculative sentences. The hedging dictionary contains a total of 190 features.

## 7 Quantifying Hedging Strength

It is clear that not all hedging devices are equally strong and that the choice of hedging device affects the strength of the speculation. However, determining the strength of a hedging device is not trivial. The fuzzy pragmatic model proposed by Hyland (1998) employs general descriptive terms such as “strong” and “weak” when discussing particular cases of hedging and avoids the need for precise quantification. Light et al. (2004) report low inter-annotator agreement in distinguishing low speculative sentences from highly speculative ones. From a computational perspective, it would be useful to quantify hedging strength to determine the confidence of the author in his or her proposition.

As a first step in accommodating noticeable differences in strengths of hedging features, we assigned weights (1 to 5, 1 representing the lowest hedging strength and 5 the highest) to all hedging features in our dictionary. Core features were assigned weights based on the discussion in Hyland (1998). For instance, he identifies modal auxiliaries, *may* and *might*, as the prototypical hedging devices, and they were given weights of 5. On the other hand, modal auxiliaries commonly used in non-epistemic contexts (*would*, *could*) were assigned a lower weight of 3. Though not as strong as *may* and *might*, core epistemic verbs and adverbs are generally good hedging cues and therefore were assigned weights of 4. Core epistemic adjectives and nouns often co-occur with other syntactic features to act as strong hedging cues and were assigned weights of 3. Terms added to the dictionary via expansion were assigned a weight one less than their seed terms. For instance, the nominalization *supposition* has weight 2, since it is expanded from the verb *suppose* (weight 3), which is further expanded from its synonym *speculate* (weight 4), a core epistemic verb. The reduction in weights of certain hedging cues reflects their peripheral nature in hedging.

Hyland (1998) notes that writers tend to combine hedges (“harmonic combinations”) and suggests the possibility of constructing scales of certainty and tentativeness from these combinations. In a similar vein, we accumulate the weights

of the hedging features found in a sentence and assign an overall hedging score to each sentence.

## 8 The Role of Syntax

Corpus analysis shows that various syntactic devices play a prominent role in hedging, both as hedging cues and for strengthening or weakening effects. For instance, while some epistemic verbs do not act as hedging cues (or may be weak hedging cues) when used alone, together with a *that* complement or an infinitival clause, they are good indicators of hedging. A good example is *appear*, which often occurs in molecular biology articles with its “come into sight” meaning (5) and becomes a good hedging cue when it takes an infinitival complement (6):

- (5) The linearity of the ommatidial arrangement was disrupted and numerous gaps appeared between ommatidia arrow.
- (6) In these data a substantial fraction of both silent and replacement DNA mutations appear to affect fitness.

On the other hand, as discussed above, words expressing strong certainty (“unhedgers”) are good indicators of hedging when negated, and strongly non-speculative otherwise.

We examined the training set and identified the most salient syntactic patterns that play a role in hedging. A syntactic pattern, or lack thereof, affects the overall score assigned to a hedging cue; a strengthening syntactic pattern will increase the overall score contributed by the cue, while a weakening pattern will decrease it. For instance, in sentence (5) above, the absence of the infinitival complement will reduce the score contribution of *appear* by 1, resulting in a score of 3 instead of 4. On the other hand, that *appear* takes an infinitival clause in example (6) will increase the score contribution of *appear* by 1. All score contributions of a sentence add up to its hedging score.

A purely syntactic case is that of *whether (if)*. Despite being a conjunction, it seems to act as a hedging cue when it introduces a clausal complement regardless of existence of any other hedging cue from the hedging dictionary. The basic syntactic patterns we identified and implemented and their effect on the overall hedging score are given in Table 2.

To obtain the syntactic structures of sentences, we used the statistical Stanford Lexicalized Parser (Klein and Manning, 2003), which provides a full parse tree, in addition to part-of-speech tagging based on the Penn Treebank tagset. A particularly useful feature of the Stanford Lexicalized Parser is typed dependency parses extracted from phrase structure parses (deMarneffe, et al. (2006)). We use these typed dependency parses to identify clausal complements, infinitival clauses and negation. For instance, the following two dependency relations indicate a clausal complement marked with *that* and identify the second syntactic pattern in Table 2.

*ccomp*(<EPISTEMIC VERB>,<VB>)  
*complm*(<VB>,that)

In these relations, *ccomp* stands for clausal complement with internal subject and *complm* stands for complementizer. *VB* indicates any verb.

Syntactic Pattern	Effect on Score
<EPISTEMIC VERB> <i>to</i> ( <i>inf</i> ) VB	+1
<EPISTEMIC VERB> <i>that</i> ( <i>comp</i> ) VB	+2
Otherwise	-1
<EPISTEMIC NOUN> followed by <i>that</i> ( <i>comp</i> )	+2
Otherwise	-1
<i>not</i> <UNHEDGING VERB>	+1
<i>no</i>   <i>not</i> <UNHEDGING NOUN>	+2
<i>no</i>   <i>not</i> immediately followed by <UNHEDGING ADVERB>	+1
<i>no</i>   <i>not</i> immediately followed by <UNHEDGING ADJECTIVE>	+1
<i>whether</i>   <i>if</i> in a clausal complement context	3

Table 2. Syntactic patterns and their effect on the overall hedging score.

## 9 Baseline

For our experiments, we used two baselines. First, we used the substring matching method reported in Light et al. (2004), which labels sentences containing one of more of the following as speculative: *suggest*, *potential*, *likely*, *may*, *at least*, *in part*, *possibl*, *further investigation*, *unlikely*, *putative*, *insights*, *point toward*, *promise* and *propose* (Baseline1). Secondly, we used the top 15 ranked

term features determined using  $P(spec|x_j)$  in training and classification models (at smoothing parameter  $\alpha=5$ ) reported in Medlock and Briscoe (2007): *suggest*, *likely*, *may*, *might*, *seems*, *Taken*, *suggests*, *probably*, *Together*, *suggesting*, *possibly*, *suggested*, *findings*, *observations*, *Given*. Our second baseline uses the substring matching method with these features (Baseline2).

## 10 Results

The evaluation results obtained using the baseline methods are given in Table 3.

Method	Precision	Recall	Accuracy	F <sub>1</sub> score
Baseline1	0.79	0.40	0.82	0.53
Baseline2	0.95	0.43	0.85	0.60

Table 3. Baseline evaluation results.

The evaluation results obtained from our system by varying the overall hedging score and using it as threshold are given in Table 4. It is worth noting that the highest overall hedging score we obtained was 16; however, we do not show the results for every possible threshold here for brevity.

Hedging Score Threshold	Precision	Recall	Accuracy	F <sub>1</sub> score
1	0.68	0.95	0.88	0.79
2	0.75	0.94	0.91	0.83
<b>3</b>	<b>0.85</b>	<b>0.86</b>	<b>0.93</b>	<b>0.85</b>
4	0.91	0.71	0.91	0.80
5	0.92	0.63	0.89	0.75
6	0.97	0.40	0.85	0.57
7	1	0.19	0.79	0.33

Table 4. Evaluation results from our system.

As seen from Table 3 and Table 4, our results show improvement over both baseline methods in terms of accuracy and F<sub>1</sub> score. Increasing the threshold (thereby requiring more or stronger hedging devices to qualify a sentence as speculative) improves the precision while lowering the recall. The best accuracy and F<sub>1</sub> score are achieved at threshold t=3. At this threshold, the differences between the results obtained with our method and baseline methods are statistically significant at 0.01 level ( $p < 0.01$ ).

Method	Recall/Precision BEP
Baseline1	0.60
Baseline2	0.76
Our system	0.85

Table 5. Recall / precision break-even point (BEP) results

With the threshold providing the best accuracy and  $F_1$  score, precision and recall are roughly the same (0.85), indicating a recall/precision BEP of approximately 0.85, also an improvement over 0.76 achieved with a weakly supervised classifier (Medlock and Briscoe, 2007). Recall/precision BEP scores are given in Table 5.

## 11 Discussion

Our results confirm that writers of scientific articles employ basic, predictable hedging strategies to soften their claims or to indicate uncertainty and demonstrate that these strategies can be captured using a combination of lexical and syntactic means. Furthermore, the results indicate that hedging cues can be gainfully weighted to provide a rough measure of tentativeness or uncertainty. For instance, a sentence with the highest overall hedging score is given below:

- (7) In one study, Liquid facets was *proposed* to target DI to an endocytic recycling compartment *suggesting* that recycling of DI *may* be required for signaling.

On the other hand, hedging is not strong in the following sentence, which is assigned an overall hedging score of 2:

- (8) There is no apparent need for cytochrome c release in *C. elegans* since CED-4 does not require it to activate CED-3.

Below, we discuss some of the common error types we encountered. Our discussion is based on evaluation at hedging score threshold of 0, where existence of a hedging cue is sufficient to label a sentence speculative.

Most of the false negatives produced by the system are due to syntactic patterns not addressed by our method. For instance, negation of “unhedgers” was used as a syntactic pattern; the pattern was able to recognize *know* as an “unhedger” in the following sentence, but not the negative quantifier (*little*), labeling the sentence as non-speculative.

- (9) *Little* was *known* however about the specific role of the roX RNAs during the formation of the DCC.

In fact, Hyland (1998) notes “negation in scientific research articles shows a preference for negative quantifiers (*few*, *little*) and lexical negation (*rarely*, *overlook*).” However, we have not encountered this pattern while analyzing the training set and have not addressed it. Nevertheless, our approach lends itself to incremental development and adding such a pattern to our rulebase is relatively simple.

Another type of false negative is caused by certain derivational forms of epistemic words. In the following example, the adjective *suggestive* is not recognized as a hedging trigger, even though its base form *suggest* is an epistemic verb.

- (10) Phenotypic differences are *suggestive* of distinct functions for some of these genes in regulating dendrite arborization.

It seems that more sophisticated lexicon expansion rules can be employed to handle such cases. For example, WordNet’s “derivationally related form” feature may be used as the basis of these expansion rules.

Regarding false positives, most of them are due to word sense ambiguity concerning hedging cues. For instance, the modal auxiliary *could* is frequently used as a past tense form of *can* in scientific articles to express the role of enabling conditions and external constraints on the occurrence of the proposition rather than uncertainty or tentativeness regarding the proposition. Currently, our system is unable to recognize such cases. An example is given below:

- (10) Also we could not find any RAG-like sequences in the recently sequenced sea urchin lancelet hydra and sea anemone genomes, which encode RAG-like sequences.

The context around the hedging cue seems to play a role in these cases. First person plural pronoun (*we*) and/or reference to objective enabling conditions seem to be a common characteristic among false positive cases of *could*.

In other cases, such as *appear*, in the absence of strengthening syntactic cues (*to*, *that*), we lower the hedging score; however, depending on the threshold, this may not be sufficient to render the sentence non-speculative. Rather than lowering the score equally for all epistemic verbs, a more

appropriate approach would be to consider verb senses separately (e.g., *appear* should be effectively unhedged without a strengthening cue, while *suggest* should only be weakened).

Another type of false positives concern “weak” hedging cues, such as epistemic deductive verbs (*conclude, estimate*) as well as adverbs (*essentially, usually*) and nominalizations (*implication, assumption*).

We have also seen a few instances, which seem speculative on the surface, but were labeled non-speculative. An example is given below:

- (11) Caspases can also be activated with the aid of Apaf-1, which in turn *appears to be* regulated by cytochrome c and dATP.

## 12 Conclusion and Future Work

In this paper, we present preliminary experiments we conducted in recognizing speculative sentences. We draw on previous linguistic work and extend it via semi-automatic methods of lexical acquisition. Using a corpus specifically annotated for speculation, we demonstrate that our linguistically oriented approach improves on the previously reported results.

Our next goal is to extend our work using a larger, more comprehensive corpus. This will allow us to identify other commonly used hedging strategies and refine and expand the hedging dictionary. We also aim to refine the weighting scheme in a more principled way.

While recognizing that a sentence is speculative is useful in and of itself, it seems more interesting and clearly much more challenging to identify speculative sentence fragments and the propositions that are being hedged. In the future, we will move in this direction with the goal of characterizing the semantics of speculative language.

## Acknowledgements

We would like to thank Thomas C. Rindfleisch for his suggestions and comments on the first draft of this paper.

## References

deMarneffe, M. C., MacCartney B., Manning C.D. 2006. Generating Typed Dependency Parses from Phrase Structure Parses. *In Proc of 5th International*

*Conference on Language Resources and Evaluation*, pp. 449-54.

DiMarco C. and Mercer R.E. 2004. Hedging in Scientific Articles as a Means of Classifying Citations. *In Exploring Attitude and Affect in Text: Theories and Applications AAAI-EAAT 2004*. pp.50-4.

Fellbaum, C. 1998. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, MA.

Friedman C., Alderson P., Austin J., Cimino J.J., Johnson S.B. 1994. A general natural-language text processor for clinical radiology. *Journal of the American Medical Informatics Association*, 1(2): 161-74.

Hyland K. 1998. *Hedging in Scientific Research Articles*. John Benjamins B.V., Amsterdam, Netherlands.

Kipper Schuler, K. 2005. *VerbNet: A broad-coverage, comprehensive verb lexicon*. PhD thesis, University of Pennsylvania.

Klein D. and Manning C. D. 2003. Accurate unlexicalized parsing. *In Proc of 41st Meeting of the Association for Computational Linguistics*. pp. 423-30.

Lakoff G. 1972. Hedges: A Study in Meaning Criteria and the Logic of Fuzzy Concepts. *Chicago Linguistics Society Papers*, 8, pp.183-228.

Light M., Qiu X.Y., Srinivasan P. 2004. The Language of Bioscience: Facts, Speculations, and Statements in between. *In BioLINK 2004: Linking Biological Literature, Ontologies and Databases*, pp. 17-24.

McCray A. T., Srinivasan S., Browne A. C. 1994. Lexical methods for managing variation in biomedical terminologies. *In Proc of 18th Annual Symposium on Computer Applications in Medical Care*, pp. 235-9.

Medlock B. and Briscoe T. 2007. Weakly Supervised Learning for Hedge Classification in Scientific Literature. *In Proc of 45<sup>th</sup> Meeting of the Association for Computational Linguistics*. pp.992-9.

Palmer F.R. 1986. *Mood and Modality*. Cambridge University Press, Cambridge, UK.

Pustejovsky J., Hanks P., Sauri R., See A., Gaizauskas R., Setzer A., Radev D., Sundheim B., Day D. Ferro L., Lazo M. 2003. The TimeBank Corpus. *In Proc of Corpus Linguistics*. pp. 647-56.

Sauri R., Verhagen M., Pustejovsky J. 2006. SlinkET: a partial modal parser for events. *In Proc of 5<sup>th</sup> International Conference on Language Resources and Evaluation*.

Wilbur W.J., Rzhetsky A., Shatkay H. 2006. New Directions in Biomedical Text Annotations: Definitions, Guidelines and Corpus Construction. *BMC Bioinformatics*, 7:356.