Negotiation, compromise, and collaboration in interpersonal and human–computer conversations

Graeme Hirst
Department of Computer Science
University of Toronto
Toronto, Ontario
Canada M5S 3G4
gh@cs.toronto.edu

> The meaning of any message from Clyte had to be negotiated the way a company of soldiers negotiates a minefield.
> —Danni Hubson, *Post-modernism: A novel*

> Thus a 'hockey game' is nothing more than a discourse between two 'teams' in order to negotiate the designation of one team as 'winner' and one as 'loser' by means of the construction of a 'final score'.
> —Fraser Stegg, *Canada as a Social Construction*

1 The repair of failure to understand

People are very adept at recognizing when something they said has been *misunderstood* by a conversational partner and at recognizing when they themselves have misunderstood something that was said earlier in the conversation. In either case, they will usually say something to *repair* the situation and regain mutual understanding. The same is true of *non-understanding*. If computers are ever to converse with humans in natural language, they must be as adept as people are in their ability to detect and repair both their own occasional misunderstandings and also those of their conversational partner—perhaps even more so, as this skill will be needed to compensate for the likely deficiencies of computers in other aspects of understanding, which will lead to frequent misunderstandings and non-understandings on each side.

The processes through which conversational repairs take place include *negotiation*, *collaboration,* and *construction* of meaning. They can be seen in examples such as the following fragment from the London–Lund *Corpus of English Conversation* (Svartvik and Quirk, 1980, S.2.4a: 1–8); here, A is a visitor in B's house.

Example 1

1. A: What's that weird creature over there?
2. B: In the corner?
3. A: *[affirmative noise]*
4. B: It's just a fern plant.
5. A: No, the one to the left of it.
6. B: That's the television aerial. It pulls out.

The conversants here collaborate simultaneously on the phrases *weird creature* and *over there*, and in the course of doing so, must also work on the phrase *in the corner*. Observe that negotiations here relate not only to the meanings of referring expressions but also to the referring expressions of meanings. A's main goal in the conversation, after all, is to find words for a meaning; he sees an object, but cannot name it and wants to know the name. So the meaning of *weird creature* (or of *weird creature over there*) is not up for negotiation; it's fixed, and for A to be willing to compromise on it (as the term "negotiation" implies) would be absurd. He may abandon his goal, but he cannot negotiate it. The same is true of the denotation of *over there*—the weird creature is where it is; so A has the sub-goal of having B understand what he means by *over there*, which she initially doesn't. Thus in both these cases, meaning (denotation) is fixed but the referring expression for the meaning is negotiated. However, the converse happens for the referring expression *in the corner*, for which the conversants negotiate the denotation. For B, the meaning of *in the corner* seems to be a small area containing only the fern plant; for A, it is a somewhat broader area that includes the weird creature. (B could have held her ground by saying, in response to line 5, something like "Well, that's not really what I'd call 'in the corner' ".)

Thus we see in this example that negotiation of meaning can work in two directions: fitting the words to the meaning or fitting the meaning to the words. In either case, the conversants' actions are *collaborative* and *constructive*. They are collaborative in that both conversants work together to reach a conclusion acceptable to both; what differs is where the possibilities for compromise and change of position are. This is not to say that negotiation of meaning cannot be *adversarial*; on the contrary, we see adversarial negotiations of meaning in much legal argument and other kinds of argument. But even such cases are nonetheless collaborative in the sense that the conversants must engage one another, and may, however unwillingly, change or compromise their position as a result of something said by the other. In addition, the conversants' actions are constructive in that they jointly build a referring expression to fit the meaning or a meaning to fit the referring expression.

Negotiation may also occur in the repair of a complete *misunderstanding*, as in this example, a fragment of conversation between a mother and her child Russ about a parent–teacher meeting (Terasaki, 1976):

Example 2

1. MOTHER: Do you know who's going to that meeting?
2. RUSS: Who?
3. MOTHER: I don't know.
4. RUSS: Oh. Probably Mrs McOwen and some of the teachers.

Russ initially interprets line 1 as expressing Mother's desire to tell, that is, as a *pretelling* or *pre-announcement* as if Mother intends to surprise him (cf *Guess who's going to that meeting!*); but Russ finds this interpretation inconsistent with her next utterance. In line 3, instead of telling him who's going, as he would expect after pretelling, Mother claims that she does not know. Russ recovers by reinterpreting line 1 as an indirect request, which his line 4 then responds to. This example demonstrates people's preference to negotiate rather than repair the problems in the utterances of others (Schegloff *et al*, 1977); Mother could have avoided negotiation by simply making an explicit repair in line 3 ("No, *I'm* asking *you*").

In our earlier research, we developed computational models of negotiation in the collaborative, constructive repair of misunderstanding and non-understanding in conversations. To address non-understanding, we developed two plan-based models of collaboration in identifying the correct referent of a description: one covers situations, such as that of Example 1 above, where both conversants know of the referent, and the other covers situations, such as direction-giving, where the recipient does not. In these models, conversants use the mechanisms of refashioning, suggestion, and elaboration, to collaboratively construct and refine a referring expression until it is successful. To address misunderstanding, we developed a model that combines intentional and social accounts of discourse. Reflecting the inherent symmetry of the negotiation of meaning, all these models can act as both speaker and hearer, and can play both the role of the conversant who is not understood or misunderstood and the role of the conversant who fails to understand. A brief outline of the work is given below; details are given in the following papers: Hirst, McRoy, Heeman, Edmonds, and Horton 1994; Heeman and Hirst 1995; McRoy and Hirst 1995; Edmonds 1994.

2 Negotiating the construction of a referring expression

Heeman and Hirst presented a computational model of collaboration on the construction of a referring expression. This model uses the kinds of conversational moves that were observed by Clark and Wilkes-Gibbs (1986) in their laboratory studies of conversants speaking about hard-to-describe tangram figures. It covers the agent who makes the initial referring expression, that is, the *initiator*, and the agent who is to understand the expression, the *recipient*. In this model, the initiator has the goal of referring to something, and constructs a plan, in the form of a sequence of linguistic actions, to try to achieve it, given a set of beliefs about what the recipient believes. The recipient, seeing only the surface linguistic actions, tries to infer the plan in order to understand the reference. Thus, referring expressions are represented by plan derivations, and an unsuccessful referring expression is an invalid plan in whose repair the agents collaborate. This collaboration takes place through the use of plans that *judge* and *refashion* (or *reconstruct*) the expression.

In addition to the planning component, the change in state of the participant is modeled through belief- and goal-adoption rules. Thus a conversational agent is modeled throughout the collaborative activity, both in its role as a speaker and its role as a hearer. In fact, two copies of the model, with different beliefs and goals, can converse with one another. (The model is implemented in Prolog. Input and output are in the form of surface-level descriptions of speech acts.) Acting as a hearer, the system performs plan inference on each set of actions that it observes, and then updates the state of the collaboration. It then takes its turn as speaker. As the new speaker, the system looks for a goal that it can adopt, and then constructs a plan to achieve it. Next, presupposing the other participant's acceptance of the plan, it updates the state of the collaboration. It repeats this until there are no more goals to adopt. The surface actions of the constructed plans form the response of the system. The system then takes its turn as hearer and waits for a response from the other copy. The model is able to engage in a simplified version of Example 1, negotiating the construction of a referring expression for a denotation that the initiator wishes to convey, although it cannot handle the full complexity of negotiation within negotiation that is seen in the example.

A crucial assumption of this model is that the recipient already has some knowledge of the referent in question. In Example 1, it is an object in the room that both speakers can see. However, a speaker sometimes has to refer to an object that is not previously known to the hearer. For example, in giving driving directions, the speaker might say "Go straight ahead until you get to a funny-looking building". The recipient has to understand the reference well enough that when she later reaches the building, she will recognize it as the intended referent. That conversants can collaborate to achieve this kind of understanding can be seen in the following portion of a telephone conversation (Psathas, 1991. p. 196):

Example 3

1. A: You just stay on 2A, until you get to Lowell Street.
2. B: Is it marked?
3. A: Yeah, I think there's a street sign there, it's an intersection with lights.
4. B: Okay.

In this dialogue, speaker B has not understood the reference to the intersection at Lowell Street, and so suggests that the intersection might be marked. Speaker A replies with an elaboration of the initial expression. Edmonds extended the model of negotiation of a referring expression to account for this type of collaboration.

The basis of Edmonds's model is that the recipient can accept the referring expression of an identification plan if *confident* that the identification plan is *adequate*. Each agent associates a numeric *confidence value* with each of the attributes in the referring expression, and computes a level of confidence in the adequacy of the complete referring expression plan. If the overall confidence value exceeds the agent's *confidence threshold*, then the agent believes the plan to be adequate. That is, if the agent is the initiator, she believes that the other will be able to understand the reference; if the agent is the recipient, he believes that he has understood the reference. The confidence value of each attribute is equivalent to its *salience* within the context of the referring expression. Salience, for our purposes in direction-giving, is primarily visual prominence. Each agent has their own beliefs about salience, and it is the difference in their beliefs that leads to the necessity for collaboration on reference.

If the recipient is not confident in the adequacy of the plan, he informs the other. Now, although he cannot refashion the expression himself, he can help the initiator by *suggesting* a good way to expand it; *suggestion* is a collaborative move in which an agent posits a new attribute that would increase his confidence in the expression's adequacy if the expression were expanded to include the attribute. By using refashioning and suggestion moves, the two agents collaborate on refashioning the referring expression until the recipient of the directions is confident that it is adequate.

3 Repairing a misunderstanding

McRoy and Hirst developed a model of interaction in which the participants might differ about the speech act that is performed by some utterance, and, upon detecting this, try to identify the goal, expectation, or misunderstanding that might have led the other agent to produce it in order to negotiate an understanding. The model uses both intentional and social sources of knowledge. Intentional information is captured by two relations: one between utterances (input forms) and speech acts, and one between utterances and the attitudes that they express. These relations are the basis for deciding whether a set of utterances is consistent. To capture socially-derived expectations, the theory includes a relation on the speech acts—for each act, which acts are expected to follow. It also contains an axiomatization of speakers' knowledge for generating appropriate utterances and for detecting and repairing misunderstandings. The model demonstrates how these decisions depend on interactions among discourse participants' beliefs, intentions, previously expressed attitudes, and knowledge of social conventions.

In particular, the interpretation of utterances and the detection of misunderstandings are both characterized as abduction problems; speech act generation is characterized as default reasoning. As a result, all three processes can be specified within a single theory of communicative interaction. The model is implemented in Prolog with the Theorist framework for abduction with prioritized defaults (Poole *et al*, 1987); two copies of the model with different goals and beliefs can negotiate the resolution of a misunderstanding, as in Example 2.

4 Negotiating context

Sometimes, what's up for negotiation is not meaning per se but context, and when context is determined, meaning follows. But context may itself be a meaning—the meaning of an earlier utterance, for example.

In many languages and cultures, such as those of Ethiopia and Somalia, the negotiation of context in conversation is much more explicit and overt than it is in English and other European languages (Levine, 1985), but of course it occurs in those languages too. We see it every day in our ordinary conversations, in advertising, political discourse, poetry, humor, allusion, persuasion, and deception. In an earlier paper (Hirst, 2000), I have shown that what may included in negotiations of context is, in principle, unconstrained. Anything attended to may be used, and anything may be attended to.

Thus context and negotiations of context in natural language are in contrast with the formalizations of context for automated reasoning that have been proposed by McCarthy and Buvač (1997). McCarthy and Buvač make axioms context-dependent in order to be able to state them at the most convenient or useful level of generality. They give an example in which the price of airplane components is computed and negotiated by agents whose databases make different assumptions about exactly what objects and services the price covers—that is, they differ in what they take the price to denote because they have a slightly different ontology. Superficially, this appears to be similar to the collaboration on *in the corner* in Example 1. However, the examples are actually very different. McCarthy and Buvač's example tacitly assumes that the assumptions made by the databases are static, that all constants are rigid designators, and that the differences in the price denotations or in the ontology have already been explicated as axioms by some previous, unspecified process that McCarthy and Buvač never explicitly refer to but which would actually be the crucial, central process of the situation that they posit. It is tempting to call this unspecified process a negotiation of context, as it creates, in its axioms, a kind of mutual understanding of terms between the agents. It's not really a negotiation, however, because neither side actually concedes anything. Rather, it just determines how each side uses terms, and the new axioms provide a translation between them.

5 The semantic interoperability of humans and machines

The models described in sections 2 and 3 above are motivated by natural language and the long-term goal of human–computer interaction in natural language conversations. But because they are couched in terms of goals, plans, intentions, and expectations, they have an applicability beyond natural language to other kinds of communication, including non-linguistic communication between machines.

Of course, the models presented here deal with the collaborative resolution of only certain very specific kinds of communication problems, and do not mirror all the kinds of problems that arise in semantic interoperability. In particular, they do not cover negotiating the alignment of ontologies or the definition of terms, which is a prerequisite for McCarthy and Buvač's example and which may be a part of both prevention of and recovery from misunderstanding. However, they provide the basis for development of more-general models. Natural language is what humans use for semantic interoperability, and it is by modeling the mechanisms for collaboration and negotiation that natural language uses that we will be able to develop mechanisms for semantic interoperability in complex non-linguistic forms of communication.

Acknowledgements

I am grateful to Geoffrey Rockwell for discussion on some of the issues raised here. Preparation of this paper was supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

References

Clark, Herbert H. and Wilkes-Gibbs, Deanna (1986). "Referring as a collaborative process", *Cognition*, 22: 1–39.

Edmonds, Philip G. (1994). "Collaboration on reference to objects that are not mutually known", *Proceedings of the 15th International Conference on Computational Linguistics (COLING-94)*, Kyoto, 1118–1122.

Heeman, Peter A. and Hirst, Graeme (1995). "Collaborating on referring expressions", *Computational Linguistics*, 21(3): 351–382.

Hirst, Graeme (2000). "Context as a spurious concept." *Proceedings, Conference on Intelligent Text Processing and Computational Linguistics (CICLing-2000)*, Mexico City, February 2000, 273–287.

Hirst, Graeme; McRoy, Susan; Heeman, Peter; Edmonds, Philip; and Horton, Diane (1994). "Repairing conversational misunderstandings and non-understandings", *Speech Communication*, 15(3–4): 213–229.

Levine, Donald N. (1985). *The flight from ambiguity: Essays in social and cultural theory*. The University of Chicago Press.

McCarthy, John and Buvač, Saša (1997). "Formalizing context (expanded notes)." In: Aliseda, Atocha; van Glabbeek, Rob; and Westerståhl, Dag (editors), *Computing natural language*. Center for the Study of Language and Information, Stanford University, pp. 13–50.

McRoy, Susan W. and Hirst, Graeme (1995). "The repair of speech act misunderstandings by abductive inference", *Computational Linguistics*, 21(4): 435–478.

Poole, David; Goebel, Randy; and Aleliunas, Romas (1987). "Theorist: A logical reasoning system for defaults and diagnosis." In: Cercone, Nick and McCalla, Gordon (editors), *The knowledge frontier: Essays in the representation of knowledge*, Springer-Verlag, New York, pp. 331–352.

Psathas, George (1991). "The structure of direction-giving in interaction." In: Boden, Deirdre and Zimemrman, Don H. (editors), *Talk and social structure: Studies in ethnomethodology and conversation analysis*, Polity Press, Cambridge, England, pp. 195–216.

Schegloff, Emanuel A.; Jefferson, Gail; and Sacks, Harvey (1977). "The preference for self-correction in the organization of repair in conversation", *Language*, 53: 361–382.

Svartvik, Jan and Quirk, Randolph (1980). *A corpus of English conversation*, Lund Studies in English 56 (C.W.K. Gleerup, Lund).

Terasaki, A. (1976). "Pre-announcement sequences in conversation." Social Science Working Paper 99, School of Social Science, University of California, Irvine.