Misunderstanding and the negotiation of meaning

Susan W. McRoy
Department of Electrical Engineering and Computer Science
University of Wisconsin-Milwaukee
Milwaukee, WI 53201
mcroy@cs.uwm.edu

Graeme Hirst
Department of Computer Science
University of Toronto
Toronto, Canada M5S 1A4
gh@cs.toronto.edu

Introduction

Most computational accounts of dialogue have assumed that once listeners have interpreted an utterance, they never change this interpretation. However, human interactions routinely violate this assumption. This is because people are necessarily limited in how much information they can make explicit. As a result, *misunderstandings* might occur—discourse participants might differ in their beliefs about the meaning of what has been said or about its relevance to the discourse. To address this possibility, participants rely in part on their expectations to determine whether they have understood each other. If a speaker fails to notice anything unusual, she may assume that the conversation is proceeding smoothly. But if she hears something that seems inconsistent with her expectations, she may hypothesize that there has been a misunderstanding and attempt to reinterpret part of the discourse, initiating a *repair*.

In other words, speakers' inferences about discourse are nonmonotonic, because speakers may learn things that conflict with their earlier reasoning and cause them to re-evaluate what happened before. Because their utterances can only make a limited amount of information explicit, discourse participants' can only surmise—*abduce*—each other's intentions. They must reason from observed utterances to causes or goals that might account for them.

The importance of detecting misunderstandings

Failing to monitor for differences in understanding can lead to trouble. Listeners expect that a speaker's utterances will have some coherent relation to the discourse so far. So, when they consider the many possible interpretations of an utterance, they may rule out alternatives that are inconsistent with the discourse. Misunderstanding can lead to a breakdown in the conversation in which one participant cannot interpret some utterance. Misunderstanding can also cause participants to form different judgements about what they have accomplished; this problem is illustrated by Example 1 (Raudaskoski 1990).

The example involves an interaction between an experimental speech-recognition system and its user. The user is left with the mistaken impression that her message for Eve has been successfully recorded, although it has not; the system has failed to understand the user's utterances T2, T4, and T6 and is desperately trying to get the user to perform a repair, because it cannot. The user was misled because the system did not produce repairs. For example, the system's reply T7 is consistent with the user's interpretation of T5 as a request for the message and her reply, T6, as compliance with it. And, after the user began a new request in T8, the system failed to generate a repair in T9, even though it presumably has expectations from the previous request that are unresolved.

Example 1 An unrepaired misunderstanding

T1	**S:**	So you want to leave a message for Eve—is that right?
T2	**U:**	A message for Eve that's correct.
T3	**S:**	What is it that you want to do then?
T4	**U:**	Can you say that Judy is at home?
T5	**S:**	Do what?
T6	**U:**	Judy is at home.
T7	**S:**	You can leave messages or listen to messages—select either.
T8	**U:**	I want to leave a message for Fred.
T9	**S:**	So you want to leave a message for Fred—is that right?
T10	**U:**	Yes, for Fred that's correct.
T11	**S:**	Please give the message.

The problem underlying this example is that the system fails to use expectations. When people speak, they expect to receive a coherent response and, when they respond, they expect to have their action taken as evidence of understanding. Misunderstandings can occur when people form different expectations. To recognize

and repair them, discourse participants must be able to detect that their expectations have been violated.

The defeasibility of interpretation

Participants in a dialogue bring to it different beliefs and goals. These differences can lead them to make different assumptions about one another's actions, construct different interpretations of discourse objects, or produce utterances that are either too specific or too vague for others to interpret as intended. These problems may in turn lead to misunderstandings. Some of these can be circumvented, for example, by a hearer's identifying and correcting others' apparent misconceptions about objects or concepts mentioned in the discourse or by a speaker's adding disambiguating information to her own descriptions before uttering them (Goodman 1985; McCoy 1985; Calistri-Yeh 1991; Eller and Carberry 1992; Zukerman 1991). Not every misunderstanding can be avoided, however, because no speaker ever has enough information, by herself, to know how her utterance will be understood, even if perfectly cooperative and considerate of the context. Nor can she know whether she herself has understood. Maxims for cooperative behavior are insufficient, because in general neither participant can know ahead of time what will be informative or relevant for the other.

Discourse participants compensate for this limitation by using the evidence provided by their utterances to verify and revise their understanding of the conversation (Clark and Schaefer 1989; Brennan 1990). If either participant disagrees with the other's interpretation, they can challenge it. Alternatively, participants may accept an interpretation and respond with an utterance that shows their understanding and acceptance of it. In effect, speakers *negotiate* the meaning of utterances. This is illustrated by Example 2 (Gumperz 1982) where a repair is used to make the negotiation explicit. In this exchange, H intends T1 as a request for the location of the newspaper, whereas W takes it as a request to fetch the paper—even after being told that he just wants the information. Alternatively, H might have accepted her interpretation by simply saying "Thank you".

Example 2 A repair

T1	**H:**	Do you know where today's paper is?
T2	**W:**	I'll get it for you.
T3	**H:**	That's okay. Just tell me where it is. I'll get it.
T4	**W:**	No, *I'll* get it.

Misunderstanding and repair

It is useful to divide speech act misunderstanding into two types: misunderstandings that are made and detected by oneself, *self-misunderstanding*, and misunderstandings that are made by one participant, but detected by the other, *other-misunderstanding*. The first type arises when a hearer finds that he cannot incorporate an utterance into the discourse consistently, unless he interprets one of the speaker's utterances differently. The second type occurs when a hearer recognizes that if one of his own acts had been interpreted differently, the speaker's utterance would have been the expected response to it. The hearer might then attempt to change the speaker's interpretation, by performing a repair. For example, he might restate his intended goal or explicitly tell the speaker that she has misunderstood. Alternatively, the hearer might choose not to make the misunderstanding public, because certain forms of third-turn repairs can easily be mistaken for a challenge (Schegloff *et al.* 1977).

After a speaker detects a misunderstanding by either participant, she may produce a repair. Conversation analysts have identified three important types of discourse-level repair, distinguished by the number of turns between the misunderstood turn and the start of the repair (Schegloff 1992). The most common type is second-turn (or next-turn) repair. These repairs occur immediately after the problematic turn, before there has been any other reply to it, as in the following example:

Example 3 A second-turn repair

T1	**B:**	Do you know where Mr. Williams is?
T2	**A:**	What?

The next most common type of repair involves correcting another speaker's interpretation of the discourse. In the simplest case, a speaker makes an utterance displaying her misunderstanding in the turn immediately following the one she misunderstood. If the other speaker then recognizes the misunderstanding and initiates a sequence to resolve the misunderstanding, this is a *third-turn* (or *third-position*) repair, so called because the repair is initiated in the third turn of the top-level sequence, counting from the misunderstood utterance.[1] Consider Example 4 from Coulthard and Brazil (1984). In this example, B has responded to T1 with an *acknowledgement*, interpreting T1 as an *inform*.

Example 4 A third-turn repair

T1	**A:**	So the meeting's on Friday.
T2	**B:**	Thanks.
T3	**A:**	No, I'm *asking* you.

However, A intended T1 to be *yes-no* question (presumably with an *inform* as the expected reply). Recognizing B's misunderstanding, A produces a third-turn repair in T3, telling B what action A had intended in T1.

[1] Schegloff(1992) distinguishes nth-turn from nth-position repairs, where the former correspond to repairs that begin exactly *n-1* turns after the problematic utterance while the latter allow an arbitrary number of intervening pairs of turns. We shall use the terms interchangeably, allowing intervening exchanges.

A could have also told B the intended goal (*e.g.,* "No, I want you to tell me.")

The third type of repair involves producing a new reply to a turn that one has apparently misunderstood. Although there is a preference for repairing one's own misunderstandings (Schegloff *et al.* 1977), these repairs are deprecated because the number of potential targets for the repair increases with each intervening exchange, making locating the target increasingly difficult (Schegloff 1992). If a conversant hears an utterance that seems inconsistent with her expectations (perhaps because she has misunderstood some previous utterance) and the inconsistency leads her to reinterpret an earlier utterance and produce a new response to it, this is a *fourth-turn* (or *fourth-position*) repair (Schegloff 1987). Such repairs not only display the alternative interpretations, but also indicate some of the information that may underlie a participant's decision to favor one of them over another. Consider the fragment of conversation shown in Example 5 (Terasaki 1976).

Example 5 A fourth-turn repair

T1	**Mother:**	Do you know who's going to that meeting?
T2	**Russ:**	Who?
T3	**Mother:**	I don't know.
T4	**Russ:**	Oh. Probably Mrs. McOwen and probably Mrs. Cadry and some of the teachers.

In this dialogue, Russ initially interprets T1 as expressing Mother's desire to tell, that is, as a *pretelling* or *preannouncement*, but finds this interpretation inconsistent with her next utterance. In T3, instead of telling him who's going (as one would expect after pretelling), Mother claims that she does not know (and therefore could not tell). Russ recovers by reinterpreting T1 as an indirect request, which his T4 attempts to satisfy. This example also demonstrates agents' reluctance to repair the problems in the utterances of others (Schegloff *et al.* 1977); although Mother might have produced a third-turn repair at T3, the manifestation of a misunderstanding provided her with an expectable option that allowed her to avoid having to produce an explicit repair.

The need for both intentional and social information

Any dialogue system must account for the detection and repair of misunderstandings as well as the interpretation and production of utterances. To consider possible misunderstandings in addition to intended interpretations would explode the number of alternatives that an interpreter would need to consider, unless there were adequate constraints. However, predominant computational approaches to dialogue, which are based on intention, already have difficulty constraining the interpretation process. Proposed sociological accounts are more constrained, but none are computational. Some synthesis of intentional and social accounts of discourse is required.

In intentional accounts, speakers use their beliefs and goals to decide what to say; when hearers interpret an utterance, they try to identify goals that might account for it. This sort of reasoning is difficult to constrain because, although beliefs can narrow the search for an interpretation, there is no principled way of constraining the depth of the search. For each motivation that a hearer considers, he must also consider any higher-level motivations that it might support. To make such an approach workable, many simplifying assumptions have to be made, including the assumption that previous parts of the conversation have been understood correctly. However, there is another way to address misunderstanding that avoids this unconstrained inference of goals: use expectations deriving from social conventions (rather than intention) to guide interpretation.

In sociological accounts provided by Ethnomethodology, both coherent discourse interactions and repairs of misunderstandings are normal activities guided by social conventions (Garfinkel 1967; Schegloff 1992). There are conventions regarding the expected range of responses to every action, for example. People then can assume that others are behaving as expected, unless they have reason to believe otherwise. In this way, the conventions give speakers a guide to possible interpretations. Reasoning is also limited, because conventions do not depend on the psychological characteristics of particular participants. What these accounts lack that computational accounts provide is an explanation of how people can identify the convention that is relevant, especially when there is no pre-existing expectation.

A possible synthesis

In our work (McRoy 1993; McRoy and Hirst 1993) we have developed a model of communicative interaction that supports the negotiation of meaning discussed in Section 1. According to the model, speakers form expectations on the basis of what they hear, and thus monitor for differences in understanding. If necessary, they also reinterpret utterances in response to new information and generate repairs. Beliefs about the discourse context and conventions for interaction are used to select speech acts that are appropriate for accomplishing the speakers' goals. Interpretation and repair attempt to retrace this selection process abductively—when speakers attempt to interpret an observed utterance, they try to identify the goal, expectation, or misunderstanding that might have led the other agent to produce it.

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.

The key features of the model that distinguish it from previous ones are the following:

- An account of the detection and repair of speech act misunderstandings and its relation to generation and interpretation. Although there has been work on identifying *potential* sources of misunderstanding, none of it addresses the problem of identifying and repairing *actual* misunderstandings. Also, unifying these tasks requires that linguistic knowledge and processing knowledge be kept distinct, improving the clarity of the model and permitting general knowledge about language to be reused.
- An integration of the socially-determined, structural conventions that have been identified by Ethnomethodology with the use of belief and intention that has been popularized within Artificial Intelligence. As a result, the model does not do extended inference about goals when it is not necessary.
- An account of the nonmonotonicity of discourse reasoning. 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.
- A reification of expectation. According to the model, agents form expectations on the basis of social conventions. They filter these expectations by considering the consistency of the Gricean intentions that they have expressed. By contrast, previous models of discourse attempt to eliminate interpretations by using some (necessarily incomplete) set of felicity conditions.
- An axiomatization in Prioritized Theorist (Poole *et al.* 1987). Theorist is a declarative framework for default and abductive reasoning. Thus, linguistic knowledge and processing knowledge are kept distinct.

The architecture of our model

In the architecture that we have formulated, producing an utterance is a default, deductive process of choosing both a speech act that satisfies an agent's communicative and interactional goals and a utterance that will be interpretable as this act in the current context. Utterance interpretation is the complementary (abductive) process of attributing communicative and interactional goals to the speaker by attributing to him or her a discourse-level form that provides a reasonable explanation for an observed utterance in the current context. Expectations deriving from social norms delimit the range of responses that can occur without additional explanation. The attitudes that speakers express provide additional constraints, because speakers are expected not to contradict themselves. We therefore attribute to each agent:

- A set $\mathcal{B}$ of prior assumptions about the beliefs and goals expressed by the speakers (including assumptions about misunderstanding).
- A set $\mathcal{M}$ of potential assumptions about misunderstandings and meta-planning decisions that agents can make to select among coherent alternatives.
- A theory $\mathcal{T}$ describing his or her linguistic knowledge, including principles of interaction and facts relating linguistic acts.

DEFINITION 1: An *interpretation* of an utterance **u** to hearer **h** by speaker **s** in discourse context **ts** is a set M of instances of elements of $\mathcal{M}$, such that

1. $\mathcal{T} \cup \mathcal{B} \cup M$ is consistent
2. $\mathcal{T} \cup \mathcal{B} \cup M \models \mathit{utter}(\mathbf{s}, \mathbf{h}, \mathbf{u}, \mathbf{ts})$
3. $\mathcal{T} \cup \mathcal{B} \cup M$ is not in conflict with any stronger defaults that might apply.[2]

DEFINITION 2: It would be *coherent* for **s** to utter **u** in discourse context **ts** if **u** is a solution to the following default reasoning problem:
$\mathcal{T} \cup \mathcal{B} \cup M^{meta} \vdash (\exists u)\ \mathit{utter}(\mathbf{s}, \mathbf{h}, u, \mathbf{ts})$
where M^{meta} is a set of assumptions about meta-planning decisions in $\mathcal{M}$, such that

1. $\mathcal{T} \cup \mathcal{B} \cup M^{meta}$ is consistent
2. $\mathcal{T} \cup \mathcal{B} \cup M^{meta} \models \mathit{utter}(\mathbf{s}, \mathbf{h}, \mathbf{u}, \mathbf{ts})$
3. $\mathcal{T} \cup \mathcal{B} \cup M^{meta}$ is not in conflict with any stronger defaults that might apply.

DEFINITION 3: A speaker *S1* is *expected* to do action R in dialogue *TS* whenever there is an action A that is active in *TS* (because it was performed earlier), R as the normal expected reply to A, and the linguistic intentions of R are compatible with the active suppositions of *TS*.

In addition, acts of interpretation and generation update the set of beliefs and goals assumed to be expressed during the discourse. The current formalization focuses on the problems of identifying how an utterance relates to a context and whether it has been understood. The update of expressed beliefs is handled in the implementation, but outside the formal language.[3]

[2] More precisely, $\mathcal{T} \cup \mathcal{B} \cup M$ satisfies the priority constraints of Prioritized Theorist.

[3] A related concern is how an agent's beliefs might change after an utterance has been understood as an act of a particular type. Although we have nothing new to add here, Perrault (1990) shows how default logic might be used to address this problem.

The following formulates our characterization of misunderstanding and repair:

Other-misunderstanding
Speaker s_1 might be attempting action a_{new} in discourse ts if:

1. Earlier, speaker s_2 performed act $a_{intended}$;
2. Actions $a_{intended}$ and $a_{similar}$ can be performed using a similar surface form;
3. If s_2 had performed $a_{similar}$, then a_{new} would be expected;
4. s_1 may have mistaken $a_{intended}$ for $a_{similar}$.

Self-misunderstanding
Speaker s_1 might be attempting action a_{new} in discourse ts if:

1. s_1 has performed action $a_{observed}$;
2. But, the linguistic intentions of a_{new} are inconsistent with the linguistic intentions of $a_{observed}$;
3. $a_{observed}$ and action $a_{intended}$ can be performed using a similar surface-level speech act; and
4. s_2 may have mistaken $a_{intended}$ for $a_{observed}$.

Third-turn repair
Speaker s_1 should tell speaker s_2 that she intended to perform $a_{intended}$ in discourse ts if:

1. s_2 has apparently mistaken an instance of act $a_{intended}$ for act $a_{observed}$; and
2. s_1 may perform a third-turn repair (*i.e.,* the act that s_1 would expect to follow her intended action has not already been done and it would be consistent for s_1 to tell s_2 that she intended to perform $a_{intended}$).

Fourth-turn repair
Speaker s_1 should do action a_{reply} in discourse ts when:

1. s_1 has mistaken an instance of act $a_{intended}$ as an instance of act $a_{observed}$.
2. A reconstruction of the discourse is possible.
3. s_1 would expect to do a_{reply} in this reconstruction.
4. And, s may perform a fourth-turn repair.

Conclusion

Participants in a dialogue are necessarily limited in the amount of information that they can make explicit. Discourse participants compensate for this limitation by using the evidence provided by their utterances to verify each other's understanding of the conversation as it progresses. To show his understanding and acceptance of an utterance, a hearer may reply with an utterance that is consistent with the speaker's expectations. Alternatively, if he disagrees with the speaker's displayed interpretation, he can initiate a repair. In this way, participants *negotiate* the meaning of utterances. Moreover, at any given time, participants are only able to abduce each other's intentions from the utterances that they have observed.

The architecture that we have developed combines intentional and social accounts of discourse to support the negotiation of meaning.[4] The approach extends intentional accounts by using expectations deriving from social conventions in order to guide interpretation. As a result, it avoids the unconstrained inference of goals that has plagued many models of discourse.

Acknowledgements

This work was supported by the University of Toronto and the Natural Sciences and Engineering Research Council of Canada. We thank Ray Reiter for his suggestions regarding abduction; James Allen for his advice; Paul van Arragon and Randy Goebel for their help on using Theorist; Hector Levesque, Mike Gruninger, Sheila McIlraith, Javier Pinto, and Steven Shapiro for their comments on many of the formal aspects of this work; and Phil Edmonds, Stephen Green, Diane Horton, Linda Peto, Suzanne Stevenson and the other members of the natural language group for their comments.

References

Susan Elise Brennan. *Seeking and Providing Evidence for Mutual Understanding.* PhD thesis, Department of Psychology, Stanford University, Stanford, CA, 1990.

Randall J. Calistri-Yeh. Utilizing user models to handle ambiguity and misconceptions in robust plan recognition. *User Modelling and User Adapted Interaction*, 1(4):289–322, 1991.

Herbert H. Clark and Edward F. Schaefer. Contributing to discourse. *Cognitive Science*, 13(2):259–294, 1989.

Malcolm Coulthard and David Brazil. Exchange structure. In Malcolm Coulthard and Martin Montgomery, editors, *Studies in Discourse Analysis*, pages 82–106. Routledge & Kegam Paul, London, 1984.

Rhonda Eller and Sandra Carberry. A meta-rule approach to flexible plan recognition in dialogue. *User Modelling and User Adapted Interaction*, 2(1–2):27–53, 1992.

Harold Garfinkel. *Studies in Ethnomethodology.* Prentice Hall, Englewood Cliffs, NJ, 1967. (Reprinted: Cambridge, England: Polity Press, in association with Basil Blackwell, 1984.).

Bradley Goodman. Repairing reference identification failures by relaxation. In *The 23rd Annual Meeting of the Association for Computational Linguistics: Proceedings of the Conference*, pages 204–217, Chicago, 1985.

John J. Gumperz. *Discourse Strategies.* Cambridge University Press, New York, New York, 1982.

[4] We have implemented the model in Prolog and the Theorist (Poole *et al.* 1987; van Arragon 1990) framework for abduction with Prioritized defaults.

Kathleen F. McCoy. The role of perspective in responding to property misconceptions. In *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, volume 2, pages 791–793, 1985.

Susan W. McRoy. *Abductive Interpretation and Reinterpretation of Natural Language Utterances*. PhD thesis, Department of Computer Science, University of Toronto, Toronto, Canada, 1993. Published as CSRI Technical Report No. NNN, University of Toronto, Department of Computer Science.

Susan W. McRoy and Graeme Hirst. Abductive explanation of dialogue misunderstandings. In *6th Conference of the European Chapter of the Association for Computational Linguistics, Proceedings of the Conference*, pages 277–286, Utrecht, The Netherlands, 1993.

C. Raymond Perrault. An application of default logic to speech act theory. In Philip R. Cohen, Jerry Morgan, and Martha Pollack, editors, *Intentions in Communication*, pages 161–186. The MIT Press, 1990. An earlier version of this paper was published as Technical Report CSLI-87-90 by the Center for the Study of Language and Information.

David Poole, Randy Goebel, and Romas Aleliunas. Theorist: A logical reasoning system for defaults and diagnosis. In Nick Cercone and Gordon McCalla, editors, *The Knowledge Frontier: Essays in the Representation of Knowledge*, pages 331–352. Springer-Verlag, New York, 1987. Also published as Research Report CS-86-06, Faculty of Mathematics, University of Waterloo, February, 1986.

Pirrko Raudaskoski. Repair work in human-computer interaction. In Paul Luff, Nigel Gilbert, and David Frolich, editors, *Computers and Conversation*, pages 151–172. Academic Press, 1990.

Emanuel A. Schegloff. Some sources of misunderstanding in talk-in-interaction. *Linguistics*, 25:201–218, 1987.

Emanuel A. Schegloff. Repair after next turn: The last structurally provided defense of intersubjectivity in conversation. *American Journal of Sociology*, 97(5):1295–1345, 1992.

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

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

Paul van Arragon. *Nested Default Reasoning for User Modeling*. PhD thesis, Department of Computer Science, University of Waterloo, Waterloo, Ontario, 1990. Published by the department as Research Report CS-90-25.

Ingrid Zukerman. Avoiding mis-communications in concept explanations. In *Proceedings of the 13th Annual Conference of the Cognitive Science Society*, pages 406–411, Chicago, IL, 1991.