

Annotating Belief in Communication: Manual

Eric Davis, Mona Diab, Lori Levin, Teruko Mitamura, Owen Rambow

November 29, 2007

Abstract

This is an annotation manual for communicated belief.

1 Introduction: Goal of Annotation

We annotate whether the speaker/writer (SW) intends the hearer/reader (HR) to interpret a stated proposition as SW's strongly held belief, as a proposition which SW does not believe strongly (but could), or as a proposition towards which SW has an entirely different cognitive attitude, such as desire or intention.

- The nature of the proposition (opinion, statement about interior state, external state) is not of interest.
- We do not annotate truth: real-world (encyclopedic) truth is not relevant.

We have three categories:

- Committed belief: the speaker or writer indicates in this utterance that he or she believes the proposition. For example, *I know Mark and Sandra have eloped.*
A subcase of committed belief concerns propositions about the future, such as *The sun will rise again.*
- Non-committed belief: the speaker or writer identifies the proposition as something which he or she could believe, but he or she happens not to have a strong belief in the proposition. For example, *Mark and Sandra may have eloped.*
A subcase of non-committed belief concerns propositions about the future, such as *John may return tomorrow.*

- Not applicable: for the speaker or writer, the proposition is not of the type in which he or she is expressing a belief, or could express a belief. Usually, this is because the proposition does not have a truth value in this world (be it in the past or in the future). For example, *I wish Mark and Sandra would finally elope.*

The interest of the annotation is clear: we want to be able to determine automatically from a given text what beliefs we can ascribe to the author, and with what strengths he or she holds them. Across languages, many different linguistic means are used to denote this attitude towards an uttered proposition, including syntax, lexicon, and morphology. To our knowledge, no systematic empirical study exists for English or Arabic, and this annotation is a step towards that goal.

2 Annotation Units

Annotation units for factivity:

- Sentence? (No)
- Proposition? (from PropBank)
- Clause?
- Event?
- Nominals? (from NomBank?)

Examples:

- (1) Losing my money made me mad (Two propositions according to PropBank)
- (2) Losing money would make me mad (Two propositions according to PropBank)
- (3) The loss of my money made me mad (*loss* might not be included in PropBank)
- (4) The hunting of deer is encouraged (*hunting* might not be included in PropBank)

DECISION: The annotation units will be the propositions identified by PropBank and NomBank.

English annotation units: Use PropBank coding manual.

ACTION ITEM (Owen and Mona): Ask Martha and Adam if they can quickly identify the annotation units in our corpus (English and Arabic).

2.1 Multi-word expressions

Some things that appear to be sentences containing verbs are actually frozen or semi-frozen multi-word expressions. We have made the following decisions so far:

- *Tell you what*: NA
- *Go ahead*: NA
- *How's that*: NA
- *I say*: as in *I say wait*: CB
- *As you know*: CB (Eric's test: this can be modified by adverbs as in *as you probably know*. The adverbs can change it from CB to NCB.)

Another construction we discuss is the correlative as in *The more we know, the better*. The speaker has a committed belief about the correlation between *more* and *better*, but does not have a committed belief about *know*. Since there is no single word or contiguous string of words that express the correlation, we decided to annotate *more* and *better* as CB, but maybe with some shading instead of a solid color. *Know* will be annotated as NA.

2.2 Auxiliary Verbs

Aspectual auxiliary verbs are not annotated. Modal auxiliary verbs (may, might, can, could, should, need to, ought to, have to) are annotated as CB. Their complements are annotated as NA. *Will* is not annotated. Its complement is NCBF.

3 Nested Sources of Propositions: Reported Speech Acts

Here we can take Sec 2.2 from UPitt TR-02-11 (Jan), but we need to modify examples, and explain what we want annotated.

4 Annotating Belief

etc)

We can use the following diagnostics:

- **I am certain that...:** If the main proposition can be embedded under *I am certain that...* (from the point of view if the WS), then that proposition should be labeled CB. For example:

- (5) a. The moon is made of cheese
 b. I am certain that the moon is made of cheese

In (5) above, in some contexts (a) expresses the same as (b). If determined that this is the case in this context, label the *be* proposition as CB.

- **I am not sure but think that...:** If the main proposition can be embedded under *I am not sure but think that...* (from the point of view if the WS), then that proposition should be labeled NCB. For example:

- (6) a. Probably, the moon is made of cheese
 b. I am not sure but I think that the moon is made of cheese

In (6) above, in some contexts (a) expresses the same as (b). If determined that this is the case in this context, label the *be* proposition as NCB.

- **I have no opinion on this, but someone claims that...:** If the main proposition can be embedded under *I have no opinion on this, but someone claims that...* (from the point of view if the WS), then that proposition should be labeled NCB. For example:

- (7) a. The AP reported that Iraq was bombed
 b. I have no opinion on this, but someone (AP) claims that Iraq was bombed

In (7) above, in some contexts (a) expresses the same as (b). If determined that this is the case in this context, label the *be bombed* proposition as NCB.

- **Matrix clause to adverb:** If you can transform the matrix clause of the target proposition into an adverb which expresses some degree of certainty, then the proposition should be labeled NCB. For example:

- (8) a. I expect John to arrive shortly
 b. John will most probably arrive shortly

In (8) above, in some contexts (a) expresses the same as (b). If you determine that this is the case in this context, label the *arrive* proposition as NCB.

- (9) a. I hope John will arrive shortly
b. # John will probably/perhaps/maybe/certainly arrive shortly

In (9) above, the meaning of (a) cannot be conveyed by (b), so you cannot assume that the *arrive* proposition is an NCB. Instead, use other diagnostics; in this case, because the matrix verb is *hope*, it is an NA.

Here are some specific rules:

- **Questions:** The main proposition in a question (*wh*- or yes/no) is always NA, since it is being queried. This does not mean that any embedded clauses are also automatically NA.

(10) Did you regret giving up your kingdom for a horse?

In (10) above, the *regret* proposition is NA, while the *giving up* proposition is CB, since the speaker clearly believes that the hearer did indeed give up her kingdom.

- **Matrix verbs of desire, hope, and longing:** Verbs which express a state that the WS wish were true (past, present, or future) are marked as NA. Example: The complements of the verbs *to hope* and *to wish*.
- **Matrix verbs of expectation:** The complements of the matrix verbs *expect*, *wait*, and *try* are NCB.
- Verbs in **purpose clauses** are marked as NA. To determine if a clause is a purpose clause, replace *to* with *in order to*. If the utterance as a whole is still grammatical, the clause is a purpose clause.

- (11) a. I am writing to ask you for money
b. I am writing in order to ask you for money

In (11) above, *to ask for* is a purpose clause, as it could successfully be replaced by *in order to ask for*. Since this is the case, the *ask for* is labeled NA.

4.1 Presuppositions

Presuppositions should be annotated as CB. The test for presupposition is that it is still asserted when the sentence is negated. The in the classic example below, *beating your wife* is presupposed when the sentence is positive and when it is negative:

- (12) a. You stopped beating your wife.
b. You didn't stop beating your wife.

Definite NPs are usually presupposed. In the examples below, the speaker presupposes the existence of the air raids even when the sentence is negative:

- (13) a. The air raids killed many people.
b. The air raids didn't kill many people.

We also discussed the following example:

- (14) a. Wait until you get July and August done.
b. Don't wait until you get July and August done.

We did not agree on whether *you get July and August done* is a presupposition. That is, we don't agree on whether the negative still presupposes that you will get July and August done.

4.2 Examples

Here are some examples:

- **Modal verbs of necessity**

- (15) I need/must/have to search you

In (15) above, the WS surely believes the *need, must, have to* propositions. As such, *need, must, have to* would all be labeled CB. However, the WS is not expressing a belief about the *search* proposition, so it would be labeled NA.

- **Epistemic certainty**

- (16) It's possible that the moon is made of cheese, but I don't know

In (16) above, the WS believes in the possibility, is not as certain about the *be made of* proposition (it is possible NOT certain), and believes in the *know* proposition. Based the definitions outlined above, the *be possible* proposition is labeled CB, the *be made of* proposition is labeled NCB, and the *know* proposition is labeled CB.

(17) I doubt that the moon is made of cheese

In (17) above, the WS believes in his or her doubt and is not as certain about the *be made of* proposition (it is doubtful NOT certain). Based the definitions outlined above, the *doubt* proposition is labeled CB, and the *be made of* proposition is labeled NCB.

- **The hypothetical & Conditional**

(18) If John dies, I will cry

In (18) above, the WS has no belief in the *die* or *cry* propositions because they are purely hypothetical. So, both *die* and *cry* would be labeled NA. However, there is a causal connection between *die* and *cry*, and this causal connection could be labeled CB.

(19) If John died, I would cry

In (19) above, the WS has no belief in the *die* or *cry* propositions because they are purely hypothetical. So, both *die* and *cry* would be labeled NA. However, there is a causal connection between *die* and *cry*, and this causal connection could be labeled CB.

(20) John was safe, but if he had died, I would have cried

This example is slightly different. The WS knows that John did not die and now has a certain belief in both the *die* and *cry* propositions - both of which would be labeled CB. The *die* and *cry* propositions are labeled CB due to negative polarity (the fact that John did not die). The causal connection still holds between *die* and *cry*, so this too would be labeled CB.

(21) I don't know if John died, but if he were to die, I would cry

In (21) above, the WS has no belief in the *die* or *cry* propositions because he or she does not know if John did die. So, both *die* and *cry* would be labeled NA. However, there is

a causal connection between *die* and *cry*, and this causal connection could be labeled CB.

(22) If John would have died, I would (have) crie(d)

Here, the *die* proposition would be labeled CB because the WS clearly has a strong belief about John dying (he did not die, so this is negative polarity). On the other hand, the WS has no belief in the *cry* proposition, so this proposition would be labeled NA. The causal connection still holds, and this connection would be labeled CB, as dying clearly entails crying.

(23) If I doubted the meeting were in Pittsburgh, I wouldn't have flown to Pittsburgh

In (23) above, the WS has strong belief in all three propositions. He or she does not doubt there was a meeting in Pittsburgh because he or she flew to Pittsburgh (a case of negative polarity). Thus, the *doubt* proposition would be labeled CB. The WS also clearly believes that the meeting is in Pittsburgh. Hence, the *be in Pittsburgh* proposition is also CB. Finally, the WS did in fact fly to Pittsburgh, so he or she must believe that there was an event taking place in Pittsburgh. As such, the *fly* proposition is also CB.

5 Annotation Tool

Annotation Tools

GATE (Sheffield) PropBank Emacs Teruko's tool ACE (University of Pennsylvania)