

# Crowdsourced Hedge Term Disambiguation

Morgan Ulinski and Julia Hirschberg

Department of Computer Science

Columbia University

New York, NY, USA

{mulinski, julia}@cs.columbia.edu

## Abstract

We address the issue of acquiring quality annotations of *hedging* words and phrases, linguistic phenomena in which words, sounds, or other constructions are used to express ambiguity or uncertainty. Due to the limited availability of existing corpora annotated for hedging, linguists and other language scientists have been constrained as to the extent they can study this phenomenon. In this paper, we introduce a new method of acquiring hedging annotations via crowdsourcing, based on reformulating the task of labeling hedges as a simple word sense disambiguation task. We also introduce a new hedging corpus we have constructed by applying this method, a collection of forum posts annotated using Amazon Mechanical Turk. We found that the crowdsourced judgments we obtained had an inter-annotator agreement of 92.89% (Fleiss' Kappa=0.751) and, when comparing a subset of these annotations to an expert-annotated gold standard, an accuracy of 96.65%.

## 1 Introduction

Hedging refers to the use of words, sounds, or constructions that add ambiguity or uncertainty to spoken or written language. Hedging can indicate a speaker's lack of commitment to what they are saying or an attempt to distance themselves from the proposition they are communicating. Identifying hedging behavior in conversational speech and text can also reveal information about social and power relations between conversants. Additionally, since hedging can be indicative of a lack of speaker commitment, identifying hedging is of interest to the information extraction community, to determine the extent to which statements have been believed by the writer or speaker.

A major challenge in identifying hedges is that many hedge words and phrases are ambiguous.

For example, in (1a), *appear* is used as a hedge word, but not in (1b).

- (1) a. The problem **appears** to be a bug in the software.
- b. A man suddenly **appeared** in the doorway.

Currently there are few corpora annotated for hedging, and these are available in a limited number of genres. In particular, there is currently no corpus of *informal* language annotated with hedge behavior. Acquiring expert annotations on text in other genres can be time consuming and may be cost prohibitive, which is an impediment to exploring how hedging can help with applications based on text in other genres. To address these issues, we have developed a method of acquiring hedge annotations through crowdsourcing, by framing the hedge identification task as a simple word sense disambiguation problem. In this paper, we describe this method and also our use of Amazon Mechanical Turk to construct a corpus of forum posts labeled with hedge information.

In Section 2, we discuss related work. In Section 3, we describe how we constructed our dictionary of hedge terms and created the *hedge* and *non-hedge* definitions for each. Section 4 describes the crowdsourcing task in more detail and discusses the resulting corpus. We conclude in Section 5.

## 2 Related Work

Currently, there is limited material available for studying hedging. The CoNLL-2010 shared task on learning to detect hedges (Farkas et al., 2010) used the BioScope corpus (Vincze et al., 2008) of biomedical abstracts and articles and a Wikipedia corpus annotated for “weasel words”. Because of the domain-specific nature of these corpora,

they can be difficult to apply to other text genres, such as social media or blogs. Additionally, the Wikipedia definition of a *weasel word* is slightly different than that of a *hedge*. Weasel words include language referring to personal opinions and subjectivity (e.g. *excellent, best*) in addition to uncertainty and lack of speaker commitment. Thus, it may be difficult to use the Wikipedia corpus to study hedging as a phenomenon that is distinct from subjectivity. Both the BioScope corpus and the Wikipedia corpus were annotated by experts and/or trained linguists; as with any annotation task, acquiring new expert-annotated data can be time- and cost-prohibitive. Our work differs from these in that we annotate a corpus of documents containing more informal language—a collection of forum posts. Additionally, rather than relying on the availability of trained linguists to annotate the corpus, our work explores how we can use crowdsourcing to obtain hedge annotations.

To facilitate annotation by non-experts, we frame the annotation task as a word sense disambiguation problem rather than asking directly about hedging. Note that there is a precedent for reformulating hedge detection in this way: as a follow-up to the CoNLL-2010 hedge classification task, [Vellidal \(2011\)](#) described a new approach to classification in which hedge detection was viewed as a simple disambiguation task, restricted to words that have previously been observed as hedge cues. [Vellidal](#) transformed the CoNLL data for the binary classification task by defining the dictionary of potential hedge terms as any tokens that appeared as hedge cues in the training data; all unlabeled instances of these terms were assumed to be non-hedge usages. A classifier trained using this approach was found to outperform the systems presented at CoNLL-2010, which relied on standard methods of token-by-token or sentence-level classification. Our work extends the word sense disambiguation approach to the problem of obtaining hedging annotations on new corpora.

Crowdsourcing has been successfully used in the past for collecting annotations for word sense disambiguation. [Chklovski and Mihalcea \(2002\)](#) had users select the WordNet sense that most closely matched the definition of a word as used in a given sentence. Likewise, [Akkaya et al. \(2010\)](#) used Amazon Mechanical Turk (AMT) to annotate Subjectivity Word Sense Disambiguation (SWSD), a coarse-grained word sense dis-

### Relational Hedges

according to, appear, arguably, assume, believe, consider, could, doubt, estimate, expect, feel, find, guess, hear, I mean, I would say, imagine, impression, in my mind, in my opinion, in my understanding, in my view, know, likely, look like, looks like, may, maybe, might, my thinking, my understanding, necessarily, perhaps, possibly, presumably, probably, read, say, seem, seemingly, should, sound like, sounds like, speculate, suggest, suppose, sure, tend, think, understand, unlikely, unsure

### Propositional Hedges

a bit, a bunch, a couple, a few, a little, a whole bunch, about, allegedly, among others, and all that, and so forth, and so on, and suchlike, apparently, approximately, around, at least, basic, basically, completely, et cetera, etc, fair, fairly, for the most part, frequently, general, generally, in a way, in part, in some ways, kind of, kinda, largely, like, mainly, more or less, most, mostly, much, occasionally, often, partial, partially, partly, possible, practically, pretty, pretty much, probable, rarely, rather, really, relatively, rough, roughly, seldom, several, something or other, sort of, to a certain extent, to some extent, totally, usually, virtually

Table 1: List of (potential) hedge words and phrases

ambiguation task. In a much easier task, [Snow et al. \(2008\)](#) had users select from among three different senses of the word *president*. Our work follows these examples by presenting hedging and non-hedging definitions and asking users to choose between them.

## 3 Hedging Dictionary

We compiled a dictionary of 117 potential hedge words and phrases. We began with the hedge terms identified during the CoNLL-2010 shared task ([Farkas et al., 2010](#)), along with synonyms of these terms. This list was further expanded and edited through consultation with the LDC and other linguists, to ensure representation of hedge terms from more informal text.

The full list of hedge words and phrases in our dictionary is shown in [Table 1](#). This hedging dictionary is divided into *relational* and *propositional* hedges. As described in [Prokofieva and Hirschberg \(2014\)](#), relational hedges have to do

Hedge Term	Hedge Definition	Non-Hedge Definition
about	<ul style="list-style-type: none"> <li>almost; approximately (“There are <u>about</u> 10 million packages in transit right now.”)</li> </ul>	<ul style="list-style-type: none"> <li>on the subject of; concerning (“We need to talk <u>about</u> Mark.”)</li> <li>located in a particular area (“He is <u>about</u> the house.”)</li> <li>on the verge of (“He was <u>about</u> to leave.”)</li> </ul>
practically	<ul style="list-style-type: none"> <li>virtually; almost; nearly (“Their provisions were <u>practically</u> gone.” “It has rained <u>practically</u> every day.”)</li> </ul>	<ul style="list-style-type: none"> <li>in a practical manner; realistically; sensibly (“<u>Practically</u> speaking, the plan is not very promising.” “He purchased as many items as he could <u>practically</u> afford.”)</li> </ul>
suppose	<ul style="list-style-type: none"> <li>to believe or assume as true (“It is generally <u>supposed</u> that his death was an accident.”)</li> <li>to think or hold as an opinion (“I <u>suppose</u> the package will arrive next week.”)</li> </ul>	<ul style="list-style-type: none"> <li>to be expected or designed; to be required or permitted (“The machine is <u>supposed</u> to make noise.” “I’m <u>supposed</u> to call if I’m going to be late.”)</li> </ul>
think	<ul style="list-style-type: none"> <li>to have an opinion, belief, or idea about someone or something (“I <u>think</u> it’s an important issue.” “John doesn’t <u>think</u> he will win the election.”)</li> <li>to have as a plan or intention (“I <u>thought</u> that I would go.”)</li> </ul>	<ul style="list-style-type: none"> <li>to use one’s mind actively to form ideas (“<u>Think</u> carefully before you begin.” “I didn’t <u>think</u> of the solution in time.”)</li> <li>to direct one’s mind toward something or someone (“I was <u>thinking</u> about you.”)</li> </ul>

Table 2: Example definitions from our hedging dictionary

with the speaker’s relation to the propositional content, while propositional hedges are those that introduce uncertainty into the propositional content itself. The examples in (2) demonstrate relational and propositional hedges.

- (2) a. I **think** the ball is blue.  
b. The ball is **sort of** blue.

In (2a), *think* is a relational hedge. In (2b), *sort of* is a propositional hedge.

For each hedge term in our dictionary, we created definitions for the hedging and non-hedging usages of the term, including examples for each case. We attempted to keep these definitions as simple as possible while still providing enough direction for workers completing the AMT task. These definitions were revised as we tested the AMT task with real-world users and received feedback pointing out ambiguities or other problems with the definitions. We did find that some words were too complicated or that the differences in senses was too nuanced to reduce definitions to short hedge and non-hedge definitions: in par-

ticular, *hear*, *read*, and *say* were identified as such. For example, the sentences in (3) differ only slightly, but *hear* is being used a hedge in the first and not in the second:

- (3) a. I **heard** that there was an arrest.  
b. I **heard** about the arrest.

For these words, it might be more effective to develop a separate AMT task that provides a more comprehensive set of definitions and examples rather than trying to reduce them to a simple binary choice. Another option would be to ask AMT workers more directly about how the speaker is using a term: e.g. whether the usage reflects uncertainty or lack of commitment to a proposition.

Table 2 shows some examples of hedging and non-hedging definitions. The complete dictionary of hedge terms, definitions, and examples is available from the authors upon request. Note that for 34 entries in our dictionary, the *non-hedge* definition is simply “Other”. These are cases where the word or phrase is generally unambiguous except for extremely rare instances (generally, typos

## Instructions

1. Read the sentence and answer the question based on the selected word.
2. Select the definition and example pair that most closely matches the meaning of the selected word as it is used in that sentence.
3. If one of the answers is 'Other', select it only if the alternative option doesn't apply at all.
4. Please answer carefully. Each HIT may contain one or more questions to which we know the correct answer; if you answer these incorrectly, we will reject the HIT. **If you have questions or feel your work was rejected unfairly, please contact us.**

Figure 1: Instructions for Amazon Mechanical Turk Task

4. It seems a bit silly now but I was **fairly** young when we met and this seemed to so important to me.

Is the meaning of the word *fairly* closer to:

- justly or honestly, impartially ("The judge decided fairly.")
- to quite a high degree ("I was fairly certain she had nothing to do with it.")
- somewhat ("This judgment passed down fairly recently.")

Figure 2: Example of AMT word sense disambiguation task

or other errors).

## 4 Corpus Annotation

We began with a collection of discussion forum posts from the 2014 Deft Committed Belief Corpora (Release No. LDC2014E55, LDC2014E106, and LDC2014E125). These posts were originally collected for the DARPA BOLT program and were selected according to a variety of criteria, including that the posts should contain primarily informal discussion and that the main focus of the threads should be discussion of dynamic events or personal anecdotes (Garland et al., 2012).

We located all instances of the hedges from our dictionary in these corpora and presented each of these instances as a potential hedge to workers on AMT. The hedge term was shown as a highlighted word or phrase within a sentence; below this sentence, we displayed definitions and examples of the hedging and non-hedging uses of the term. We asked workers which definition they felt most closely matched the meaning of the word highlighted in the sentence. To avoid bias based on the placement of the choices, we varied the order in which the hedging and non-hedging definitions appeared. Each Human Intelligence Task (HIT) asked for judgments on 10 sentences, with one being a gold-standard check judgment. If the worker failed to answer the check judgment correctly, we discarded their data and republished the

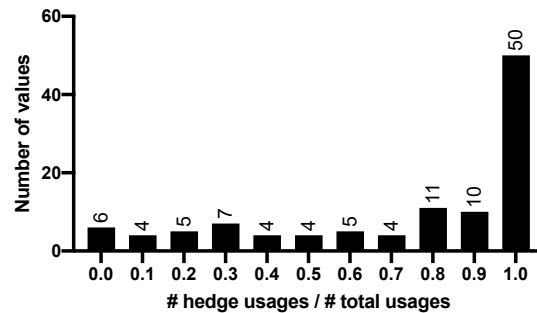


Figure 3: Distribution of proportion of hedge usages out of all occurrences of each term.

HIT. We obtained 5 judgments for every potential hedge word and picked the majority vote as the label for that instance. Figure 1 shows the instructions given to workers. An example of the task for the word *fairly* is shown in Figure 2.

The resulting corpus has a total of 20,683 annotated potential hedge terms, although the data set is very unbalanced, with some hedge terms appearing many more times than others. For example, *about* appears 2,124 times but *in some ways*, *et cetera*, and *to a certain extent* each appear only once. The number of hedge usages vs. non-hedge usages for each term also varied. Figure 3 shows the distribution of the proportion of times a term was used a hedge out of all occurrences of that term. Overall agreement among the AMT workers was 92.89%, with Fleiss' Kappa equal to 0.751.

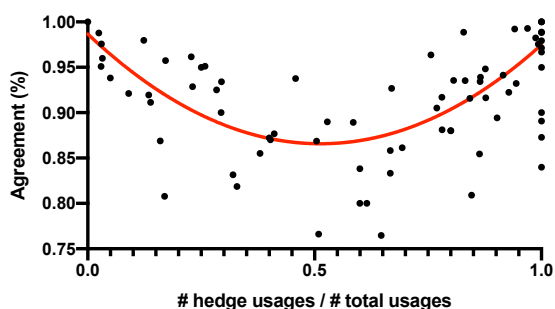


Figure 4: Plot of agreement vs. proportion of hedge usages out of total occurrences for each term.

The agreement varied depending on the hedge term. Figure 4 shows a scatterplot of the agreement percentage vs. how often each term is used a hedge. As one might expect, the general trend shows that agreement is higher for terms that are almost always used as hedges (or as non-hedges) than for the more ambiguous terms.

To get a sense of the quality of the crowdsourced judgments, we annotated a subset of the corpus ourselves. This subset was constructed by randomly selecting two instances for each hedge term. Each instance received two judgments, one by each of the two authors of this paper. As one would expect, inter-annotator agreement was higher, 94.73% overall, with Cohen’s Kappa equal to 0.857. For most hedge terms, agreement was 100%; however, 11 hedge terms had an agreement of 50%. We adjudicated the questions for which we disagreed to create a single gold standard answer. We then compared our gold standard answers for this subset to the majority vote judgments obtained from AMT workers for the same questions. The crowdsourced majority vote judgment differed from the gold standard on only 7 questions, for an overall accuracy of 96.65%.

## 5 Summary

We have described a new method of using crowdsourcing to annotate a corpus with hedging information, by framing the hedge detection task as a word sense disambiguation problem. We have used this method to annotate a corpus of forum posts, which we hope to make generally available through the LDC. We have shown that annotations obtained using this method can in fact be very accurate; when comparing the crowdsourced judgments to an expert-annotated subset of the corpus, we obtained an accuracy of 96.65%.

## Acknowledgments

This paper is based upon work supported by the DARPA DEFT program. The views expressed here are those of the author(s) and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

## References

- Cem Akkaya, Alexander Conrad, Janyce Wiebe, and Rada Mihalcea. 2010. [Amazon Mechanical Turk for Subjectivity Word Sense Disambiguation](#). In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon’s Mechanical Turk*, pages 195–203, Los Angeles. Association for Computational Linguistics.
- Timothy Chklovski and Rada Mihalcea. 2002. [Building a Sense Tagged Corpus with Open Mind Word Expert](#). In *Proceedings of the ACL-02 Workshop on Word Sense Disambiguation: Recent Successes and Future Directions*, pages 116–122. Association for Computational Linguistics.
- 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*, pages 1–12, Uppsala, Sweden. Association for Computational Linguistics.
- Jennifer Garland, Stephanie Strassel, Safa Ismael, Zhiyi Song, and Haejoong Lee. 2012. [Linguistic resources for genre-independent language technologies: User-generated content in BOLT](#). In *@NLP Can u Tag #user\_generated\_content ?! Via Lrec-Conf.Org (NLP4UGC 2012)*.
- Anna Prokofieva and Julia Hirschberg. 2014. [Hedging and speaker commitment](#). In *Proceedings of the 5th International Workshop on Emotion, Social Signals, Sentiment & Linked Open Data*, pages 10–13, Reykjavik, Iceland.
- Rion Snow, Brendan O’Connor, Daniel Jurafsky, and Andrew Ng. 2008. [Cheap and Fast – But is it Good? Evaluating Non-Expert Annotations for Natural Language Tasks](#). In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pages 254–263, Honolulu, Hawaii. Association for Computational Linguistics.
- Erik Velldal. 2011. [Predicting speculation: A simple disambiguation approach to hedge detection in biomedical literature](#). *Journal of Biomedical Semantics*, 2(5):S7.
- Veronika Vincze, György Szarvas, Richárd Farkas, György Móra, and János Csirik. 2008. [The Bio-Scope corpus: Biomedical texts annotated for uncertainty, negation and their scopes](#). *BMC Bioinformatics*, 9(11):S9.