

A High-Precision Approach to Detecting Hedges and Their Scopes

Halil Kilicoglu and Sabine Bergler

Department of Computer Science and Software Engineering

Concordia University

1455 de Maisonneuve Blvd. West

Montréal, Canada

{h.kilico,bergler}@cse.concordia.ca

Abstract

We extend our prior work on speculative sentence recognition and speculation scope detection in biomedical text to the CoNLL-2010 Shared Task on Hedge Detection. In our participation, we sought to assess the extensibility and portability of our prior work, which relies on linguistic categorization and weighting of hedging cues and on syntactic patterns in which these cues play a role. For Task 1B, we tuned our categorization and weighting scheme to recognize hedging in biological text. By accommodating a small number of vagueness quantifiers, we were able to extend our methodology to detecting vague sentences in Wikipedia articles. We exploited constituent parse trees in addition to syntactic dependency relations in resolving hedging scope. Our results are competitive with those of closed-domain trained systems and demonstrate that our high-precision oriented methodology is extensible and portable.

1 Introduction

Natural language is imbued with uncertainty, vagueness, and subjectivity. However, information extraction systems generally focus on extracting factual information, ignoring the wealth of information expressed through such phenomena. In recent years, the need for information extraction and text mining systems to identify and model such extra-factual information has increasingly become clear. For example, online product and movie reviews have provided a rich context for analyzing sentiments and opinions in text (see Pang and Lee (2008) for a recent survey), while tentative, speculative nature of scientific writing, particularly in biomedical literature, has provided

impetus for recent research in speculation detection (Light et al., 2004). The term *hedging* is often used as an umbrella term to refer to an array of extra-factual phenomena in natural language and is the focus of the CoNLL-2010 Shared Task on Hedge Detection.

The CoNLL-2010 Shared Task on Hedge Detection (Farkas et al., 2010) follows in the steps of the recent BioNLP'09 Shared Task on Event Extraction (Kim et al., 2009), in which one task (speculation and negation detection) was concerned with notions related to hedging in biomedical abstracts. However, the CoNLL-2010 Shared Task differs in several aspects. It sheds light on the pervasiveness of hedging across genres and domains: in addition to biomedical abstracts, it is concerned with biomedical full text articles as well as with Wikipedia articles. Both shared tasks have been concerned with scope resolution; however, their definitions of scope are fundamentally different: the BioNLP'09 Shared Task takes the scope of a speculation instance to be an abstract semantic object (an *event*), thus a normalized logical form. The CoNLL-2010 Shared Task, on the other hand, defines it as a textual unit based on syntactic considerations. It is also important to note that hedging in scientific writing is a core aspect of the genre (Hyland, 1998), while it is judged to be a flaw which has to be eradicated in Wikipedia articles. Therefore, hedge detection in these genres serves different purposes: explicitly encoding the factuality of a scientific claim (*doubtful*, *probable*, etc.) versus flagging unreliable text.

We participated in both tasks of the CoNLL-2010 Shared Task: namely, detection of sentences with uncertainty (Task 1) and resolution of uncertainty scope (Task 2). Since we pursued both of these directions in prior work, one of our goals in participating in the shared task was to assess how our approach generalized to previously unseen texts, even genres. Towards this goal, we adopted

an open-domain approach, where we aimed to use previously developed techniques to the extent possible. Among all participating groups, we distinguished ourselves as the one that fully worked in an open-domain setting. This approach worked reasonably well for uncertainty detection (Task 1); however, for the scope resolution task, we needed to extend our work more substantially, since the notion of scope was fundamentally different than what we adopted previously. The performance of our system was competitive; in terms of F-measure, we were ranked near the middle in Task 1, while a more significant focus on scope resolution resulted in fourth place ranking among fifteen systems. We obtained the highest precision in tasks focusing on biological text. Considering that we chose not to exploit the training data provided to the full extent, we believe that our system is viable in terms of extensibility and portability.

2 Related Work

Several notions related to hedging have been previously explored in natural language processing. In the news article genre, these have included *certainty*, *modality*, and *subjectivity*. For example, Rubin et al. (2005) proposed a four dimensional model to categorize *certainty* in news text: *certainty level*, *focus*, *perspective* and *time*. In the context of TimeML (Pustejovsky et al., 2005), which focuses on temporal expressions in news articles, *event modality* is encoded using subordination links (SLINKs), some of which (MODAL,EVIDENTIAL) indicate hedging (Saurí et al., 2006). Saurí (2008) exploits modality and polarity to assess the *factuality degree* of events (whether they correspond to facts, counter-facts or possibilities), and reports on FactBank, a corpus annotated for event factuality (Saurí and Pustejovsky, 2009). Wiebe et al. (2005) consider subjectivity in news articles, and focus on the notion of *private states*, encompassing speculations, opinions, and evaluations in their subjectivity frames.

The importance of speculative language in biomedical articles was first acknowledged by Light et al. (2004). Following work in this area focused on detecting speculative sentences (Medlock and Briscoe, 2007; Szarvas, 2008; Kilicoglu and Bergler, 2008). Similar to Rubin et al.'s (2005) work, Thompson et al. (2008) proposed a categorization scheme for *epistemic modality* in

biomedical text according to the type of information expressed (e.g., *certainty level*, *point of view*, *knowledge type*). With the availability of the BioScope corpus (Vincze et al., 2008), in which negation, hedging and their scopes are annotated, studies in detecting speculation scope have also been reported (Morante and Daelemans, 2009; Özgür and Radev, 2009). Negation and uncertainty of bio-events are also annotated to some extent in the GENIA event corpus (Kim et al., 2008). The BioNLP'09 Shared Task on Event Extraction (Kim et al., 2009) dedicated a task to detecting negation and speculation in biomedical abstracts, based on the GENIA event corpus annotations.

Ganter and Strube (2009) elaborated on the link between *vagueness* in Wikipedia articles indicated by *weasel words* and hedging. They exploited word frequency measures and shallow syntactic patterns to detect weasel words in Wikipedia articles.

3 Methods

Our methodology for hedge detection is essentially rule-based and relies on a combination of lexical and syntactic information. Lexical information is encoded in a simple dictionary, and relevant syntactic information is identified using the Stanford Lexicalized Parser (Klein and Manning, 2003). We exploit constituent parse trees as well as corresponding collapsed dependency representations (deMarneffe et al., 2006), provided by the parser.

3.1 Detecting Uncertainty in Biological Text

For detecting uncertain sentences in biological text (Task 1B), we built on the linguistically-inspired system previously described in detail in Kilicoglu and Bergler (2008). In summary, this system relies on a dictionary of lexical speculation cues, derived from a set of core surface realizations of hedging identified by Hyland (1998) and expanded through WordNet (Fellbaum, 1998) synsets and UMLS SPECIALIST Lexicon (McCray et al., 1994) nominalizations. A set of lexical certainty markers (*unhedgers*) are also included, as they indicate hedging when they are negated (e.g., *know*). These hedging cues are categorized by their type (*modal auxiliaries*, *epistemic verbs*, *approximative adjectives*, etc.) and are weighted to reflect their central/peripheral contribution to hedging, inspired by the fuzzy model of Hyland (1998). We use a scale

of 1-5, where 5 is assigned to cues most central to hedging and 1 to those that are most peripheral. For example, the modal auxiliary *may* has a weight of 5, while a relatively weak hedging cue, the epistemic adverb *apparently*, has a weight of 2. The weight sum of cues in a sentence in combination with a predetermined threshold determines whether the sentence in question is uncertain. Syntax, generally ignored in other studies on hedging, plays a prominent role in our approach. Certain syntactic constructions act as cues (e.g., *whether-* and *if-*complements), while others strengthen or weaken the effect of the cue associated with them. For example, a *that-*complement taken by an epistemic verb increases the hedging score contributed by the verb by 2, while lack of any complement decreases the score by 1.

For the shared task, we tuned this categorization and weighting scheme, based on an analysis of the biomedical full text articles in training data. We also adjusted the threshold. We eliminated some hedging cue categories completely and adjusted the weights of a small number of the remaining cues. The eliminated cue categories included *approximative adverbs* (e.g., *generally*, *largely*, *partially*) and *approximative adjectives* (e.g., *partial*), often used to “manipulate precision in quantification” (Hyland, 1998). The other eliminated category included verbs of *effort* (e.g., *try*, *attempt*, *seek*), also referred to as *rationalising narrators* (Hyland, 1998). The motivation behind eliminating these categories was that cues belonging to these categories were never annotated as hedging cues in the training data. The elimination process resulted in a total of 147 remaining hedging cues. Additionally, we adjusted the weights of several other cues that were not consistently annotated as cues in the training data, despite our view that they were strong hedging cues. One example is the epistemic verb *predict*, previously assigned a weight of 4 based on Hyland’s analysis. We found its annotation in the training data somewhat inconsistent, and lowered its weight to 3, thus requiring a syntactic strengthening effect (an infinitival complement, for example) for it to qualify as a hedging cue in the current setting (threshold of 4).

3.2 Detecting Uncertainty in Wikipedia Articles

Task 1W was concerned with detecting uncertainty in Wikipedia articles. Uncertainty in this

context refers more or less to *vagueness* indicated by *weasel words*, an undesirable feature according to Wikipedia policy. Analysis of Wikipedia training data provided by the organizers revealed that there is overlap between *weasel words* and hedging cues described in previous section. We, therefore, sought to adapt our dictionary of hedging cues to the task of detecting vagueness in Wikipedia articles. Similar to Task 1B, changes involved eliminating cue categories and adjusting cue weights. In addition, however, we also added a previously unconsidered category of cues, due to their prominence in Wikipedia data as *weasel words*. This category (*vagueness quantifiers* (Lapin, 2000)) includes words, such as *some*, *several*, *many* and *various*, which introduce imprecision when in modifier position. For instance, in the example below, both *some* and *certain* contribute to vagueness of the sentence.

- (1) Even today, *some cultures* have *certain instances* of their music intending to imitate natural sounds.

For Wikipedia uncertainty detection, eliminated categories included verbs and nouns concerning *tendencies* (e.g., *tend*, *inclination*) in addition to verbs of *effort*. The only modal auxiliary consistently considered a *weasel word* was *might*; therefore, we only kept *might* in this category and eliminated the rest (e.g., *may*, *would*). *Approximative adverbs*, eliminated in detecting uncertainty in biological text, not only were revived for this task, but also their weights were increased as they were more central to vagueness expressions. Besides these changes in weighting and categorization, the methodology for uncertainty detection in Wikipedia articles was essentially the same as that for biological text. The threshold we used in our submission was, similarly, 4.

3.3 Scope Resolution for Uncertainty in Biological Text

Task 2 of the shared task involved hedging scope resolution in biological text. We previously tackled this problem within the context of biological text in the BioNLP’09 Shared Task (Kilicoglu and Bergler, 2009). That task defined the scope of speculation instances as abstract, previously extracted *bio-events*. Our approach relied on finding an appropriate syntactic dependency relation between the bio-event trigger word identified in

earlier steps and the speculation cue. The category of the hedging cue constrained the dependency relations that are deemed appropriate. For example, consider the sentence in (2a), where *involves* is a bio-event trigger for a *Regulation* event and *suggest* is a speculation cue of epistemic verb type. The first dependency relation in (2b) indicates that the epistemic verb takes a clausal complement headed by the bio-event trigger. The second indicates that *that* is the complementizer. This cue category/dependency combination licenses the generation of a speculation instance where the event indicated by the event trigger represents the scope.

- (2) (a) The results *suggest* that M-CSF induction of M-CSF *involves* G proteins, PKC and NF kappa B.
- (b) *ccomp(suggest,involves)*
complm(involves,that)

Several other cue category/dependency combinations sought for speculation scope resolution are given in Table 1. *X* represents a token that is neither a cue nor a trigger (*aux*: auxiliary, *obj*: direct object, *neg*: negation modifier).

Cue Category	Dependency
Modal auxiliary (<i>may</i>)	<i>aux(Trigger,Cue)</i>
Conditional (<i>if</i>)	<i>complm(Trigger,Cue)</i>
Unhedging noun (<i>evidence</i>)	<i>dobj(X,Cue)</i> <i>ccomp(X,Trigger)</i> <i>neg(Cue,no)</i>

Table 1: Cue categories with examples and the dependency relations to search

In contrast to this notion of scope being an abstract semantic object, Task 2 (BioScope corpus, in general) conceptualizes hedge scope as a continuous textual unit, including the hedging cue itself and the biggest syntactic unit the cue is involved in (Vincze et al., 2008). This fundamental difference in conceptualization limits the direct applicability of our prior approach to this task. Nevertheless, we were able to use our work as a building block in extending scope resolution heuristics. We further augmented it by exploiting constituent parse trees provided by Stanford Lexicalized Parser. These extensions are summarized below.

3.3.1 Exploiting parse trees

The constituent parse trees contribute to scope resolution uniformly across all hedging cue categories. We simply determine the phrasal node that dominates the hedging cue and consider the tokens within that phrase as being in the scope of the cue, unless they meet one of the following exclusion criteria:

1. Exclude tokens within post-cue sentential complements (indicated by S and SBAR nodes) introduced by a small number of discourse markers (*thus, whereas, because, since, if, and despite*).
2. Exclude punctuation marks at the right boundary of the phrase
3. Exclude pre-cue determiners and adverbs at the left boundary of the phrase

For example, in the sentence below, the verb phrase that included the modal auxiliary *may* also included the complement introduced by *thereby*. Using the exclusion criteria 1 and 2, we excluded the tokens following *SPACER* from the scope:

- (3) (a) ... motifs *may* be easily compared with the results from BEAM, PRISM and SPACER, thereby extending the SCOPE ensemble to include a fourth class of motifs.
- (b) CUE: *may*
SCOPE: motifs may be easily compared with the results from BEAM, PRISM and SPACER

3.3.2 Extending dependency-based heuristics

The new scope definition was also accommodated by extending the basic dependency-based heuristics summarized earlier in this section. In addition to finding the trigger word that satisfies the appropriate dependency constraint with the hedging cue (we refer to this trigger word as *scope head*, henceforth), we also considered the other dependency relations that the *scope head* was involved in. These relations, then, were used in *right expansion* and *left expansion* of the scope. *Right expansion* involves finding the rightmost token that is in a dependency relation with the *scope head*. Consider the sentence below:

- (4) The surprisingly low correlations between Sig and accuracy may *indicate* that the objective functions employed by motif finding

programs are only a first *approximation* to biological *significance*.

The epistemic verb *indicate* has as its *scope head* the token *approximation*, due to the existence of a clausal complement dependency (*ccomp*) between them. On the other hand, the rightmost token of the sentence, *significance*, has a prepositional modifier dependency (*prep_to*) with *approximation*. It is, therefore, included in the scope of *indicate*. Two dependency types, adverbial clause modifier (*advcl*) and conjunct (*conj*), were excluded from consideration when the rightmost token is sought, since they are likely to signal new discourse units outside the scope.

In contrast to *right expansion*, which applies to all hedging cue categories, *left expansion* applies only to a subset. Left expansion involves searching for a subject dependency governed by the *scope head*. The dependency types descending from the subject (*subj*) type in the Stanford dependency hierarchy are considered: *nsubj* (nominal subject), *nsubjpass* (passive nominal subject), *csbj* (clausal subject) and *csbjpass* (passive clausal subject). In the following example, the first token, *This*, is added to the scope of *likely* through left expansion (*cop*: copula).

- (5) (a) This is most *likely* a conservative estimate since a certain proportion of interactions remain unknown . . .
 (b) *nsubj(likely,This)*
cop(likely,is)

Left expansion was limited to the following cue categories, with the additional constraints given:

1. *Modal auxiliaries*, only when their *scope head* takes a passive subject (e.g., *they is added to the scope of may in they may be annotated as pseudogenes*).
2. Cues in adjectival categories, when they are in copular constructions (e.g., Example (5)).
3. Cues in several adjectival ad verbal categories, when they take infinitival complements (e.g., *this is added to the scope of appears in However, this appears to add more noise to the prediction without increasing the accuracy*).

After scope tokens are identified using the parse tree as well as via left and right expansion, the algorithm simply sets as scope the continuous textual unit that includes all the scope tokens and the

hedging cue. Since, *likely* is the hedging cue and *This* and *estimate* are identified as scope tokens in Example (5), the scope associated with *likely* becomes *This is most likely a conservative estimate*.

We found that citations, numbers and punctuation marks occurring at the end of sentences caused problems in scope resolution, specifically in biomedical full text articles. Since they are rarely within any scope, we implemented a simple stripping algorithm to eliminate them from scopes in such documents.

4 Results and Discussion

The official evaluation results regarding our submission are given in Table 2. These results were achieved with the threshold 4, which was the optimal threshold on the training data.

	Prec.	Recall	F-score	Rank
Task 1B	92.07	74.94	82.62	12/24
Task 1W	67.90	46.02	54.86	10/17
Task 2	62.47	49.47	55.21	4/15

Table 2: Evaluation results

In Task 1B, we achieved the highest precision. However, our relatively low recall led to the placement of our system in the middle. Our system allows adjusting precision versus recall by setting the threshold. In fact, setting the threshold to 3 after the shared task, we were able to obtain overall better results (Precision=83.43, Recall=84.81, F-score=84.12, Rank=8/24). However, we explicitly targeted precision, and in that respect, our submission results were not surprising. In fact, we identified a new type of hedging signalled by coordination (*either . . . or . . .* as well as just *or*) in the training data. An example is given below:

- (6) (a) It will be *either* a sequencing error *or* a pseudogene.
 (b) CUE: *either-or*
 SCOPE: *either a sequencing error or a pseudogene*

By handling this class to some extent, we could have increased our recall, and therefore, F-score (65 out of 1,044 cues in the evaluation data for biological text involved this class). However, we decided against treating this class, as we believe it requires a slightly different treatment due to its special semantics.

In participating in Task 1W, our goal was to test the ease of extensibility of our system. In that regard, our results show that we were able to exploit the overlap between our hedging cues and the weasel words. The major difference we noted between hedging in two genres was the class of vagueness quantifiers, and, with little effort, we extended our system to consider them. We also note that setting the threshold to 3 after the shared task, our recall and F-score improved significantly (Precision=63.21, Recall=53.67, F-score=58.05, Rank=3/17).

Our more substantial effort for Task 2 resulted in a better overall ranking, as well as the highest precision in this task. In contrast to Task 1, changing the threshold in this task did not have a positive effect on the outcome. We also measured the relative contribution of the enhancements to scope resolution. The results are presented in Table 3. Baseline is taken as the scope resolution algorithm we developed in prior work. These results show that: a) scope definition we adopted earlier is essentially incompatible with the BioScope definition b) simply taking the phrase that the hedging cue belongs to as the scope provides relatively good results c) left and right expansion heuristics are needed for increased precision and recall.

	Prec.	Recall	F-score
Baseline	3.29	2.61	2.91
Baseline+ Left/ right expansion	25.18	20.03	22.31
Parse tree	49.20	39.10	43.58
Baseline+ Parse tree	50.66	40.27	44.87
All	62.47	49.47	55.21

Table 3: Effect of scope resolution enhancements

4.1 Error Analysis

In this section, we provide a short analysis of the errors our system generated, focusing on biological text.

Since our dictionary of hedging cues is incomplete and we did not attempt to expand it for Task 1B, we had a fair number of recall errors. As we mentioned above, *either-or* constructions occur frequently in the training and evaluation data, and we did not attempt to handle them. Additionally, some lexical cues, such as *feasible* and *imPLICATE*, do not appear in our dictionary, causing

further recall errors. The weighting scheme also affects recall. For example, the adjective *apparent* has a weight of 2, which is not itself sufficient to qualify a sentence as uncertain (with a threshold of 4) (7a). On the other hand, when it takes a clausal complement, the sentence is considered uncertain (7b). The first sentence (7a) causes a recall error.

- (7) (a) An *apparent* contradiction between the previously reported number of cycling genes ...
 (b) ... it is *apparent* that the axonal termini contain a significantly reduced number of varicosities ...

In some cases, syntactic constructions that play a role in determining the certainty status of a sentence cannot be correctly identified by the parser, often leading to recall errors. For example, in the sentence below, the clausal complement construction is missed by the parser. Since the verb *indicate* has weight 3, this leads to a recall error in the current setting.

- (8) ... *indicating* that dMyc overexpression can substitute for PI3K activation ...

Adjusting the weights of cues worked well generally, but also caused unexpected problems, due to what seem like inconsistencies in annotation. The examples below highlight the effect of lowering the weight of *predict* from 4 to 3. Examples (9a) and (9b) are almost identical on surface and our system predicted both to be uncertain, due to the fact that *predicted* took infinitival complements in both cases. However, only (9a) was annotated as uncertain, leading to a precision error in (9b).

- (9) (a) ... include all protein pairs *predicted* to have posterior odds ratio ...
 (b) Protein pairs *predicted* to have a posterior odds ratio ...

The error cases in scope resolution are more varied. Syntax has a larger role in this task, and therefore, parsing errors tend to affect the results more directly. In the following example, during left-expanding the scope of the modal auxiliary *could*, *RNAi screens*, rather than the full noun phrase *fruit fly RNAi screens*, is identified as the passive subject of the scope head (*associated*), because an appropriate modifier dependency cannot

be found between the noun phrase head *screens* and either of the modifiers, *fruit* and *fly*.

- (10) ... was to investigate whether fruit fly RNAi screens of conserved genes *could* be associated with similar tick phenotypes and tick gene function.

In general, the simple mechanism to exploit constituent parse trees was useful in resolving scope. However, it appears that a nuanced approach based on cue categories could enhance the results further. In particular, the current mechanism does not contribute much to resolving scopes of adverbial cues. In the following example, parse tree mechanism does not have any effect, leading to both a precision and a recall error in scope resolution.

- (11) (a) ... we will consider tightening the definitions and *possibly* splitting them into different roles.
(b) FP: possibly
FN: possibly splitting them into different roles

Left/right expansion strategies were based on the analysis of training data. However, we encountered errors caused by these strategies where we found the annotations contradictory. In Example (12a), the entire fragment is in the scope of *thought*, while in (12b), the scope of *suggested* does not include *it was*, even though on surface both fragments are very similar.

- (12) (a) ... the kinesin-5 motor is *thought* to play a key role.
(b) ... it was *suggested* to enhance the nuclear translocation of NF- κ B.

Post-processing in the form of citation stripping was simplistic, and, therefore, was unable to handle complex cases, as the one shown in the example below. The algorithm is only able to remove one reference at the end.

- (13) (a) ... it is possible that some other signalling system may operate with Semas to confine dorsally projecting neurons to dorsal neuropile [3],[40],[41].
(b) FP: may operate with Semas to confine dorsally projecting neurons to dorsal neuropile [3],[40],
FN: may operate with Semas to confine dorsally projecting neurons to dorsal neuropile

5 Conclusions

Rather than developing a dedicated methodology that exclusively relies on the data provided by organizers, we chose to extend and refine our prior work in hedge detection and used the training data only in a limited manner: to tune our system in a principled way. With little tuning, we achieved the highest precision in Task 1B. We were able to capitalize on the overlap between hedging cues and weasel words for Task 1W and achieved competitive results. Adapting our previous work in scope resolution to Task 2, however, was less straightforward, due to the incompatible definitions of scope. Nevertheless, by refining the prior dependency-based heuristics with left and right expansion strategies and utilizing a simple mechanism for parse tree information, we were able to accommodate the new definition of scope to a large extent. With these results, we conclude that our methodology is portable and easily extensible.

While the results show that using the parse tree information for scope resolution benefited our performance greatly, error analysis presented in the previous sections also suggests that a finer-grained approach based on cue categories could further improve results, and we aim to explore this extension further.

References

- Marie-Catherine deMarneffe, Bill MacCartney, and Christopher D. Manning. 2006. Generating typed dependency parses from phrase structure parses. In *Proceedings of the 5th International Conference on Language Resources and Evaluation*, pages 449–454.
- 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 (CoNLL-2010): Shared Task*, pages 1–12, Uppsala, Sweden, July. Association for Computational Linguistics.
- Christiane Fellbaum. 1998. *WordNet: an electronic lexical database*. MIT Press, Cambridge, MA.
- Viola Ganter and Michael Strube. 2009. Finding Hedges by Chasing Weasels: Hedge Detection Using Wikipedia Tags and Shallow Linguistic Features. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, pages 173–176.

- Ken Hyland. 1998. *Hedging in scientific research articles*. John Benjamins B.V., Amsterdam, Netherlands.
- Halil Kilicoglu and Sabine Bergler. 2008. Recognizing speculative language in biomedical research articles: a linguistically motivated perspective. *BMC Bioinformatics*, 9 Suppl 11:s10.
- Halil Kilicoglu and Sabine Bergler. 2009. Syntactic dependency based heuristics for biological event extraction. In *Proceedings of Natural Language Processing in Biomedicine (BioNLP) NAACL 2009 Workshop*, pages 119–127.
- Jin-Dong Kim, Tomoko Ohta, and Jun'ichi Tsujii. 2008. Corpus annotation for mining biomedical events from literature. *BMC Bioinformatics*, 9:10.
- Jin-Dong Kim, Tomoko Ohta, Sampo Pyysalo, Yoshinobu Kano, and Jun'ichi Tsujii. 2009. Overview of BioNLP'09 Shared Task on Event Extraction. In *Proceedings of Natural Language Processing in Biomedicine (BioNLP) NAACL 2009 Workshop*, pages 1–9.
- Dan Klein and Christopher D Manning. 2003. Accurate unlexicalized parsing. In *Proceedings of the 41th Meeting of the Association for Computational Linguistics*, pages 423–430.
- Shalom Lappin. 2000. An intensional parametric semantics for vague quantifiers. *Linguistics and Philosophy*, 23(6):599–620.
- Marc Light, Xin Y. Qiu, and Padmini Srinivasan. 2004. The language of bioscience: facts, speculations, and statements in between. In *BioLINK 2004: Linking Biological Literature, Ontologies and Databases*, pages 17–24.
- Alexa T. McCray, Suresh Srinivasan, and Allen C. Browne. 1994. Lexical methods for managing variation in biomedical terminologies. In *Proceedings of the 18th Annual Symposium on Computer Applications in Medical Care*, pages 235–239.
- Ben Medlock and Ted Briscoe. 2007. Weakly supervised learning for hedge classification in scientific literature. In *Proceedings of the 45th Meeting of the Association for Computational Linguistics*, pages 992–999.
- Roser Morante and Walter Daelemans. 2009. Learning the scope of hedge cues in biomedical texts. In *Proceedings of the BioNLP 2009 Workshop*, pages 28–36.
- Arzucan Özgür and Dragomir R. Radev. 2009. Detecting speculations and their scopes in scientific text. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 1398–1407.
- Bo Pang and Lillian Lee. 2008. *Sentiment Analysis and Opinion Mining*. Now Publishers Inc, Boston, MA.
- James Pustejovsky, Robert Knippen, Jessica Littman, and Roser Saurí. 2005. Temporal and event information in natural language text. *Language Resources and Evaluation*, 39(2):123–164.
- Victoria L. Rubin, Elizabeth D. Liddy, and Noriko Kando. 2005. Certainty identification in texts: Categorization model and manual tagging results. In James G. Shanahan, Yan Qu, and Janyce Wiebe, editors, *Computing Attitude and Affect in Text: Theories and Applications*, volume 20, pages 61–76. Springer Netherlands, Dordrecht.
- Roser Saurí and James Pustejovsky. 2009. FactBank: a corpus annotated with event factuality. *Language Resources and Evaluation*, 43(3):227–268.
- Roser Saurí, Marc Verhagen, and James Pustejovsky. 2006. Annotating and recognizing event modality in text. In *Proceedings of 19th International FLAIRS Conference*.
- Roser Saurí. 2008. *A Factuality Profiler for Eventualities in Text*. Ph.D. thesis, Brandeis University.
- György Szarvas. 2008. Hedge classification in biomedical texts with a weakly supervised selection of keywords. In *Proceedings of the 46th Meeting of the Association for Computational Linguistics*, pages 281–289.
- Paul Thompson, Giulia Venturi, John McNaught, Simonetta Montemagni, and Sophia Ananiadou. 2008. Categorising modality in biomedical texts. In *Proceedings of LREC 2008 Workshop on Building and Evaluating Resources for Biomedical Text Mining*.
- 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.
- Janyce Wiebe, Theresa Wilson, and Claire Cardie. 2005. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, 39(2):165–210.