Hedging and Speaker Commitment

Anna Prokofieva, Julia Hirschberg
Columbia University
prokofieva@cs.columbia.edu, julia@cs.columbia.edu

Abstract

Hedging is a behavior wherein speakers or writers attempt to distance themselves from the proposition they are communicating. Hedge terms include items such as "***I think*** *X*" or "*It's* ***sort of*** *Y*". Identifying such behaviors is important for extracting meaning from speech and text, and can also reveal information about the social and power relations between the conversants. Yet little research has been done on the automatic identification of hedges since the CONLL 2010 Shared Task. In this paper, we present our newly expanded and generalized guidelines for the annotation of hedge expressions in text and speech. We describe annotation and automatic extraction experiments using these guidelines and describe future work on the automatic identification of hedges.

Keywords: hedging, annotation guidelines, crowd-sourced annotation

1. Introduction

Hedging is a phenomenon in which a speaker communicates a lack of commitment to what they are saying. For example:

(1) "I ***think*** *it's a* ***little*** *odd."*

This phrase contains two hedges,"think" and "a little"; one indicating the speaker's lack of commitment to the proposition "it's a little odd" and the other indicating lack of commitment to the quality of oddness.

Hedges occur quite commonly in text and speech: Prince et al. (1982) noted that hedges occurred about every 15 seconds in their 12-hour medical corpus. Since people may hedge for many reasons - for example, to save face (Prince et al., 1982), to show politeness (Ardissono et al., 1999), or to appear more cooperative (Vasilieva, 2004) - the study of hedging behaviors can give us important insight into conversational dynamics. They are also thought to correlate with power relations between conversational participants in domains such as the medical hierarchy. Our goal is to develop procedures for automatically classifying hedges in text and speech corpora so that we can better define speaker commitments and relationships. To this end we have developed hedging annotation guidelines expanding upon previous work, which we are using for semi-automated corpus annotation.

2. Previous work

Lakoff (1975) originally defined hedges as words "whose job it is to make things fuzzier". Prince et al. (1982) noted that this 'fuzziness' could be manifested in two ways: as fuzziness within the propositional content, or as fuzziness in the relationship between the propositional content and the speaker. These two types of hedges are thus termed *propositional* and *relational*.

Others have expanded this notion of 'fuzziness' to encompass words that signal uncertainty, a lack of precision or non-specificity, or an attempt to downplay speakers' commitment to elements in an utterance. Previous studies of hedging have found that the phenomenon is correlated with many discourse functions, such as attempting to evade questions and avoid criticism (Crystal, 1988). de Figueiredo-Silva (2001) proposed viewing hedging as a manifestation of the speaker's attitude towards a claim and towards their audience. As such, hedging can be viewed as an expression of the speaker's inner state.

On the other hand, we can also look at hedging from the listeners' perspective, since the use of hedge words (or the lack thereof) can shape the listeners' opinion of the speaker and of their argument (Blankenship and Holtgraves, 2005; Hosman and Siltanen, 2006; Erickson et al., 1978). In this way, hedges are part of a feedback loop in conversational dynamics.

To date, most of the exploration of hedging in text has been focused on the domain of academic writing (Meyer, 1997; Hyland, 1998; Varttala, 1999). The organizers of the CONLL 2010 Shared Task investigated hedging in the BioScope corpus, which contains abstracts and articles in the biomedical field. This corpus, along with a Wikipedia corpus annotated for "weasel words" (words that equivocate without communicating a precise claim), were used in the Shared Task to investigate techniques for the automatic detection of hedges (Farkas et al., 2010). This Shared Task produced the first set of detailed guidelines on hedge annotation. However, these guidelines are somewhat domain and genre-dependent.

There has also been some investigation of hedging in other corpora, although to date no additional hedge annotations have been made public (Aijmer, 1986; Poos and Simpson, 2002). There has been little work on hedging in speech, beyond Prince et al. (1982)'s study of conversations between medical personnel and patients; even in that study, the audio data was not made available to the researchers so no specific analysis of the speech itself was possible.

3. Defining Hedges

Given the prevalence and importance of hedging behavior to the interpretation of speaker commitment and other social aspects of dialogue, we have begun a study of hedging behavior with the goal of creating a more general tool for identifying hedges in text and speech. Ultimately, we want to create a corpus annotated for hedging. To this end, we have created a new set of Hedging Annotation Guidelines which are more comprehensive than the CONLL 2010

Guidelines and are applicable to both text and speech from various domains and of various levels of formality.

3.1. Domain and Genre Specificity

These guidelines have been developed and refined using several diverse corpora: the CONLL BioScope Corpus (Vincze et al., 2008), the SCOTUS Supreme Court Corpus, and the NIST Meeting Corpus (Garofolo et al., 2004). In the process, we have explored a number of challenges faced in identifying and annotating the phenomenon.

Our investigations of hedging in multiple domains and genres have shown that many terms clearly used as hedges in other corpora were not included in the CONLL guidelines. Some of the hedge terms we discovered appear to be specific to the domains our corpora represent and the linguistic conventions in those domains. In our new Guidelines, we have thus considerably expanded the set of potential hedge terms based on the hedging behaviors we have observed in these different corpora. For example, "in my opinion" is not mentioned in the CONLL guidelines as a hedge, probably because it did not appear in the corpus, but appears quite frequently in the SCOTUS Corpus as a hedge. This is due to the fact that the CONLL guidelines were meant for annotation on academic text, where expressing a personal opinion is often discouraged, whereas in the Supreme Court arguments of the SCOTUS corpus, the lawyers often hedged their views by stating something as opinion rather than fact in order to avoid criticism from the judges.

Additionally, it became clear that other hedge terms found in our corpora were specific to spoken conversation. We thus added our own observations from the SCOTUS Corpus together with those observed in other speech-focused studies to the guidelines (Prince et al., 1982). A pilot annotation on the more informal NIST meeting corpus (Garofolo et al., 2004) led us to further broaden the guidelines to include hedging instances from other selections of conversational speech. In particular, we were able to add many new multi-word hedge constructions, such as "and all that" and "something or other" to our list of hedges; these were not present in the more formal SCOTUS or BioScope corpora. This illustrates our finding that hedging is quite domain-specific and depends on the level of formality, as well as any established conventions of the domain.

3.2. Hedging and Disfluency

The CONLL Guidelines, developed for text annotation, did not include mechanisms for dealing with speech phenomena such as hesitations, self-repairs, and other disfluencies.

> *(3) "**I think** it's – I **think** it's an extremist group that's trying to make us move faster."*

In (3), there is a repetition of the hedge word; to be consistent with the standard for disfluency annotation, both instances would be marked as hedges. Our pilot annotations of the Supreme Court Corpus showed that these conversational phenomena and others, including interruptions, ungrammatical phrases and incomplete utterances, all require special handling in the annotation guidelines.

Specifically, we annotate the hedge word wherever it is at least partially formed, based on the speaker's intention as far as we can determine such from the context. It is the hope that broadening the scope of our annotation in such a way will allow a more in-depth investigation into the relationship between disfluency and hedging.

3.3. Relational vs Propositional Hedges

Based on Prince et al. (1982), we have expanded and clarified distinctions between relational and propositional hedges. Using Prince et al. (1982)'s definitions, we identify relational hedges as those that have to do with the speaker's relation to the propositional content, and propositional hedges as those that introduce uncertainty into the propositional content itself. Since these distinctions themselves can sometimes be confusing, we have provided additional questions annotators may ask themselves to make such a determination. In particular, the annotator can try to preface a potentially hedged sentence with "I'm certain" to see whether the hedge contained therein is relational or propositional.

> (4) "I'm certain that ... *his feet are **sort of** blue." (propositional hedge)*

> (5) # "I'm certain that ... *I **guess** John is right." (relational hedge)*

In (4), inserting "I'm certain" does not change the meaning of the sentence; however, in (5), such an insertion is infelicitous.

However, there is one type of relational hedge for which this test fails: this is the *attributive* hedge. In attributive hedges, a speaker attributes information to some other source in order to downplay its force (as in (6)) or to garner authoritative power for their statement (as in (7)).

> *(6) "**People I've talked to say** "Lincoln" was okay."*

> *(7) "Well, **the Encyclopedia Britannica says** that, so it must be true."*

We mark these as relational hedges, since in either case such attribution indicates a lack of commitment on the part of the speaker with respect to an entire proposition. These sorts of hedges are difficult to annotate automatically, but are nonetheless important for showing a lack of the speaker's personal investment in what they are saying.

3.4. Multi-word Hedges

Hedges can be single cue words or combinations of words. In some cases words which would not normally function as hedges do so in combination with other words. For example, the phrase "in my understanding" can serve as a hedge even though each individual word, when placed in a different context, would not. "In my mind", "my thinking is" and "if I'm understanding you correctly" are other examples of multi-word relational hedges. Multi-word propositional hedges include "and so forth" and "or something like that". Attributive hedges are most often multi-word hedges as well, since both the source to which the information is being attributed, along with the accompanying verb, are included in the hedge.

3.5. Ambiguity

One of the major difficulties in detecting hedges is that potential hedge words are inherently ambiguous. For example:

*(1) "I **think** it's a **little** odd."*

(2) "I think about you all the time."

In (1),"think" is a hedge, but not so in (2). This is true for most hedge verbs and distinguishing whether the verb is being used in a hedging context is a difficult task even for trained annotators. Moving forward, we plan to address these issues using word sense disambiguation techniques. Yarowsky (2000) successfully utilized hierarchical decision lists for a word sense disambiguation task and achieved a precision of 78.9%; we believe that such an approach, which would use lexical and syntactic features to distinguish hedge senses from non-hedge senses, would be adequate to resolve this issue.

3.6. Hedges in Questions

Due to the inherent uncertainty that questions themselves convey, the CONLL 2010 guidelines did not mark hedges in questions. However, we have found that it is in fact possible to find hedges that are independent of the overall uncertainty conveyed by the question. For example:

*(6) "What about the argument that the plaintiff **may not** have been harmed by the disclosure?"*

*(7) "Is this the type of statute that depends **largely** on private enforcement to implement it?"*

We find hedges in both wh- and yes-no questions. In (6), the speaker is questioning the validity of "the argument", but the argument itself contains a hedge ("may") that is independent of the overall uncertainty inherent in the question. In (7), the question itself expresses the speaker's uncertainty about the type of the statute, but the presence of the hedge "largely" is independent of that uncertainty.

In general, hedges should be identified in questions when the hedge words themselves do not identify the statement as a question. For example, auxiliaries that might serve as hedges in statements are not marked in questions, because their use in questions is dictated by rules of grammar rather than a desire to hedge. For example, in: *"Could you clarify this for me?"*, "could" is not marked as a hedge.

In the specific case of statements followed by tag questions, such as: *"It **might** rain, might it not?"*, "might" would be marked as a hedge in the first part of the statement (which can stand as a statement by itself), but not in the tag.

4. Data

Major revisions were necessary to make the guidelines appropriate for annotating text as well as speech, which suggests that hedging may be domain specific. To that end, we wanted to compare whether hedging was more or less prevalent in formal speech as compared to informal speech. We obtained gold standard annotations as per our latest iteration of the annotation guidelines on the Supreme Court Corpus (an instance of less conversational, more formal speech) to compare the presence of hedging therein to the hedging found in the NIST Meeting Corpus (arguably a much more informal, conversational setting).

	SCOTUS	NIST
% Turns with Hedges	38.5%	23.5%
% Sentences with Hedges	23.0%	16.9%
% hRel	71.4%	53.4%
% hProp	28.6%	46.6%

Table 1: Presence of hedges in the SCOTUS and NIST Meeting corpora.

These results were surprising given that we expected more hedging in informal speech. However, the high percentage of relational hedges in the SCOTUS corpus can be explained by the fact that lawyers frequently used "I think" when responding to the judges' queries; this can also account for the higher percentage of hedging in general in that corpus.

5. Automatic Hedge Detection

While our guidelines focus on the lexical items which **may** serve as hedges, they rely upon human interpretation of the context in which potential hedge terms occur in order to determine whether an item is being used as a hedge or not. To understand the importance of this disambiguation process to the identification of hedges, we performed a small experiment in automatic hedge detection.

Our pilot annotation of meetings from the NIST Meeting Corpus has given us a small seed of gold standard data. To motivate the necessity of creating a smart algorithm for the automatic detection of hedges, as opposed to a keyword-search approach, we ran a simple lexical-based search for potential hedges on those meetings. The keywords used were hedges mentioned in the CONLL 2010 Guidelines and those found in a previous annotation exercise we had done on the Supreme Court Corpus.

	NIST Corpus
Precision	0.45
Recall	0.66
F-score	0.53

Table 2: Keyword search approach to hedge detection.

These results provide some evidence that hedge detection requires more than simple key-word search. In the majority of cases, words that are identified by the lexical search as hedges are actually not hedges in that particular context. Moreover, only two-thirds of the hedge terms identified by our labelers in the NIST Meeting Corpus had been previously seen in other corpora. Thus, successful hedge detection will need to involve not only disambiguation of potential hedge terms but also methods to identify new ways of expressing this phenomenon.

Given that annotating hedging can be complicated and time-consuming, we are exploring the potential for crowd-

sourcing hedge annotation, using Amazon Mechanical Turk (AMT). However, as with any complex task, this will require careful planning in order to obtain reliable annotations from untrained labelers. Currently we are developing a multi-stage strategy to incorporate crowd-sourcing into the process of creating a large corpus annotated for hedging. We are building a rule-based algorithm from our guidelines to identify potential hedges syntactically, using terms identified by simple keyword search. These can then be checked by AMT labelers to distinguish hedge uses from non-hedge uses in a series of simple word sense disambiguation tasks. Specifically, annotators would be presented with a sentence containing a potential hedge and asked whether that word could be replaced by a synonym representing one of its potential senses.

*(1) "It's **sort of** diagonal here."*

*Does **sort of** in this sentence mean 'type of'?*

In this case, the correct answer would be 'no' and that would inform us that "sort of" was being used in a hedging sense in this sentence.

Snow et al. (2008) conducted a similar word sense disambiguation task on Amazon Mechanical Turk and were able to obtain 100% accuracy using majority voting based on 10 annotations of each word. Those sentences that are verified by multiple labelers as containing hedges in this first stage will then be passed along to the second stage of annotation. In this stage, annotators will be asked to identify the type of hedge, relational or propositional, by answering questions about the role of the hedge in the matrix sentence. We also hope to reduce the amount of annotation necessary in the first verification stage by using an active learning algorithm trained on a small seed set of gold standard annotated data in order to select the most ambiguous and difficult cases for annotation. We plan to use this additional annotated data to train a statistical classifier to disambiguate hedge uses automatically.

6. Conclusion

In this paper, we have described newly expanded and generalized guidelines for the annotation of hedge expressions in text and speech. We present a more detailed description of this phenomenon, some preliminary experimental results on annotation and automatic detection of hedges, and discuss future plans for disambiguating potential hedge terms using crowd-sourcing and, eventually, automatic machine learning methods.

7. References

Aijmer, K. (1986). Discourse variation and hedging. *Corpus Linguistics II. New studies in the analysis and exploitation of computer corpora*, pages 1–18.

Ardissono, L., Boella, G., and Lesmo, L. (1999). Politeness and speech acts. *Proc. Workshop on Attitude, Personality and Emotions in User-Adapted Interaction*, pages 41–55.

Blankenship, K. and Holtgraves, T. (2005). The role of different markers of linguistic powerlessness in persuasion. *Journal of Language and Social Psychology*, 24(1):3–24.

Crystal, D. (1988). On keeping one's hedges in order. *English Today*, 15:46–47.

de Figueiredo-Silva, M. I. R. (2001). Teaching academic reading: Some initial findings from a session on hedging. *Postgraduate Conference of the University of Edinburgh*.

Erickson, B., Lind, E., Johnson, B., and O'Barr, W. (1978). Speech style and impression formation in a court setting: The effects of "powerful" and "powerless" speech. *Journal of Experimental Social Psychology*, 14(3):266–279.

Farkas, R., Vincze, V., Mora, G., Csirik, J., and Szarvas, G. (2010). The conll-2010 shared task: learning to detect hedges and their scope in natural language text. *Proceedings of the Fourteenth Conference on Computational Natural Language Learning—Shared Task*, pages 1–12.

Garofolo, J. S., Laprun, C., Michel, M., Stanford, V., and Tabassi, E. (2004). The nist meeting room pilot corpus. *LREC*.

Hosman, L. A. and Siltanen, S. (2006). Powerful and powerless language forms their consequences for impression formation, attributions of control of self and control of others, cognitive responses, and message memory. *Journal of Language and Social Psychology*, 25(1):33–46.

Hyland, K. (1998). *Hedging in scientific research articles*, volume 54. John Benjamins Publishing.

Lakoff, G. (1975). *Hedges: A study in meaning criteria and the logic of fuzzy concepts*. Springer, Netherlands.

Meyer, P. G. (1997). Hedging strategies in written discourse: Strengthening the argument by weakening the claim. In *Hedging and Discourse: Approaches to the Analysis of a Pragmatic Phenomenon in Academic Texts*, Berlin. Walter de Gruyter.

Poos, D. and Simpson, R. (2002). Cross-disciplinary comparisons of hedging. *Using corpora to explore linguistic variation*, 9(1).

Prince, E. F., Frader, J., and Bosk, C. (1982). On hedging in physician-physician discourse. *Linguistics and the Professions*, pages 83–97.

Snow, R., O'Connor, B., Jurafsky, D., and Ng, A. (2008). Cheap and fast—but is it good?: evaluating non-expert annotations for natural language tasks. In *Proceedings of the conference on empirical methods in natural language processing*, pages 254–263. Association for Computational Linguistics.

Varttala, T. (1999). Remarks on the communicative functions of hedging in popular scientific and specialist research articles on medicine. *English for Specific Purposes*, 18(2):177–200.

Vasilieva, I. (2004). Gender-specific use of boosting and hedging adverbs in english computer-related texts–a corpus-based study. *International Conference on Language, Politeness and Gender*, pages 2–5.

Vincze, V., Szarvas, G., Farkas, R., Mora, G., and Csirik, J. (2008). The bioscope corpus: biomedical texts annotated for uncertainty, negation and their scopes. *BMC bioinformatics*, 9(Suppl 11):S9.

Yarowsky, D. (2000). Hierarchical decision lists for word sense disambiguation. *Computers and the Humanities*, 34(1-2):179–186.