

# 1 Hedging

This section includes guidelines on the annotation of hedges, defined as speculative cues, and their respective scopes. The most typical hedges are listed below and their scopes are illustrated with examples. (Note: These guidelines are adapted from the CONLL 2010 Shared Task guidelines).

Hedges are words “whose job it is to make things fuzzier” (Lakoff, 1972). More specifically, they are single cue words or combinations of words that are used by a speaker to mitigate the strength of their utterance. Their use has been correlated with many discourse functions, such as trying to save face (Prince et al, 1980), indicating politeness (Ardissono et al, 1999) and cooperative intent (Vasilieva, 2004), as well as attempting to evade questions and avoid criticism (Crystal, 1987). In general, hedging can be seen as a manifestation of the speaker’s attitude towards a claim and towards their audience (Isabel, 2001). The use of hedge words (or the lack thereof) can shape an audience’s opinion of the speaker and of their argument (Blankenship and Holtgraves 2005; Hosmon and Siltamen 2006; Erickson et al, 1978). As such, locating hedge words and identifying their scope (that is, the propositional content that is being ‘hedged’) may help us identify when such dialogue actions are taking place.

## 1.1 Hedges: What to Annotate

Only sentences with some instance of speculative language are to be annotated. If a sentence does not include any speculative element or any element that refers to uncertainty (i.e. it contains only a statement of fact), the absence of hedging behavior need not be indicated. Note that not all speculative language is considered hedging. Hypotheticals such as *‘If it rains, I won’t go to the game’* contain instances of speculative language (if/then) but are not considered instances of hedging behavior.

In general, when in doubt over whether something is a hedge or not, the following questions can be asked:

- Is the speaker being deliberately uninformative (or under-informative)?
- Is the speaker uncertain?
- Is the speaker trying to downplay the force of their utterance?

If the answer is yes to any of these, then there is a much higher likelihood that the utterance under consideration contains a hedge.

In this task, we annotate both the **hedge cues** (a word or words that signal the presence of a hedge) and their **scopes** (the content that is being hedged by each hedge cue).

### 1.1.1 Hedge Cues

A hedge cue can be a single word or a combination of multiple words that signal uncertainty, a lack of precision or non-specificity, or an attempt to soften or downplay the force of the speaker’s utterance.

I **guess** John's right.

Jane was **probably** drunk.

I **think** it's an important issue.

So that **may** be an effect of bioterrorism in the United States.

It is **largely** known that this university has a good reputation and an excellent track record.

In each of the examples above, the word in bold is a hedge cue. More details about hedge cues and how to identify them are provided in sections 1.1.3 through 1.1.5, as well as section 1.2.

### 1.1.2 Scope

The scope of a hedge cue is the propositional content that is the subject of the speaker's speculative language. The scope of a hedge is the object of the speculation – the material over which the hedge can be interpreted. When determining scope, it may be useful to ask: '<hedge word> what?' What is the speaker hedging about? For example, in '*I think that statement is untrue,*' (where 'think' is a relational hedge) the scope of the hedge may be determined by asking 'I think *what?*' Note that the hedge word itself is always included within its scope, while the subject of the sentence usually is not.

The scope of a hedge can also be determined in part from the utterance's syntax. Generally speaking, the scope should include the hedge and should extend to the end of the smallest syntactic unit which contains the proposition being hedged – a clause or a noun phrase, for example.

I **guess** *John's right.*

Jane was **probably** *drunk.*

I **think** *it's an important issue.*

So that **may** *be an effect of bioterrorism in the United States.*

In the examples above, and those in the rest of this document, the hedge cue is indicated by text in bold and the scope of each hedge is indicated by the text in italics.

More detail on determining scope is provided in section 1.3.

### 1.1.3 Hedge Types: Relational and Propositional

We categorize hedges themselves into two types, based on Prince et al's (1980) definition: *relational hedges* that have to do with the speaker's relation to the propositional content, and *propositional hedges* that introduce uncertainty into the propositional content itself.

A hedge cue marks a **relational hedge** when there is uncertainty in the commitment of the speaker towards the proposition, as in:

I **guess** John's right.

Jane was **probably** drunk.

In the first case, the speaker is uncertain about whether John is right. In the second example, the speaker indicates that Jane may or may not have been drunk, and although the speaker is leaning towards the former, they are still uncertain.

On the other hand, a hedge cue marks a propositional hedge when there is uncertainty about some part of the propositional content itself, as, for example, in:

It's **kind of** hard to read them straight up and down like that.

In this example, there is some difficulty in accomplishing the task of reading but it is the degree of difficulty rather than its existence that is in question.

Propositional hedges create uncertainty in the propositional content by marking non-prototypicality with respect to class behavior (as in 'His feet are **sort of blue**', where the color of the feet is being marked as not a prototypical shade of blue). They can also be used to introduce fuzziness into the degree or quantity of an action (as in 'It's **kind of hard**' and '**Sometimes**, it's difficult').

One way to tell propositional and relational hedges apart is that one can insert "I'm certain" before a sentence containing a propositional hedge without changing the meaning of the sentence, as in:

I'm certain (that) ... his feet are **sort of blue**

because propositional hedges do not imply uncertainty on the part of the speaker. Such an insertion is less plausible before a sentence containing a relational hedge:

#I'm certain (that) ... I **guess** John's right.

One type of relational hedge for which the above test does not work is the *attributive hedge* – when the speaker attributes information to some other source in order to downplay its force (as in 'I heard someone say ...') or to garner authoritative power for their statement (as in 'Well, the Encyclopedia Britannica says that ...'), rather than committing to the proposition themselves. 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 their utterance.

I **read** that this place is bad.

**They say** it's impossible to get a job there.

In the first example, 'read' is a hedge because it attributes information to some written source. In the second example, 'They say' is marked as a multi-word hedge, because although 'say' in and of itself is not a hedge word, the attribution of a proposition to someone else is signaled by the presence of a subject other than a first person pronoun. When the subject **does** represent the speaker, for example, 'I say it's impossible to get a

job there' there is, in contrast, no hedge.

### 1.1.4 Typical Relational Hedges

Many words may be ambiguous between a hedge use and a non-hedge use depending on the context in which the word appears. Some of the words which may signify relational hedges include the following:

Typical Relational Hedge Cues				
	Hedge		Non-Hedge	
	Definition	Example	Definition	Example
<b>Verbs</b>				
think	to believe	I <b>thought</b> you were wrong . I <b>believe</b> I met him at last year's picnic.	to come up with	I thought of the solution.
believe	to think, to hold as an opinion	I <b>consider</b> the failure to be intentional.	to accept as true/truthful	I believe you. We can consider the case based on this cause of action.
consider	to regard as, to think/believe	I <b>assumed</b> this position was temporary .	to reflect on, to pay attention to to take on, to take over, to take upon oneself	I assumed this position upon his retirement.
assume	to take without evidence	I <b>understand</b> it to be in the range of \$30-40 million.	to comprehend	I understand what you are saying.
understand	to perceive, to hold a point of view	I <b>found</b> it curious that you were objecting to it .	to locate	I found my keys under the couch.
find	to perceive	I <b>felt</b> it was uncalled for .	to sense/be affected by	I felt sick afterwards.
feel	to think	I <b>don't know</b> if that's a good idea .	not possess information	I don't know her phone number.
don't know	not sure	The problem <b>appears</b> to be a bug in the software .	to come into sight, to be part of a stage performance	A bird I didn't recognize appeared in our yard on Sunday.
appear (also seem, look like, etc.)	to give the impression of being	I <b>suppose</b> the package will arrive next week .	to be required/obligated	I'm supposed to call if I'm going to be late.
suppose	to assume, to take without evidence			
guess, estimate, speculate, etc.	to form a theory or conjecture to cause one to think that something is the case, to put forward for consideration	I <b>guess</b> they're not coming .  The results <b>suggest</b> that the procedure is effective .	N/A  N/A	
<b>Auxiliaries</b>				
may	expressing possibility	That <b>may</b> be an effect of bioterrorism .	expressing permission	The action may be brought at any time within two years.
could	expressing possibility	You <b>could</b> be right .	having the ability to	I could touch my toes easily when I was younger.
should	expressing likelihood	It <b>should</b> be rainy tomorrow . It <b>might</b> be a problem with the server .	expressing duty/obligation	A good system should be able to handle any airport.
might	expressing possibility		expressing advisability	You might just give a very short explanation of what it is.
<b>Adjectives and Adverbs</b>				
not necessarily, surely, probably, likely, maybe, perhaps, unsure, etc.	expressing uncertainty	That is <b>not necessarily</b> the case .	N/A	

#### Notes:

1. Verbs are not marked as hedges if they follow the pronoun 'you' and are being used in a rhetorical sense (eg. 'So you think they would announce it on the news, but they didn't.') If the pronoun 'you' can be replaced by 'one', then it is usually not a case of hedging.
2. There are many difficult cases involving 'should': for example, 'They should be entitled to that right' could mean either 'they aren't but they ought to be' or 'it's not quite certain whether they are or not'. When the context does not disambiguate, one should annotate 'should' as a hedge.
3. *would* (simple past tense and past participle of *will*; used to express the future in past sentences; used in place of *will* to make a statement less blunt; used to express repeated action in the past; used to express an intention or inclination) is difficult to classify as a hedge. Annotators should **not** mark 'would' as a hedge.

This is not intended to be an exhaustive list, but is presented here simply to give the

annotator an indication of words that are often used as hedges and ways to disambiguate hedge uses from non-hedge uses.

### 1.1.5 Typical Propositional Hedges

Propositional hedges show uncertainty within the propositional content. Some of the typical propositional hedges are adjectives and adverbs that show uncertainty by conveying a lack of precision or non-specificity to frequency, quantity or degree:

1. frequency: *generally, often, rarely, sometimes, frequently, occasionally, seldom, usually*
2. quantifiers: *not much, most, a whole bunch/a bunch, several, a couple, a few, a little/a little bit*

In general, if a quantifier is an intensifier (eg. *a lot, much, etc*), it is less likely to be a hedge (and may only be a hedge if used in a negated sense – eg. *not much*)

3. degree: *almost, practically, apparently, virtually, basically, approximately, roughly, somewhat, somehow, partially, not really*

*Note: sort of/kind of* is marked as a hedge when it conveys inexactness or vagueness, as in

This is a graph that **sort of** compares the performance of speech recognition in 2000 and 2001.

It is not considered a hedge when it means ‘type of’, as in

It is the sort of thing that it would be nice if it worked.

4. non-specificity: *some/somewhat/someone/somehow, etc.*

*Note:* marked as a hedge if the speaker is uncertain or appears to be deliberately underspecifying, but not if it can be construed as a grammatical necessity or when specification would not be necessary in the context of the conversation. So in (a) *somebody* is marked as a hedge and in (b) it is not.

- a. Well, **somebody** already answered those questions um I won't say who.
- b. It would be so easy for somebody to disturb something here.

## 1.2 How to Annotate Hedge Cues

Because hedge cues can consist of one or more words, we mark the beginning and end of each hedge cue with a tag - <hRel> and </hRel>, <hProp> and </hProp>. Each tag has an id number that corresponds to the number of the hedge, counted from the beginning of the document. Each hedge cue has an associated scope; each pair (hedge, scope) will be uniquely identified with an id number. The scope of each hedge is marked with <xcope> and </xcope> tags (more detail about how to determine the scope is

provided in section 1.3). The hedge itself is always included in its scope:

But I recall you saying **several things** that **seemed very important** that weren't listed on the slide.

But I recall you saying <xcope id="1"><hProp id="1">**several**</hProp> things</xcope> that <xcope id="2"><hRel id="2">**seemed**</hRel> very important</xcope> that weren't listed on the slide.

All of the examples in this section and its subsections will contain the example sentence followed by an annotated version of that sentence.

### 1.2.1 Multi-Word Hedges

Hedges may also be comprised of more than one word. There are several different cases to consider when annotating these:

- 1) a sequence of words expresses uncertainty together, but none of the words do so separately;

*They had National Guards around **and all that**.*

<xcope id="3">They had National Guards around <hProp id="3">**and all that**</hProp></xcope>.

The phrase '*and all that*' is marked as a multi-word hedge, although none of the words individually would be considered hedges. Other such multi-word hedges include 'and so forth' and 'et cetera'. Complex hedges also include constructions that derive from the verbs mentioned above and imply a personal view as opposed to a statement of fact, such as: 'in my mind', 'in my opinion', 'in my understanding', 'my thinking is', 'my understanding is', 'in my view', 'if I'm understanding you correctly', and so on.

- 2) some of the words within the complex hedge can express uncertainty by themselves, but not all do;

In the case where one or more (but not all) words in the complex hedge *can* express uncertainty by themselves, but the complex phrase also expresses uncertainty, the words in the phrase which can separately be used as hedges should be marked separately, as well as part of the complex hedge. The following examples illustrate case (2):

Well, **somebody said** the answer to the question is no.

Well, <xcope id="5"><hRel id="5"><xcope id="4"><hProp id="4">**somebody**</hProp></xcope> **said**</hRel> the answer to the question is no</xcope>.

Oh ~HEPA? *High efficiency particular **something or other**.*

Oh ~HEPA? <xcope id="7">High efficiency particular<hProp id="7"><xcope id="6"><hProp id="6">**something**</hProp></xcope> **or other**</hProp></xcope>.

In the first example, the use of 'somebody' by itself is a propositional hedge, but 'somebody said' is a relational (attributive) hedge as well; as such, we mark both 'somebody said' as a hedge and 'somebody' as a nested hedge within the multi-word hedge. Similarly, in the second example, 'something' is a hedge itself (since the speaker is conveying their uncertainty about what the 'A' stands for in HEPA), but 'something or other' is also a multi-word hedge; both are marked.

- 3) each of the words, or subsequences of words, can express uncertainty by itself.

Complex hedges are not to be confused with two or more sequential hedges, each of which expresses uncertainty on its own. For example, 'so far' and 'at least' are marked separately in the following example of case (3), since each can stand alone as a hedge term:

Well *it hasn't, **so far at least**, hasn't affected me at all.*

Well <xcope id="8">it hasn't, <xcope id="9"><hProp id="8">**so far**</hProp> <hProp id="9">**at least**</hProp></xcope>, hasn't affected me at all</xcope>.

When marking multi-word hedges, annotate the minimal unit that expresses uncertainty as the hedge. In other words, a phrase should not be marked as a complex hedge if only a single word in that phrase expresses the speculative content, independent of the other word(s) in the phrase. On the other hand, if the entire phrase is required to express the speculative content, the entire phrase should be marked as a hedge.

### 1.2.2 Negated Hedges

Note that while we do *not* distinguish negated hedges from non-negated hedges, if there is a negating particle present, it should be included within the hedge itself (eg. 'not much', 'weren't really' for propositional hedges, 'don't know' for relational hedges, and so on). If the negating particle is contracted, the verb that it is attached to is included within the hedge cue.

He **wasn't really** clear on what the assignment was.

He <xcope id="10"><hRel id="10">**wasn't really**</hRel> clear</xcope> on what the assignment was.

### 1.2.3 Hedges in Questions

Due to the inherent uncertainty that the question itself conveys, it can be difficult to detect hedges in questions. It is however possible to annotate hedges within questions that appear to be independent of the overall uncertainty the speaker is conveying via the question. For example:

What about the argument that *the plaintiff **may not** have been harmed by the disclosure?*

What about the argument that *the plaintiff* **may not** *have been harmed by the disclosure*?

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

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

In the first example, 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 the second example, 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.

The following represent cases of questions in which words that might be hedges in statements would not be marked as hedges:

Do you think it’s wrong?

Could you clarify this for me?

In both examples, it is impossible to distinguish between the uncertainty introduced by the potential hedge word (‘think’, ‘could’) and the uncertainty inherent in the question itself. Hence, we do not mark these words as hedges.

Also, 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 question.

It **might** rain, might it not?

It *might* rain, might it not?

This is because ‘might’ in the tag is grammatically necessary to the creation of the tag question itself.

#### 1.2.4 Marking Disfluent Hedges

In spoken language, utterances may be disfluent, containing *filled pauses* (‘um’, ‘er’) and *self repairs* (‘I th- think...’).

I **think** it’s – I **think** it’s an extremist group that’s trying to make us move faster.

I *think* it’s – I *think*



*id="15"><hRel id="15">think</hRel> it's an extremist group that's trying to make us move faster</xcope>.*

Here, the first production of 'think' is in the *reparandum* (the portion of the utterance that will subsequently be repaired) and the second is in the *repair* (the repaired utterance). Both occurrences of 'think' should be marked as hedges.

Elements of a hedge phrase may also be separated by disfluencies or filler phrases like "you know" such that the entire disfluent or filler word or phrase is preceded and followed by words of the hedge phrase. In such cases the entire phrase including the disfluency or filler should be included in the hedge phrase.

*They had National Guards around and you know all that.*

*<xcope id="16">They had National Guards around <hProp id="16">and you know all that</hProp></xcope>.*

*They had National Guards around and um all that.*

*<xcope id="17">They had National Guards around <hProp id="17">and um all that</hProp></xcope>.*

### 1.3 How to Annotate Scope

The scope of a hedge is the object of the speculation – the material over which the hedge can be interpreted. For relational hedges the object of speculation is a proposition. For propositional hedges, the object of speculation can be an action, attribute or object. When determining scope, it may be useful to ask: '<hedge word> what?' What is the speaker hedging about? For example, in 'I think that statement is untrue,' (where 'think' is a relational hedge) the scope of the hedge may be determined by asking 'I think what?' Note that the hedge word itself is always included within its scope, while the subject of the sentence usually is not.

The scope of a hedge can also be determined in part from the utterance's syntax. Generally speaking, the scope should include the hedge and should extend to the end of the smallest syntactic unit which contains the proposition being hedged – a clause or a noun phrase, for example.

In the case of auxiliary and main verbs, adjectives and adverbs, the scope would normally start at the hedge word, and continue to enclose the rest of the proposition that is the subject of the speculation. In the case of verbal elements, the scope typically ends at the end of the current clause or sentence; thus, all complements and adjuncts are included in the hedge's scope.

*I think it's an important issue.*

*I <xcope id="18"><hRel id="18">think</hRel> it's an important issue</xcope>.*

*So that may be an effect of bioterrorism in the United States.*

So that <xcope id="19"><hRel id="19">**may**</hRel> be an effect of bioterrorism in the United States</xcope>.

It is **largely** known that this university has a good reputation and an excellent track record.

It is <xcope id="20"><hProp id="20">**largely**</hProp> known that this university has a good reputation and an excellent track record</xcope>.

If a verb or adverb appears at the end or within the clause or sentence, the scope is marked over the entire proposition that is being hedged, even if this proposition includes words that precede the hedge word. So, in the first example below, the scope includes the clause preceding the hedge word 'think' and in the second, it includes the entire sentence before the hedge word 'basically'.

*There are other considerations here, I **think** that need to be addressed.*

<xcope id="21">There are other considerations here, I <hRel id="21">**think**</hRel> that need to be addressed</xcope>.

Consider the sentence as: I **think** there are other considerations here that need to be addressed.

*So you add up your your ~R squareds and whatever's left is unexplained variability, **basically**.*

So <xcope id="22">you add up you're your ~R squareds and whatever's left is unexplained variability, <hProp id="22">**basically**</hProp></xcope>.

Consider the sentence as: So **basically** you add up you're your ~R squareds and whatever's left is unexplained variability.

The scope of attributive adjectives (those that ascribe a quality to a noun and appear just before it) generally extends to the following noun phrase. However, the scope of predicative adjectives (complements of the verb) includes the entire clause or sentence, as can be seen below.

It's a case of assault and **possible** theft.

It's a case of assault and <xcope id="23"><hRel id="23">**possible**</hRel> theft</xcope>.

*The decline of profits due to the negative reviews of the product is **possible**.*

<xcope id="24">The decline of profits due to the negative reviews of the product is <hRel id="24">**possible**</hRel>.

Sentential adverb hedges (those that modify the main verb in the sentence or clause) have scope over the rest of that entire sentence, while the scope of other adverbs usually begins with the hedge word and ends at the end of the clause.

The popular opinion **probably** affects the success of a political candidate.

The popular opinion <xcope id="25"><hRel id="25">**probably**</hRel> affects the success of a political candidate</xcope>.

He had a cough and **probably** pneumonia.

He had a cough and <xcope id="26"><hRel id="26">**probably**</hRel> pneumonia</xcope>.

Noun and pronoun hedges (the most common examples being 'somebody' and 'something') have a scope that extends over just the noun phrase that contains the hedge. In most cases, the scope will extend over just the hedge itself, as in:

I can't give you the document because um I put it down **somewhere**.

I can't give you the document because um I put it down <xcope id="27"><hProp id="27">**somewhere**</hProp></xcope>.

If the hedge noun or pronoun has a complement, then that is included in the scope as well:

And I saw **somebody** who shouldn't have been there walking down the hall.

And I saw <xcope id="28"><hProp id="28">**somebody**</hProp> who shouldn't have been there</xcope> walking down the hall.

### 1.3.1 Scope and the Passive Voice

In active sentences, the scope of a hedge usually begins with the hedge and excludes the subject, except in certain cases identified above. However, for passive sentences, the subject is treated as the object of the verb. So the subject is typically included in the scope of the hedge in passive sentences where it would not be in active ones.

*Dormant anger issues have also been **suggested** as a factor.* (passive)

<xcope id="29">Dormant anger issues have also been <hRel id="29">**suggested**</hRel> as a factor</xcope>. (passive)

John has also **suggested** dormant anger issues as a factor. (active)

John has also <xcope id="30"><hRel id="30">**suggested**</hRel> dormant anger issues as a factor</xcope>. (active)

In a similar manner, relative pronouns such as 'which' and 'what' are also included in the scope of the auxiliary hedge in the case of passive voice.

This will go on his academic record, *which **may** be a problem during the college application process.*

This will go on his academic record, <xcope id="31">which <hRel id="31">**may**

</hRel> *be a problem during the college application process*</xcope>.

What we **assumed** was that you would finish this last month.

<xcope id="32">What we <hRel id="32">**assumed**</hRel> was that you would finish this last month</xcope>.

In the case of *raising verbs* such as *seem*, *appear*, *be expected*, *be likely*, which can take 'it' as their subject, there are two different syntactic patterns with respect to the scope of these hedges:

It **seems** that the treatment is successful.

It <xcope id="33"><hRel id="33">**seems**</hRel> that the treatment is successful</xcope>.

The treatment **seems** to be successful.

<xcope id="34">The treatment <hRel id="34">**seems**</hRel> to be successful</xcope>.

In the first case, where the verb takes a sentential complement (that is, a clause), the scope of *seems* begins with the verb. In the second, in which the verb takes an infinitival complement, the scope should include the subject.

### 1.3.2 Marking Scopes of Disfluent Hedges

In the case of disfluencies, the scope continues through the completion of the speaker's intended utterance as in:

**Their impression** is no State has a statue of limitation that is longer I'm sorry no State has a record keeping requirement that exceeds the federal record keeping requirement.

<xcope id="35"><hRel id="35">**Their impression**</hRel> is no State has a statue of limitation that is longer I'm sorry no State has a record keeping requirement that exceeds the federal record keeping requirement</xcope>.

If the hedge word itself (or a portion of it) is repeated in both the reparandum and the repair, both instances are marked; the scope of the first extends only to the end of the reparandum, while the scope of the hedge word in the repair extends until the end of that utterance.

I **think** it's – I **think** it's an extremist group that's trying to make us move faster.

I <xcope id="36"><hRel id="36">**think**</hRel> it's</xcope> --- I <xcope id="37"><hRel id="37">**think**</hRel> it's an extremist group that's trying to make us move faster</xcope>.