

Annotating Committed Belief

Interim Report: March 25, 2008

Mona Diab, Owen Rambow
Center for Computational Learning Systems, Columbia University
`mdiab@ccls.columbia.edu, rambow@ccls.columbia.edu`

March 26, 2008

Contents

1	Interim Report: March 25, 2008	2
1.1	Administrative Status	2
1.2	Status	2
1.3	Motivating Utility Example: Committed Belief	2
1.4	Issues and Difficult cases: Committed Belief	3
1.5	Future Thoughts	4
1.5.1	Can it be automated?	4

Chapter 1

Interim Report: March 25, 2008

1.1 Administrative Status

Columbia is still not under contract. The main issue is the fact that the subcontract proposed by Johns Hopkins included language related to classified research which Columbia cannot accept (it may not perform classified research). Columbia Research Administration responded to Johns Hopkins on January 24, 2008, and is waiting for a response.

We have been working on the project, but on an at-risk basis, with the Center for Computational Learning Systems at Columbia carrying the risk. Incurred expenses to date include annotator expenses.

Therefore, this report is not an official final report, but rather an interim report.

1.2 Status

Jointly with the CMU team (Lori Levin, Teruko Mitamura, Eric Davis) we have created an annotation manual and annotated the core corpus. We have not yet done inter-annotator agreement.

The annotation manual has been submitted by the CMU team.

1.3 Motivating Utility Example: Committed Belief

We will present two examples that motivated the utility of the committed belief annotation.

Example 1: Suppose an IR system is given a task to show documents that discuss incidents of peasants being robbed of their land. The first sentence below would be relevant to this query and the second would not. In our annotation scheme, *robbing* is tagged as committed belief in the first sentence, but is tagged as “not applicable” in the second.

- (1) The people robbing Iraqi peasants of their land should be punished.
- (2) Robbing Iraqi peasants of their land would be wrong.

Example 2: Suppose the following question was posed to a QA system: Did the humanitarian crisis end? The first sentence below answers the question. The second does not.

- (3) He arrived on Tuesday, bringing an end to the humanitarian crisis.
- (4) He arrived on Tuesday, calling for an end to the humanitarian crisis.

In general, we would use our committed belief annotation system as a filter on a traditional IR or QA system. After the traditional system returned results, the filter would apply in order to discard examples that do not contain committed beliefs.

1.4 Issues and Difficult cases: Committed Belief

A number of issues came up when deciding on an annotation scheme for committed belief. The first issue was what to annotate. Our group (Owen Rambow, Mona Diab) met with the group from CMU (Lori Levin, Teruko Mitamura, Eric Davis) and Jan Wiebe to discuss possibilities. We decided to annotate propositions identified by PropBank and nominals identified by NomBank.

The next issue was distinguishing between Committed Belief and Non-Committed Belief. The goal was to make a clear distinction to cut down on ambiguity and enable inter-coder agreement. The final meeting came from the above-mentioned meeting, and we determined that the write/speaker had to be nearly fully committed in his or her belief for a unit to be annotated Committed Belief. If a belief scale were created out of 10, Committed Belief would be 9 or higher.

Next, we had to consider multi-word expressions that come up in casual dialogs such as the switchboard document. The five that came up were: *Tell you what*, *Go ahead*, *How’s that*, *I say*, and *as you know*. We decided to label the former three as Not Applicable, since they are speech acts that do not necessarily address belief, whereas the latter 2 are Committed Belief, since depending on context, they can

have some bearing on the writer/speaker's belief. One final multi-word expression is something like *the more we know, the better...* We decided that the speaker truly believes that there is a correlation between *more* and *better* so both are labeled Committed Belief. In order to increase inter-coder agreement, we came up with 7 diagnostics. These are listed and explained in our manual about Annotating Belief in Communication.

The final issue was how to annotate presuppositions such as *beating your wife* in *Did you stop beating your wife?* We decided that such presuppositions should be labeled as Committed Belief, and our test for presuppositions is that they still are asserted even under negation. In addition, we agreed that definite NPs are usually presupposed. To further clarify our annotation scheme, we outlined 10 examples in our annotation manual. The examples cover such cases as modal verbs of necessity, epistemic certainty, and the hypothetical and conditional. Then final issue was whether we should use an annotation tool or not. Initially, as a proof of concept, we highlighted text spans in Microsoft Word. Then, we experimented with the ACE annotation tool from UPenn and GATE from Sheffield. We determined that GATE was easier to use and more intuitive, so we decided to use GATE as the annotation tool.

1.5 Future Thoughts

1.5.1 Can it be automated?

We believe that annotation of committed belief can be partially automated with rule based construction analyzers or automatic classifiers.

CMU has been experimenting with rule based construction analyzers on another project. The input to a construction analyzer is a dependency parse tree. The constructions themselves are form-meaning pairs. For example, the form *why not V-infinitive* as in *Why not go?* is associated with the meaning of making a suggestion. The form part of the form-meaning pair is represented as a template that matches a dependency tree. When the template matches, the tree is transformed into a semantic representation. For example, “why not” is removed and is replaced with “speech act: suggestion”.

Much of the committed belief annotation could be implemented with a set of constructions: main clauses with tensed verbs, clauses embedded under modal auxiliary verbs, clauses in certain sequence of tense relations with other clauses, and so on.

However, a small percentage of our decisions about committed belief may not be easily captured in rules. For example, in the sentence *Theft will become a bigger problem*, we decided that *theft* is an event that happened, based on the words

become bigger, which imply that theft is already big. This may not be a general or common rule, and may produce many false positives.

The syntactic patterns which map to certain values for committed belief can be learned from the annotated corpus. We can experiment with tree kernel-based learning methods (SVMs), or the use of supertags, which are tags which capture the syntactic behavior of a lexical head. Lexical and morphological features can also be added, thus perhaps capturing the complex cases just mentioned (*theft will become a bigger problem*). We expect to be able to achieve a fairly high but not perfect performance.