Defining Propositional Similarity

Systemizing identification of presuppositional binding

Jennifer Spenader

Computational Linguistics
Department of Linguistics
Stockholm University, 106 91 Sweden
jennifer@ling.su.se

Abstract

When can we say that two propositions in a spoken dialogue are similar enough that the one can function as an antecedent for the other? This study looks at the problem of determining meaning equivalence or meaning relatedness in spoken utterances taken from the London-Lund Corpus of Spoken English. Propositions judged to be equivalent or extremely similar in meaning were categorized according to the type of semantic relationship holding between them. The results here can be used to systemize the identification of presuppositional binding within the anaphoric theory of presupposition.

1. Introduction

Briefly, in the anaphoric theory (van der Sandt, 1992), presuppositions[1] are treated as anaphora. All presuppositional usage falls into one of two categories, binding or accommodation. Either a proposition which can function as an antecedent for the presupposition exists in the previous discourse, in which case binding occurs, or the presupposition must be accommodated into the discourse record. The accommodation process has been the subject of much discussion. The seemingly simpler problem of finding an antecedent for bound presuppositions is, however, deceptively difficult. With most other anaphora, a semantically poor anaphoric element, such as a pronoun, is bound to a semantically rich antecedent. Presuppositions are propositions, which means that determining an antecedent involves identifying a proposition earlier in the discourse that can be said to be sufficiently equivalent to the presupposed proposition to be able to bind with it. In many cases, the antecedent is actually a set of propositions, often related to the presupposition through linguistic relationships and based on world-knowledge, as well as inferential information implicit in the text.

This work attempts to give a more precise characterization of the relationships found between the propositional anaphor and propositional antecedents. It is hoped that a more explicit definition can then be used to automate the identification of binding when working with new examples.

2. Background

If presuppositions are semantically-loaded anaphoric expressions, then examining anaphoric binding in other types of anaphoric relationships should be relevant for identifying propositional binding.

First, what is the nature of the anaphor-antecedent relationship between nouns? Van Deemter (1992) developed a more general definition of anaphora by incorporating many non-traditional anaphoric NP relations that are central in establishing discourse coherence. He points out that referential identity has been the relationship traditionally identified between anaphors and their antecedents, in part because this is the *only* relationship possible for pronominal anaphora. Non-identity relationships are difficult to characterize. Possible relationships include subsumption, which he defines as "subset of a set, part of a quantity, substructure in a given structure" (p. 35) and those generated by relational nouns, (e.g. *the book-the author*, books always have an author). Many of these are sometimes termed bridging anaphora, though definitions differ widely. Van Deemter concentrated on full NP-anaphora but he also mentions anaphoric predicate relations between verbs. He does not, however, go in to how these could be handled.

Non-identity anaphoric relationships are problematic. Kramer & van Deemter (1998) develop a procedure for identifying non-identity relationships, which they term *partial matching*, in their study of definite noun phrases. They associate each discourse referent with a value set, which is the set of characteristics common to the common noun type to which the discourse referent belongs. If the disjunction of the value set of the two discourse referents is non-empty, then we have a partial match. How value sets are determined, and how we determine if the disjunction of two value sets is non-empty, is not elaborated on, which makes it difficult to apply this work to empirical data.

Bos et al. (1995) go further in developing a workable procedure for identifying partial matching by incorporating lexical information into the process. They use qualia structure information (Pustejovsky 1991), which contains information on both lexical relationships and world-knowledge, (an example they give is *bar-barkeeper*). By relating the anaphor to the relevant part of the qualia structure, bridging anaphora can be

[1] Following current usage, all propositions triggered by lexical or syntactic presuppositional triggers are called presuppositions.

resolved. Their work is also limited to definite noun phrases.

One draw-back of Bos et al.'s work is that it has difficulty modeling bridging instances where lexical relationships are not central. Asher & Lascarides (1998) attempt to remedy this by handling bridging by identifying the rhetorical relation that holds between the antecedent and the anaphoric expression. Identifying the rhetorical relationship allows them to compute other semantic information that in turn helps to make the binding relation more clear. They also briefly apply their method to an example of an it-cleft presupposition and an example with the temporal adverbial "again" as a presuppositional trigger. In both cases they discuss the presupposed proposition as an event, and identify a rhetorical relation that helps relate the event to the potential antecedent. Using their method requires that there is already a mechanism for unambiguously determining rhetorical relations.

In summary, work has focused on identifying what types of relationships can license anaphoric binding between full NPs: identity relationships, which are full matches, and several kinds of partial matches. How can we apply the methods found for NP-anaphora to propositions? There are identity propositions, that is when the two propositions are exactly the same. Additionally, we can probably safely allow the applications of some of the laws of logical equivalence found in statement and predicate logic, e.g. De Morgan's laws, laws of commutativity, etc., to also be considered propositional identify. Propositions that are expressed with clear synonyms should also be considered here.

But because even partial matching can license anaphoric binding, we also need to define degrees of similarity between propositions. Relationships such as hyponymy, subsumption, and relational noun inferences have been shown to license anaphoric binding between NPs. One way of applying this information to propositions is by allowing propositions to be considered similar enough for the one to function as an antecedent of the other when they differ in one argument. The relationship between the two arguments must then be one of the relationships defined as partial matching in full NP anaphora. For example, the propositions "The car is broken" should be considered similar enough to "The motor is broken" to allow the first to function as an antecedent for the latter, because of the relationship between *car* and *motor*. This type of relationship could be automated by having access to a lexicon, such as Pustejovsky's generative lexicon, which Bos et al. (1995) argued works with NP-anaphora.

If the relationships discussed here are found in the examples previously judged to be anaphoric binding, then we have the beginning of a method that could be automated to identify presuppositional binding.

3. Method

31 examples previously classified by the author as cases of presuppositional binding for another study were used as data. This earlier study looked at presuppositions with factive verbs, aspectual verbs and it-clefts, so examples are limited to these three trigger types. All examples were taken from the multi-speaker dialogues found in the London-Lund Corpus of Spoken English[2]. For each example, the entire discourse up until the presupposed proposition was examined for a possible antecedent. Examples were categorized according the relationship held between the presupposition and the propositions in the relevant parts of the previous discourse. When necessary, DRT-like representations of the propositions compared are given to make the relationships more clear.

4. Results

Examples of binding could be divided into four main categories that correlated positively with the authors intuitions of the ease of identification of the presupposition as bound. Each category is described below followed by examples. Examples are treated as if all other, non-propositional anaphora have been resolved.

The first two groups were by far the largest, and both deal with propositions where the relationship is one of identity.

1. Sense identity: a proposition/propositions earlier in the discourse have the same meaning as the presupposed proposition.

EXAMPLE 1 >> SHE IS UNUSUAL (1 3 1190)

Speaker A: But at the same time she seems unusual$_{antecedent}$, doesn't she.
Speaker B: Yes. And everybody notices that **she's unusual**.

This type of binding is most trivial and merely involves matching. There is no need to have access to additional linguistic information that that in the text.

EXAMPLE 2 >> "SOMEONE INVITED ME (= SPEAKER B)" (2 1@1086)

Speaker A:...was the invitation to York for which I did not apply. I was just invited$_{antecedent}$. (35 LINES LATER)
Speaker A: It was he who **invited me** .

A was invited differs from *someone invited A* only in that the former is passive and the later active. What the two sentences have in common is the fact that speaker A was invited, more explicitly, e.g. [x, A]:[invited(x,A)][3], where the value of x (the invitee) is uninstantiated in the antecedent, as well as in the presupposed proposition.

2. Sense identify by synonyms: Propositions differ only in lexical choice, where the words used were in a

[2] More information on the London-Lund Corpus of Spoken English can be obtained at http://www.hit.uib.no/icame/icame.html

[3] Examples given in DRT-format are simplified to only contain discourse referents and predicates relevant to the example.

well known lexical relationship such as synonymy or hyponymy.

EXAMPLE 3 >>"SOMEONE/SOME ARE ELITISTS" (2 9 630)

Speaker B: Oh no, it's very elitist$_{antecedent}$.
Speaker A: I thought it was the specialists who **are elitists**

This requires access to a lexicon that defines relationships between similar words of different parts-of-speech to identify the connection between "being elitist" and "being an elitist".

EXAMPLE 4 >> "SOMEONE LIFTED IT (=THE CONCEPT OF TRANSFORMATIONS) FROM THEM" (=MATHEMATICIANS) (2 5A 615+C)

Speaker A: I mean every transformation word that I've heard is in at the moment in [dhi] course for mathematics.
Speaker B: That's right. Well, that's where it all comes from$_{antecedent}$.
Speaker A: Yes. And it's {so} fascinating to see the analogy and it's much better in the mathematics than it is in grammar, I think.
Speaker B: But it's us that **lifted it from them**, not vice versa.

Here, B has already pointed out that the concept of transformations originated in mathematics, but seems to feel that A has not really understood that. He therefore reiterates with a more forceful it-cleft construction that linguists took the concept from mathematicians. Understanding that the presupposition is already present requires understanding the similarity between *"it" all comes from (mathematics) = someone lifted "it" from mathematics*, or more simply the synonymy of [X,Y,Z]:[[lifted_from(X,Y,Z)] = [originates(Y,Z)]] .

EXAMPLE 5 >> YOU ARE GOING TO KNOCK OUT AN EXPECTANT MOTHER (1 8 993)

Speaker A It was lethal to expectant mothers$_{antecedent}$ with small children.
(38 more lines)
Speaker A After all, I mean you can't go down and shop if you know that **you're going to knock out an expectant mother**-it was some <violent> streptococcus {that he'd got}

Finding the propositional antecedent in this example involves a kind of recursive process of identifying the relationship between the group of expectant mothers and a single instance of an expectant mother, followed by relating "it was lethal" to "knock out".

3. Non-identity/partial match of arguments:

1) The predicate of the proposition is the same, or a synonym, but one of the arguments to the predicate is different or

2) the predicate and arguments are the same, but the role of each argument differs.

EXAMPLE 6 >> "SOMEONE SAID I WANTED TO SELL OUT" (113 855)

Speaker A: James, it was no good. You didn't tell me to sell out$_{antecedent}$, it was I who **said I wanted to sell out.**

Here we must in some way characterize [B,A]:[tell_to(B,A,sellout)]=[B,A]:[want(say(A,A,sellout)]. The action of sell-out is common to both, but the difference lies in *tell* and *say* with a volitional operator *want*. Roughly, the expression want(say(X,Y,Z)) can only have the agent of "sell-out" as its subject, whereas the subject of tell can only be the agent of "sell-out" if a reflexive pronoun is used ("I told myself to sell out"). What the two propositions have in common is the relation of *A* as the agent to *sell-out*. Here, the change of polarity does not seem to be a hinder for an anaphoric interpretation.

4. Extended differences: Understanding the relationship requires tracking information over several propositions, and several speakers and then synthesizing this information as well as using extended lexical knowledge, knowledge about the context and world knowledge.

EXAMPLE 7 >> "HE IS ENTHUSIASTIC (HE = PROF. PITT)" (2 1 @ 120)

Speaker A: Also Pitt has talked about it a good deal. Professor Pitt here, and he has supported. (SEVERAL LINES) Yes, he has supported you is it with the Cambridge Press. He has supported you quite strongly and we had
(UNTIL LINE 1015)
Speaker A: You could indeed but I should continue also to give Professor Pitt's since I know that
Speaker B: **Know** that **he is enthusiastic**.
Speaker A: Yes, quite. He supported you very strongly.

Being supportive and *being enthusiastic* are not synonyms. However, we can understand them as synonyms in the context of this example because both are used to refer to Prof. Pitt's attitude towards Speaker B, and because the expression of *being supportive* is iterated several times, giving it a stronger import that makes it more possible to consider it synonymous with enthusiasm. Note the gap of almost 900 lines of dialogue from the time Prof. Pitt's support is mentioned until the statement that he is enthusiastic, illustrating the necessity of examining a very large context in order to correctly identify possible cases of binding.

EXAMPLE 8 >> PEOPLE GAVE YOU THINGS BEFORE (MARRIAGE) (210 927)

Speaker A doesn't seem much different except for trying to answer these awful conundrums about what shall we give you - and trying to fit it in (laughs)
Speaker B That's nice.
Speaker C That's the bit Debbie enjoyed enormously.
Speaker B Oh that's much the <cos> that won't, that won't happen again so I should enjoy it very much.
Speaker A Yes, that's true.
Speaker B Cos that stops very rapidly after you get married actually. **People** stop **giving you things**.

This example illustrates the necessity of tracking the individual parts of a possible propositional antecedent, and the need to use world knowledge. Speaker A's general statement about the difficulties involved in finding gifts is followed by an expression by Speaker C of how enjoyable receiving the gifts are. This "idea" is then referred to with different situational pronouns down until Speaker B's final statement, which according to my intuitions the giving of gifts before marriage was clearly part of the common ground, but pinpointing exactly where and how this is said is very difficult, and beyond the present study.

5. Discussion

Many of the examples studied could be accounted for by identifying simple lexical relationships between the anaphor and the antecedent. Some examples did require relating syntactically different sentences, for example passive sentences related to active counterparts, or ignoring certain types of extra information (such as the "seems" in example 1). Still, all these types of transformations are well understood and could probably be automated.

Typical bridging examples were not found. In the case of propositions this would be a bridging anaphoric relationship between one of the arguments in the antecedent with one of the arguments in the anaphoric proposition. There are three possible reasons for this result. The initial classification of examples as binding or accommodation may have been too strict, and examples with a possible bridging interpretation were placed into the accommodation category. Krahmer & van Deemter (1998) argued that partial match examples, including bridging, should be considered ambiguous between an interpretation as accommodation and an interpretation of binding, meaning that there is a greater chance that individual intuitions play a larger role in the interpretation of examples with bridging. A second possibility is that bridging between arguments in propositions may just be too difficult to take in, because the propositions compared often differ in other ways, e.g. there may be a temporal difference or a change of speaker in the dialogue. At the level of the individual word, the bridging relationship may be clear, but not as a smaller part of a proposition. A third possibility is that bridging is too infrequent to have been found in the limited number of examples studied here.

Not all cases involved a mere matching of synonymous arguments. In particular, group three illustrates some of the possible relationships that can arise when working with full propositions, rather than noun phrases. Here, actual switching of arguments had occurred between the two propositions judged to function in an anaphoric relation. In the data there were only two examples of this type of relationship, but immediately two questions arise. What other kinds of argument switching relationships license anaphoric binding, and where do we draw the line and acknowledge that we are dealing with two different propositions? The examples below give some ideas of what we may have to deal with.

(a) Mike gave Mary a package.
(b) No, Mary gave Mike a package.
(c) No, John gave Mary a package.
(d) No, John gave Mike a package.
(e) No, Mary took a package from Mike.

(ab) and (ac) are sequences similar to those found in the data. In (ab) the actors are the same, but the roles have been switched. In (ac) Mary's receiving a package is static, but the agent of the action changes. Both these examples could be argued to support an anaphoric relationship, and this is partly shown by our ability to use ellipse for c. (e.g. for c': "No, it was John".), so there is definitely an important coherency relation here. On the other hand, (ad) only keeps the action of giving static, but both actors have changed, making the jump too much to support a binding interpretation. (ae) keeps the actors and the initial and the resulting state static, but changes the import of the action. In terms of very basic concepts of transfer of ownership, (a) and (e) are very similar, but my intuitions on whether or not it would really license binding are not very clear here.

Another difficulty unique to propositional anaphora is that parts of the presupposed proposition may have been introduced into the discourse at different times, by different speakers, as illustrated by example 8. Tracking these, and then combing them to make the antecedent proposition requires a good understanding of the discourse.

6. Conclusions

Identifying propositional binding is easy for examples where lexical relationships known to license NP-anaphoric binding can be used. It is however difficult when arguments differ, or are switched, or when the antecedent proposition is a synthesis of several statements. Future work should look at more empirical data. Also, more work could be done working with idealized examples of the relationships found in group three and four.

References

Asher, N. & Lascarides, A. (1998). Bridging. Journal of Semantics,15 (1), 83--113.

Bos, J., Buitelaar, P. & Mineur, A. (1995). Bridging as Coercive Accommodation. In Proceedings of CLNLP'95, Workshop on Computational Logic for Natural Language Processing, Edinburgh.

van Deemter, K. (1992). Towards a Generalization of Anaphora. Journal of Semantics, 9(1), 27--51.

Krahmer, E. & van Deemter, K. (1998). On the Interpretation of Anaphoric Noun Phrases: Towards a Full Understanding of Partial Matches, *Journal of Semantics*, 15(4), 355--392.

Pustejovsky, J. (1991). The Generative Lexicon. Computational Linguistics, 17 (4), 409-441.

van der Sandt, R. A. (1992). Presupposition projection as Anaphora Resolution. Journal of Semantics, 9(4), 333--377.