The repair of speech act misunderstandings by abductive inference

Susan W. McRoy*
University of Wisconsin-Milwaukee

Graeme Hirst†
University of Toronto

During a conversation, agents can easily come to have different beliefs about the meaning or discourse role of some utterance. Participants normally rely on their expectations to determine whether the conversation is proceeding smoothly: if nothing unusual is detected, then understanding is presumed to occur. Conversely, when an agent says something that is inconsistent with another's expectations, then the other agent may change her interpretation of an earlier turn and direct her response to the reinterpretation, accomplishing what is known as a fourth-turn repair.

Here we describe an abductive account of the interpretation of speech acts and the repair of speech act misunderstandings. Our discussion considers the kinds of information that participants use to interpret an utterance, even if it is inconsistent with their beliefs. It also considers the information used to design repairs. We describe a mapping between the utterance-level forms (semantics) and discourse-level acts (pragmatics), and a relation between the discourse acts and the beliefs and intentions that they express. We give a relation on the discourse acts—for each act which acts are expected to follow. We also describe our account of belief and intention, distinguishing the beliefs agents actually have from the ones they act as if they have when they perform a discourse act. To support repair, we model how misunderstandings can lead to unexpected actions and utterances and describe the processes of interpretation and repair. To illustrate the approach, we show how it accounts for an example repair.

1. Introduction

Speech act misunderstandings occur when two participants differ in their understanding of the discourse role of some utterance. For example, one speaker might take an utterance as an assertion while another understands it to be a request. Although many researchers have considered the problem of *avoiding* misunderstanding (*e.g.,* by correcting misconceptions), previously none had addressed the problem of identifying and repairing misunderstandings once they occurred. Here, we will consider a general model of dialogue that also accounts for the detection and repair of speech act misunderstandings.

1.1 The difference between misunderstanding and misconception

The notions of misunderstanding and misconception are easily confounded, so we shall begin by explicating the distinction. Misconceptions are errors in the prior knowledge of

* Department of Electrical Engineering and Computer Science, Milwaukee, WI 53201, *mcroy@cs.uwm.edu*
† Department of Computer Science, Toronto, Canada M5S 1A4, *gh@cs.toronto.edu*

a participant; for example, believing that Canada is one of the United States. McCoy (1989), Calistri-Yeh (1991), Pollack (1986b),Pollack (1990), and others have studied the problem of how one participant can determine the misconceptions of another during a conversation (see Section 5.3 below). Typically such errors can be recognized immediately when an expression is not interpretable with respect to the computer's (presumedly perfect!) knowledge of the world.

By contrast, a participant is not aware, at least initially, when *misunderstanding* has occurred. In misunderstanding, a participant obtains an interpretation that she believes is complete and correct, but which is, however, not the one that the other participant intended her to obtain. At the point of misunderstanding, the interpretations of the two participants begin to diverge. It is possible that a misunderstanding will remain unnoticed in a conversation and the participants continue to talk at cross-purposes. Alternatively, the conversation might break down, leading one participant or the other to decide that a misunderstanding has occurred and (possibly) attempt to resolve it.

1.2 The use of repair in the negotiation of meaning

Although they might not always recognize a misunderstanding when it occurs, discourse participants are aware that misunderstandings can occur. So, participants, rather than just passively hoping that they have understood and have been understood, actively listen for trouble and let each other know whether things seem okay. Each participant will use the subsequent discourse itself in order to judge whether previous discourse has been understood correctly. When one participant produces a response that is consistent and coherent with what the other has just said, then the other will take it as a *display* of understanding. Otherwise, it might be taken as evidence of misunderstanding. In either case, the response is used as an indication of how the second participant interpreted the first, as presumably his response must have some rational explanation; the indicated interpretation is called the *displayed interpretation*. When a participant notices a discrepancy between her own interpretation and the one displayed by the other participant, she can choose to initiate a repair or to let it pass. By their choice of repairing or accepting a displayed interpretation, speakers in effect *negotiate* the meaning of utterances.[1]

Repairs can take many forms, depending on how and when a misunderstanding becomes apparent. Conversation analysts classify repairs according to how soon after the problematic turn a participant initiates a repair (Schegloff, 1992). The most common type occurs within the turn itself or immediately after it, before the other participant has had a chance to reply. These are called first-turn repairs. The next most common type, second-turn repairs, occur as the reply to the problematic turn (*e.g.,* as a request for clarification). We will not consider these two types of repairs further, because they do not involve misunderstanding *per se*. Rather, they are used to correct misconceptions, misspeakings, nonhearings, etc.

Third-turn and fourth-turn[2] repairs address actual misunderstandings. If a display of misunderstanding occurs in the turn immediately following the one that was misunderstood, and the speaker notices the problem immediately and acts to resolve it, then we say that they have made a third-turn repair (see Example 1).
If a display of misunderstanding occurs during a subsequent turn by the same speaker

1 Note that this choice allows for a speaker feigning the occurrence of a misunderstanding in order to achieve some social goal.
2 Schegloff (1992) distinguishes nth-*turn* repair from nth-*position* repair. The former corresponds to repairs that begin exactly $n-1$ turns after the problematic utterance, while the latter allows an arbitrary number of intervening pairs of turns. We shall use "nth-turn" to refer to both types, allowing intervening exchanges.

Example 1

T1	**S:**	Where do you do this?
T2	**H:**	To make the crops grow.
T3	**S:**	I said where do you do it.
T4	**H:**	In a tin hut in Greeba.

who generated the misunderstood turn, and the hearer then reinterprets the earlier turn and produces a new response to it, then we say that they have made a fourth-turn repair. The fragment of conversation shown in Example 2 (Terasaki, 1976) includes a fourth-turn repair. Initially, Russ 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 a 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 Fox (1987). points out that such repairs involve, in effect, a *reconstruction* of the initial utterance. From an AI perspective, these reconstructions resemble the operation of a truth-maintenance system upon an abductive assumption that has proved to be incorrect.[3]

Example 2

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.

1.3 The need for both intentional and social information

The problem of interpreting an utterance involves deciding what actions the speaker is doing or trying to do. This process involves not only looking at the surface form of an utterance—for example, was it stated as a declarative?—but also at the context in which it was uttered. This context includes the tasks that the participants are involved in, the prior beliefs that they had, and the discourse itself. Context is important because it allows speakers to use the same set of words, for example "Do you know what time it is?", to request the time, to express a complaint, or to ask a yes–no question. Intentional information can rule out some of these readings; for example, a belief that the speaker already knows the time, might rule out the 'request' interpretation.

The difficulty in considering misunderstandings in addition to intended interpretations is that it greatly increases the number of alternatives that an interpreter needs to consider, because one cannot simply ignore the interpretations that seem inconsistent. However, predominant computational approaches to dialogue, which are based solely on inference of intention, already have difficulty constraining the interpretation process.

3 This is distinct from the kind of plan repair described by Spencer (1990), which he models using an assumption-based truth-maintenance system. In his work, "repair" addresses the problem of incompleteness in a taxonomy of plans, rather than errors in interpretation.

Sociological accounts suggest a more constrained approach to interpretation and the recognition of misunderstanding, but none are computational. Our model extends the intentional and social accounts of discourse, combining the strengths of both.

In the intentional accounts, speakers use their beliefs, goals, and expectations to decide what to say; when they interpret an utterance, they identify goals that might account for it. For example, a speaker who wants someone to know that she lacks a pencil might say "I don't have a pencil." A hearer might then interpret this utterance as an attempt to convey the information. However, for any goal that would explain an utterance, the reasons for having that goal would also be potential interpretations of the utterance. Thus, for the above utterance, intentional accounts might also consider interpretations corresponding to an attempt to express a need for a pencil, a request to be given the pencil, an incomplete attempt to fill out a questionnaire, and so on.[4] The inherent difficulty with this approach is thus knowing when to stop searching for potential meanings.

According to the ethnomethodological account of human communication known as Conversation Analysis (CA), agents design their behavior with the understanding that they will be held *accountable* for it. Agents know that their utterances will be taken to display their understanding of some (culturally determined) rules of conversation and the situation prior to the utterance. Agents, aware of some rule or norm that is relevant to their current situation, choose to follow (or not follow) the rule, depending on how they view the consequences of their choice. One important convention is the *adjacency pair*. Adjacency pairs are sequentially constrained pairs of utterances, (such as question–answer), in which an utterance of the first type creates an expectation for one of the second. A hearer is not bound to produce the expected reply, but if he does not, he must be ready to justify his action and to accept responsibility for any inferences that the speaker might make (Schegloff and Sacks 1973). Where the CA approach is weakest is in its explanation of how the recipient of an utterance is able to understand an utterance that is the first part of an adjacency pair. For this, an agent needs linguistic knowledge that links the features of an utterance to a range of speech acts that can form adjacency pairs. Agents also need to have some idea of the beliefs and intentions that particular actions express, so they can make judgments about their appropriateness in the context.

1.4 Overview

The aim of our research is to construct a model of communicative interaction that will be able to support the negotiation of meaning. In particular, we want to develop a general model of conversation that is flexible enough to handle misunderstandings. To support this degree of flexibility, the agents that we model form expectations on the basis of what they hear, monitor for differences in understanding, and, when appropriate, change their own interpretations in response to new information. The model specifies the relationship between this reasoning and discourse participants' beliefs, intentions, and previously expressed attitudes, as well as their knowledge of social conventions.

In the account, speakers select speech acts on the basis of both their goals and their knowledge of which speech acts are expected to follow upon a given speech act. They must select an utterance form that both parties would agree (in the current discourse context) could accomplish the desired goal. Interpretation and repair attempt to apply

4 The amount of reasoning is a function of the size of one's plan hierarchy. So, if it is believed that questionnaires are used to obtain a driver's license, which is needed to drive a car, which is needed to get to California, then this same utterance could even be interpreted as an incomplete attempt to get to California. Thus, the hearer must also assume that he and the speaker share the same plan hierarchy.

this process in reverse, working back from an observed utterance to the underlying goal. Such reasoning is clearly nonmonotonic; here we suggest that it can be characterized quite naturally as abduction. The model is expressed as a logical theory in the Prioritized Theorist framework (Poole, Goebel, and Aleliunas, 1987; van Arragon, 1990) .

2. The structured intentional approach

We now introduce a model of dialogue that extends both intentional and social accounts of discourse. The model unifies theories of speech act production, interpretation, and the repair of misunderstandings. This unification is achieved by treating production as default reasoning, while using abduction to model interpretation and repair. In addition, the model avoids open-ended inference about goals by using expectations derived from social norms to guide interpretation. As a result, the model provides a constrained, yet principled, account of interpretation; it also links social accounts of expectation with other mental states.

In this section, we will discuss how the model addresses the following concerns:

- The need to control the inference from observed actions to expected replies. Extended inference about goals is usually unnecessary and a waste of resources.
- The need to account for nonmonotonicity in both the interpretation and production of utterances. This nonmonotonicity takes two forms. First, utterances can make only a part of the speaker's goals explicit to the hearer, so hearers must reason abductively to account for them. Second, expectations are defeasible. At any given moment, speakers may differ in their beliefs about the dialogue and hence can only *assume* that they understand each other. Speakers manage the nonmonotonicity by *negotiating* with each other to achieve understanding.
- The need to detect and correct misunderstandings. Speakers rely on their expectations to decide whether they have understood each other. When hearers identify an apparent inconsistency, they can reinterpret an earlier utterance and respond to it anew. However, if they fail to identify a misunderstanding, the communication might mislead them into prematurely believing that their goals have been achieved.
- The need for an alternative to the notion of mutual belief. Typically, models rely on mutual beliefs without accounting for how speakers achieve them or for why speakers should believe that they have achieved them.

2.1 Using social conventions to guide interpretation and repair

Our account of interpretation avoids the extended inference required by plan-based models by reversing the standard dependency between an agent's expectations and task-related goals. Plan-based approaches (Allen and Perrault 1979; Litman 1986; Carberry 1990; Lambert and Carberry 1991) start by applying context-independent inference rules to identify the agent's task-related plan, possibly favoring alternatives that extend a previously recognized plan. By contrast, our approach begins with an expectation, using it to premise both the analysis of utterance meaning and any inference about an agent's goals. Moreover, our approach treats apparent conflicts with expectations as meaningful; for example, if an utterance is inconsistent with expectations, then the reasoner will try to explain the inconsistency.

The model focuses on two convention-based sources of expectation. The first is conventions about what attitudes (belief, desire, intention, etc) each speech act expresses;[5] we call these the *linguistic intentions* of the speech act. The second is conventions for each speech act about what act should follow; we call these *linguistic expectations*. Speakers will expect each other to display their understanding of these conventions and how they apply to their conversation. Thus, they can expect each other to be consistent in the attitudes that they express and to respond to each act with its conventional reply, unless they have (and can provide) a valid reason not to.

Linguistic intentions are based on Grice's (1957) notion of reflexive intention. For example, an *inform(S,H,P)* expresses the linguistic intentions whose content is P and *intend(S,know(H,P))* (*i.e.,* the speaker intends the hearer to believe (1) that P is true and (2) that the speaker intends that the hearer know P). Linguistic expectations capture the notion of adjacency pairs.[6]

In defining linguistic intentions, which are shown in Figure 1, we have followed existing speech act taxonomies, especially those given by Bach and Harnish (1979), Allen (1983), and Hinkelman (1990).[7] Thus, when a speaker produces an *askref* about P she expresses (and thereby intends the hearer to recognize that she expresses) that she does not know the referent of some description in P, intends to find out the referent of that description, and intends the hearer to tell her that referent. If the speaker is sincere, she actually believes the content of what she expresses; if the hearer is trusting, he might come to believe that she believes it.

Following Schegloff's (1988) analysis of Example 2, we provide a speech act definition for *pretell*.[8] In order to capture the linguistic intentions of *pretelling*, we also add a new attitude, *knowsBetterRef(S, H, P)* that is true if the knowledge of S is strictly better than the knowledge of P—for example because S is the expert or S has had more recent experience with P.

We allow that individuals might not all share the same taxonomy of speech acts and linguistic intentions and that certain social groups or activities might have their own specialized sets of linguistic expectations.[9] Our theory supports this flexibility by having each speaker evaluate the coherence of all utterances within her own view of the discourse. Thus, where we refer to the "displayed interpretation" of an utterance, we mean displayed given the perspective of a particular speaker.[10]

The figure shows a list of attitudes that each act expresses; the lists are assumed to be exhaustive with respect to the theory (but not to the various connotations that

---

5 We assume that these attitudes are a function of discourse or illocutionary level of speech acts, rather than the surface or locutionary level. This approach has worked well for us, but, as one reviewer remarked, it is an interesting issue as to whether they are also a function of the locutionary level.

6 Note that although linguistic intentions often express that an action is intended (*e.g.,* questions express an intention that the hearer answer), the two conventions are independent. For example, while an invitation to visit at 6pm might create an expectation that dinner will be served, it does not express an intention to serve it.

7 In the figure, we have used the symbol *intend* to name both the intention to achieve a situation in which a property holds and the intention to do action.

8 Schegloff actually argues *against* representing such sequences as speech acts; however, as in the computational work cited above, we have used notion of "discourse-level speech act" to represent the functional relationship between the surface form of an utterance, the context, and the attitudes expressed by the speaker.

9 Reithinger and Maier (1995) have used n-gram dialogue act probabilities to induce the adjacency pairs from a corpus of dialogues for appointment scheduling.

10 Communication can occur despite such differences because speakers with similar linguistic experiences presumably will develop similar expectations about how discourse works. Differences in expectations might very well be one thing that new acquaintances must resolve in order to avoid social conflict.

might be associated with each act). The set of acts itself is not necessarily exhaustive, but sufficient to handle the examples that we consider. While our taxonomy might seem small, most other acts appear to be specializations of those that we selected. Similarly, the model incorporates only a small number of linguistic expectations; these are shown in Figure 2).[11]

Act type	Speech act name	Linguistic intentions
informative	*assert(S, H, P)*	*know(S, P)*
	assertref(S, H, P)	*knowref(S, P)*
	assertif(S, H, P)	*knowif(S, P)*
	inform(S, H, P)	*know(P)* *intend(S, know(H, P))*
	informref(S, H, P)	*knowref(S, P)* *intend(S, knowref(H, P))*
	informif(S, H, P)	*knowif(S, P)* *intend(S, knowif(H, P))*
inquisitive	*askref(S, H, P)*	*not knowref(S, P)* *intend(S, knowref(S, P))* *intend(S, do(H, informref(H, S, P)))*
	askif(S, H, P)	*not knowif(S, P)* *intend(S, knowif(S, P))* *intend(S, do(H, informif(H, S, P)))*
requestive	*request(S, H, do(H, P))*	*intend(S, do(H, P))*
	pretell(S,H, P)	*knowref(S, P)* *knowsBetterRef(S, H, P)* *intend(S, do(S, informref(S, H, P)))* *intend(S, knowref(H, P))*
	testref(S, H, P)	*knowref(S, P)* *intend(S, do(H, assertref(H, S, P)))*
	testif(S, H, P)	*knowif(S, P)* *intend(S, do(H, assertif(H, S, P)))*

Figure 1
Linguistic intentions

First turn	Expected reply
askref	*informref*
askif	*informif*
request	*comply*
pretell	*askref*
testref	*assertref*
testif	*assertif*

Figure 2
Adjacency pairs (Linguistic expectations)

11 Quantitative results by Jose (1988) and Nagata and Morimoto (1993) provide evidence for these adjacency pairs. In addition, we have used pairs discovered by Conversation Analysis from real dialogues (Schegloff, 1988).

	Example	**Metaplan type**
1	A: Do you have a quarter?	Plan adoption
2	B: No.	Acceptance
3	B: I never lend money.	Challenge
4	A: No, I meant to offer you one.	Repair
5	B: Oh. Thanks.	Repair
6	A: Bye.	Closing

Figure 3
Examples of different types of coherence strategies

2.2 Characterizing interpretation, production, and repair

Our model unifies the fundamental tasks of interpreting, producing, and repairing speech act interpretations within a nonmonotonic framework. In particular, speakers' knowledge about language is represented as a set of default rules. The rules describe conventional strategies for producing coherent utterances, thereby displaying understanding, and strategies for identifying misunderstanding. As a result, speakers' decisions about what utterances they might coherently generate next correspond to default inference over this theory, while decisions about possible interpretations of utterances (including recognizing misunderstanding) correspond to abductive inference over the theory.

Definition 1: Given a theory $\mathcal{T}$ and a goal proposition $\mathcal{G}$, we say that one can *abduce* a set of assumptions Δ from $\mathcal{G}$ if $\mathcal{T} \cup \Delta \models \mathcal{G}$ and $\mathcal{T} \cup \Delta$ is consistent.

Abduction has been applied to the solution of local pragmatics problems (Hobbs et al. 1988; 1993) and to story understanding (Charniak and Goldman 1988).

The model incorporates five strategies, or *metaplans*, for generating coherent utterances: *plan adoption, acceptance, challenge, repair*, and *closing* (the model treats opening as a kind of plan adoption). Figure 3 contains a conversation that includes an example for each of the five types. In *plan adoption*, speakers simply choose an action that can be expected to achieve a desired illocutionary goal, given social norms and the discourse context. (The goal itself must originate within the speaker's non-linguistic planning mechanism.) The first utterance in the figure is a plan adoption. The second utterance in the figure, if it occurs immediately after an utterance such as the first one, would be an acceptance. With *acceptance* of an utterance, agents perform actions that have been elicited by a discourse partner. That is, the hearer displays his understanding and acceptance of the appropriateness of a speaker's utterance (independent of whether he actually agrees with it). *Challenges* display understanding of an utterance, while denying its appropriateness. For example, an agent might challenge the presuppositions of previous action. The third utterance, if it occurs immediately after an utterance such as the first one, would be a challenge. *Repairs* display non-acceptance of a previously displayed interpretation (see Section 1.2). The fourth utterance, occurring after an exchange such as $\langle 1, 3\rangle$, would be a third-turn repair by A; the fifth utterance, occurring after $\langle 1, 3, 4\rangle$, would be a fourth-turn repair by B.[12] *Closings* signal that the participants are ready to terminate the conversation (and that they accept the conversation as a whole). The last utterance in the figure is a closing.

Misunderstandings are classified according to which participant recognizes that the misunderstanding has occurred and who she thinks has misunderstood. *Self-misunderstandings* are those in which a hearer finds that a speaker's current utterance is inconsistent with something that that speaker said earlier and decides that his own interpretation of the earlier one must be incorrect. Conversely, *other-misunderstandings* are those in which the hearer attributes a misunderstanding to the speaker. Fourth-turn repairs may occur after a self-misunderstanding is recognized; third-turn repairs may occur after other-misunderstanding.

The model addresses both classes of misunderstanding (see Section 3.3.3), but is limited to misunderstandings that appear as misrecognized speech acts.[13] Such misun-

12 Non-understanding, which entails non-acceptance (or deferred acceptance), is signaled by second-turn repair. This type of repair will not be considered here.

13 Other misunderstandings are possible; for example there can be disagreement about what object a speaker is trying to identify with a referring expression (*cf.* Heeman and Hirst 1995; Hirst et al. 1994).

derstandings are especially important to detect, because the discourse role attributed to an utterance creates expectations that influence the interpretation of subsequent ones. These misunderstandings are also difficult to prevent, because they can result from many common sources, including intra-sentential ambiguity and mishearing.

2.3 Building a model discourse interpretations

For a hearer to interpret an utterance as a particular metaplan or as a manifestation of misunderstanding, he needs a model of his understanding of the prior discourse. The typical way to model interpretations has been to represent the discourse as a partially completed plan corresponding to the actual beliefs (perhaps even mutual beliefs) of the participants (*cf.* Carberry 1990). This representation incorporates two assumptions that must be relaxed in any model that accounts for the negotiation of meaning: first, that hearers are always credulous about what the speaker says, and second, that neither participant makes mistakes. To relax these assumptions, the hearer's model distinguishes the beliefs that speakers claim or act as if they have during the dialogue from those that the hearer actually believes they have.[14] The model also represents the alternative interpretations that the hearer has considered as a result of repair.[15] We will now consider an axiomatization of the model.

3. The architecture of the model

Our model characterizes a participant in a dialogue, alternately acting as speaker and hearer. In this section we will give both the knowledge structures that enable the participant's behavior and the reasoning algorithms that produce it. (Section 4 and Appendix A present machine-to-machine dialogues involving two instantiations of the implemented model.)

3.1 The reasoning framework: Prioritized Theorist

The model has been formulated using the Prioritized Theorist framework (Poole, Goebel, and Aleliunas, 1987; Brewka, 1989; van Arragon, 1990), because it supports both default and abductive reasoning. Theorist typifies what is known as a "proof-based approach" to abduction because it relies on a theorem prover to collect the assumptions that would be needed to prove a given set of observations and to verify their consistency. Our reasoning algorithm is based on Poole's implementation of Theorist, which we extended to incorporate preferences among defaults as suggested by van Arragon (1990).[16] A Prioritized Theorist reasoner can assume any default d that the programmer has designated as a potential hypothesis, unless it can prove $\neg d$ from some overriding fact or hypothesis. This makes the reasoning nonmonotonic, because the addition of a new fact or overriding default may make less preferable ones underivable.

The syntax of Theorist is an extension of the predicate calculus. It distinguishes two types of formulae, *facts* and *defaults*. In Poole's implementation, facts are given by "FACT w.", where w is a wff. A default can be given either by "DEFAULT (p, d)." or "DEFAULT $(p, d) : w$.", where p is a priority value, d is an atomic formula with only free variables as

14 This distinction is similar to the one made by Luperfoy (1992).
15 For present purposes, we also assume that the complete model is accessible to the hearer; one could better simulate the limitations of working memory by limiting access to only the most recent utterances.
16 Poole's Theorist implements a full first-order clausal theorem prover in Prolog. Like Prolog, it applies a resolution-based procedure, reducing goals to their subgoals using rules of the form $goal \leftarrow subgoal_1 \wedge \cdots \wedge subgoal_n$. However, unlike Prolog, it incorporates a model-elimination strategy (Loveland, 1978; Stickel, 1989; Umrigar and Pitchumani, 1985) to reason by cases.

arguments, and w is a wff. For example, we can express the default that birds normally fly, as:

DEFAULT $(2, birdsFly(b)) : bird(b) \supset fly(b)$.

If $\mathcal{F}$ is the set of facts and Δ^p is the set of defaults with priority p, then an expression DEFAULT$(p, d) : w$ asserts that $d \in \Delta^p$ and $(d \supset w) \in \mathcal{F}$. The language lacks explicit quantification; as in Prolog, variable names are understood to be universally quantified.

Facts are taken as true in the domain, whereas defaults correspond to the hypotheses of the domain (*i.e.,* formulae that can be assumed true when the facts alone are insufficient to explain some observation). A priority value is an integer associated with a given default (and all ground instances of it), where a default with priority i is stronger than one with priority j, if $i < j$. When two defaults conflict, the stronger one (*i.e.,* the one having the lower priority value) takes precedence. For sets of defaults Δ^i and Δ^j such that $i < j$, no $d \in \Delta^j$ can be used in an explanation if $\neg d \in \Delta^i$ and $\neg d$ is consistent with defaults usable from any Δ^h, $h < i$.

In the Theorist framework, *explanation* is a process akin to scientific theory formation—if a closed formula representing an observation is a logical consequence of the facts and a consistent set of default assumptions, then it can be explained:

DEFINITION 2: An *explanation* from the set of facts $\mathcal{F}$ and the sets of prioritized defaults $\Delta^1, \ldots, \Delta^n$ of a closed formula g is a set $\mathcal{F} \cup D^1 \cup \cdots \cup D^n$, where each D^i is a set of ground instances of elements of Δ^i, such that:

1. $\mathcal{F} \cup D^1 \cup \cdots \cup D^n$ is consistent
2. $\mathcal{F} \cup D^1 \cup \cdots \cup D^n \models g$
3. For all D^i such that $2 \leq i \leq n$, there is no $\mathcal{F} \cup D'^{\,1} \cup \cdots \cup D'^{\,i-1}$ that satisfies the priority constraints and is inconsistent with D^i.

Priority constraints require that no ground instance of $d \in \Delta^i$ can be in D^i if its negation is explainable with defaults usable from any Δ^j, $j < i$.

Priorities enable one to specify that one default is stronger than another, perhaps because it represents an exception. In our model, defaults will have one of three priority values: *strong*, *weak*, or *very weak*. The strongest value is reserved for attitudes about the prior context, whereas assumptions about expectations are given as weak defaults and assumptions about unexpected actions or interpretations are given as very weak defaults. This allows us to specify a preference for expected analyses when there is an ambiguity.

3.2 The language of our model

The model is based on a sorted first-order language, where every term is either an agent, a turn, a sequence of turns, an action, a description, or a supposition. The language includes an infinite number of variables and function symbols of every sort and arity. We also define several special ones to characterize suppositions, actions, and sequences of turns.

3.2.1 Suppositions. Suppositions are terms that name propositions that agents believe or express. Suppositions can be thought of as quoted propositions, but with a limited syntax and semantics. We define the following functional expressions:

- **do**(s, a) expresses that agent s has performed the action a;
- **mistake**(s, a_1, a_2) expresses that agent s has mistaken an act a_1 for act a_2;

- **and**(p_1, p_2) expresses the conjunction of suppositions p_1 and p_2, where p_1 must be simple (*i.e.,* not formed from others using the function symbol **and**);
- **not** p expresses the negation of a simple supposition p.[17]

We also define several suppositions for expressions of knowledge and intention.

Two suppositions are equivalent if and only if they are syntactically identical. To capture the notion that speakers are normally consistent in the suppositions that they choose to express, we need to know how different suppositions relate to each other. More to the point, we need to know when the expressing of two simple suppositions is or is not consistent. A complete account must take into account possible entailments among expressed propositions; however, no such account yet exists. As a placeholder for such a theory, there is a compatibility relation for expressed suppositions. Our approach is to make compatibility a default and define axioms to exclude clearly incompatible cases, such as these:

- The suppositions Q and **not** Q
- The supposition of an intention to make Q true when Q is already true in the agent's interpretation of the discourse.
- The supposition of the performance of some act that expresses, via a linguistic intention, any supposition that would be incompatible with (another supposition of) the agent's interpretation of the discourse.
- The supposition of an intention to perform some act expressing any supposition that is incompatible with the agent's interpretation of the discourse.
- The supposition of an intention to **knowif** Q if either Q or **not** Q is already true in the agent's interpretation of the discourse.

When suppositions are not simple, we check their compatibility by verifying that each of the conjuncts of each supposition is compatible. (In the system, this is implemented as a special predicate, *inconsistentLI*).

There is a danger in treating compatibility as a default in that one might miss some intuitively incompatible cases and hence some misunderstandings might not be detectable. An alternative would be to base compatibility on the notion of consistency in the underlying logic, if a complete logic has been defined.[18]

3.2.2 Speech acts. For simplicity, we represent utterances as surface-level speech acts in the manner first used by Perrault and Allen (1980).[19] Following Cohen and Levesque

---

17 The function **not** is distinct from boolean connective $\neg$. We use it to capture the supposition expressed by an agent who says something negative, *e.g.,* "I do not want to go," which might be represented as **inform(s, h, not WantToGo)**.

18 Note that human behavior lays somewhere in between these two extremes; in particular, people do not seem to express all the entailments of what they utter (Walker, 1991).

19 Other representation languages, such as one based on case semantics, would also be compatible with the approach and would permit greater flexibility. The cost of the increased flexibility would be increased difficulty in mapping surface descriptions onto speech acts; however, because less effort would be required in sentence processing, the total complexity of the problem need not increase. Using a more finely grained representation, one could reason about sentence type, particles, and prosody explicitly, instead of requiring the sentence processor to interpret this information (*cf.* Hinkelman 1990; Beun 1990).

(1985), we limit the surface language to the acts **surface-request**, **surface-inform**, **surface-informref**, and **surface-informif**. Example 3 shows the representation of the literal form of Example 2, the fourth-turn repair example. (We abbreviate "**m**" for "Mother", "**r**" for "Russ", and "**whoIsGoing**" for "who's going".)

Example 3

T1 m: **surface-request(m, r, informif(r, m, knowref(r, whoIsGoing)))**

T2 r: **surface-request(r, m, informref(m, r, whoIsGoing))**

T3 m: **surface-inform(m, r, not knowref(m, whoIsGoing))**

T4 r: **surface-informref(r, m, whoIsGoing)**

We assume that such forms can be identified by the parser, for example treating all declarative sentences as **surface-informs**.[20]

The theory includes the discourse-level acts **inform, informif, informref, assert, assertif, assertref, askref, askif, request, pretell, testref,** and **warn**, which we represent using a similar notation.[21] [22]

3.2.3 Turn sequences. A turn sequence represents the interpretations of the discourse that a participant has considered up to a particular time. It is structured as a tree, where each level below the root corresponds to a single turn in the sequence, ordered as they occurred in time. Each path from the root to a leaf represents a single interpretation of the dialogue. Nodes that are siblings (*i.e.,* that have the same parent) correspond to different interpretations of the same turn. Nodes at the same level, but having different parents, represent repairs. The currently active interpretation is defined by its most recent turn, which we shall call the *focus* of the sequence.

The purpose of this tree structure is to capture the sequential structure of the dialogue and, for each state of the dialogue, what attitudes the participants are accountable for having expressed.[23] Branches in the sequential structure enable the participants to retract attitudes via repair and to reason about the alternatives that they have achieved.

We will call the turn sequence whose focus is the current turn the "discourse context". In order to consider previous states of the context, such as before a possible misunderstanding occurred, we define a successor relation on turn sequences:

DEFINITION 3: A turn sequence *TS2* is a *successor* to turn sequence *TS1* if *TS2* is identical to *TS1* except that *TS2* has an additional turn t that is not a turn of *TS1* and t is the successor to the focused turn of *TS1.*

---

20 We also presume that a parser can recognize **surface-informref** and **surface-informif** syntactically when the input is a sentence fragment, but it would not hurt our analysis to input them all as **surface-inform**.

21 In the utterance language, a yes–no question is taken to be a **surface-request** to **informif** and a wh-question is taken to be a **surface-request** to **informref**. We then translate these request forms into the discourse-level actions **askif** and **askref**. An alternative would be to identify them as **surface-askif** or **surface-askref** during sentence processing, as Hinkelman (1990) does.

22 Speech act names that end with the suffix **-ref** take a description as an argument; speech act names that end with **-if** take a supposition. The act **inform**(s, p) asserts that the proposition is true. The act **informif**(s, p) asserts the truth value of the proposition named by p (*i.e.,* **informif** is equivalent to "**inform** $\vee$ **inform-not**").

23 Tree structures are often used to represent discourse, but usually the hierarchical structure of the discourse, rather than its temporal structure (see Lambert and Carberry 1991; 1992).

3.3 The characterization of a discourse participant

We will now consider the knowledge structures that enable a participant's behavior and the reasoning algorithms that produce it. We divide our specification of a participant into three subtheories:

- 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 metaplanning decisions.
- A theory $\mathcal{T}$ describing his or her linguistic knowledge, including principles of interaction and facts relating linguistic acts.

Given these three subtheories, an interpretation of an utterance is a set of ground instances of assumptions that explain the utterance. An utterance would be a coherent reply to the first utterance if it would logically follow, given the selection of some metaplan:

DEFINITION 4: An *interpretation* of an utterance $\mathbf{u}$ to hearer $\mathbf{h}$ by speaker $\mathbf{s}$ in discourse context $\mathbf{ts}$ 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. $\mathcal{T} \cup \mathcal{B} \cup M$ satisfies the priority constraints; that is, $\mathcal{T} \cup \mathcal{B} \cup M$ is not in conflict with any stronger defaults that might apply.

DEFINITION 5: It would be *coherent* for $\mathbf{s}$ to utter $\mathbf{u}$ in discourse context $\mathbf{ts}$ if the utterance can be derived from an agent's linguistic knowledge, assuming some set M^{meta} of metaplanning decisions, 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. $\mathcal{T} \cup \mathcal{B} \cup M^{meta}$ satisfies the priority constraints.

That is, $\mathbf{u}$ is a solution to the following default reasoning problem:

$$\mathcal{T} \cup \mathcal{B} \cup M^{meta} \vdash (\exists u)\ \mathit{utter}(\mathbf{s},\ \mathbf{h},\ u,\ \mathbf{ts})$$

In the language of the model, the predicate *shouldTry* is used for discourse actions that are coherent (M^{meta}) and the predicate *try* is for actions that are explainable (M). If *shouldTry(S1,S2,A,TS)* is true, it means that, given discourse context *TS* (which corresponds to a particular agent's perspective), it would be appropriate for speaker *S1* to address speaker *S2* with discourse-level speech act *A* (*i.e.,* according to social conventions, here represented by the linguistic expectations and the meta-plans, *S1* should do *A* next).

By contrast, *try(S1,S2,A,T2)* would mean that, given a discourse context *TS*, *S1* has performed the discourse-level act *A*. Discourse-level acts are related to surface-level acts by the following default:

DEFAULT $(3, pickForm(s_1, s_2, a_{surfaceForm}, a, ts))$:[24]
$decomp(a_{surfaceForm}, a)$
$\wedge\ try(s_1, s_2, a, ts)$
$\supset\ utter(s_1, s_2, a_{surfaceForm}, ts)$.

This says that the fact that the surface form $a_{surfaceForm}$ can be used to perform discourse act a in some context and the apparent occurrence of a would be a reason for agent s_1 to utter $a_{surfaceForm}$.

The predicates *shouldTry* and *try* are related because the appropriateness of a potential interpretation is taken as (default) evidence that it is, in fact, the correct interpretation:

DEFAULT $(1, intentionalAct(s_1, s_2, a, ts))$:
$shouldTry(s_1, s_2, a, ts)$
$\supset try(s_1, s_2, a, ts)$.

The key difference is that *try* allows that the best interpretation might be contextually inappropriate. See Figure 4.

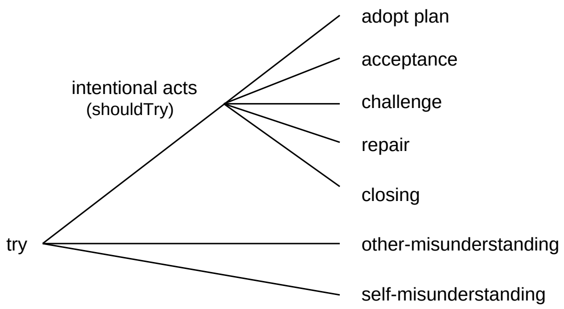

Figure 4
The relationship between *try* and *shouldTry* and their possible explanations

Interpretation corresponds to the following problem in Theorist:

EXPLAIN $utter(\mathbf{s1}, \mathbf{s2}, \mathbf{u}, \mathbf{ts})$.

Generation corresponds to the following problem in Theorist:

EXPLAIN $shouldTry(\mathbf{s1}, \mathbf{s2}, a_d, \mathbf{ts}) \wedge \ decomp(a_s, a_d)$.

In addition, acts of interpretation and generation update the set of beliefs and goals assumed to be expressed during the discourse.[25]

3.3.1 The discourse context. The first component of the model, $\mathcal{B}$, represents the beliefs and goals that the participants have expressed during their conversation. We assume that an agent will maintain a record of these expressed attitudes, represented as a turn sequence. To keep track of the current interpretation of the dialogue we introduce the notion of *activation* of a supposition with respect to a turn sequence. If during a turn T, a supposition is expressed by an agent through the utterance of some speech act or

24 The model does not discriminate between equally acceptable alternatives. The default *pickForm* allows us to account for the fact that the same surface form can perform several discourse acts and the same discourse act might be accomplished by one of several different surface forms. In our system, this default is also used as an oracle, allowing us to see how different interpretations affect the participants' understanding of subsequent turns. Because the default has very weak priority, it can be overridden by user input, without influencing other defaults.

25 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 display of misunderstanding, then we say it becomes *active* in the turn sequence that has T as its focus (see Section 3.2.3). Moreover, once active, a supposition will remain active in all succeeding turn sequences, unless it is explicitly refuted.

Individual turns are represented by a set of facts of the form *expressed(P, T)* and *expressedNot(P, T)*, where P is an unnegated supposition that has not been formed from any simpler suppositions using the function **and**.[26]

3.3.2 Possible hypotheses. The second component of the model is $\mathcal{M}$, the set of potential assumptions about misunderstandings and metaplanning decisions. This is given by a set of Theorist defaults:[27]

intentionalAct, expectedReply, acceptance, adoptPlan, challenge, makeFourthTurnRepair, makeThirdTurnRepair, reconstruction, otherMisunderstanding, selfMisunderstanding, and *done*.

The theorem prover may assume ground instances of any of these predicates if they are consistent with all facts and any defaults with higher priority. As mentioned in Section 3.1, each of these defaults will have one of three priority values, *strong*, *weak*, or *very weak*. The strongest level is reserved for attitudes about beliefs and suppositions. Assumptions about expectations (*i.e., expectedReply, acceptance, makeThirdTurnRepair*, and *makeFourthTurnRepair*) are given as weak defaults. Assumptions about unexpected actions or interpretations (*i.e., adoptPlan, challenge, done, selfMisunderstanding*, and *otherMisunderstanding*) are given as very weak defaults, so that axioms can be written to express a preference for expected analyses when there is an ambiguity. We will consider each of these predicates in greater detail when we discuss the third component of the model.

3.3.3 A speaker's theory of language. The third component of the model is $\mathcal{T}$, a speaker's theory of communicative interaction. This theory includes strategies for expressing beliefs and intentions, for displaying understanding, and for identifying when understanding has broken down. The strategies for displaying understanding suggest performing speech acts that have an identifiable, but defeasible, relationship to other speech acts in the discourse (or to the situation). Misunderstandings are recognized when an utterance is inconsistent or incoherent; strategies for repair suggest reanalyzing previous utterances or making the problem itself public.

Relations on linguistic knowledge. There are three important linguistic knowledge relations: *decomp*, *lintention*, and *lexpectation*. They are shown as circles in Figure 5; the boxes in the figure are the objects that they relate.

The *decomp* relation links surface-level forms to the discourse-level forms that they might accomplish in different contexts. It corresponds to the *body* relation in STRIPS-based approaches.[28] Two speech acts are *ambiguous* whenever they can be performed with the same surface level form. *Lintentions* relate discourse acts to the linguistic intentions that they conventionally express (see Section 2.1). The *lexpectation* relation

26 The intended meaning of $expressedNot(P, T)$ is that during turn T speakers have acted as if the supposition P were false. Although $expressed(\mathbf{not}(P), T)$ and $expressedNot(P, T)$ represent the same state of affairs, the latter expression avoids infinite recursion by Theorist.

27 The theory also contains defaults to capture the persistence of activation (*persists*), and the willingness of participants to assume that others have a particular belief or goal (*credulousB* and *credulousI*, respectively).

28 Pollack (1986a) calls this the "is-a-way-to" relation.

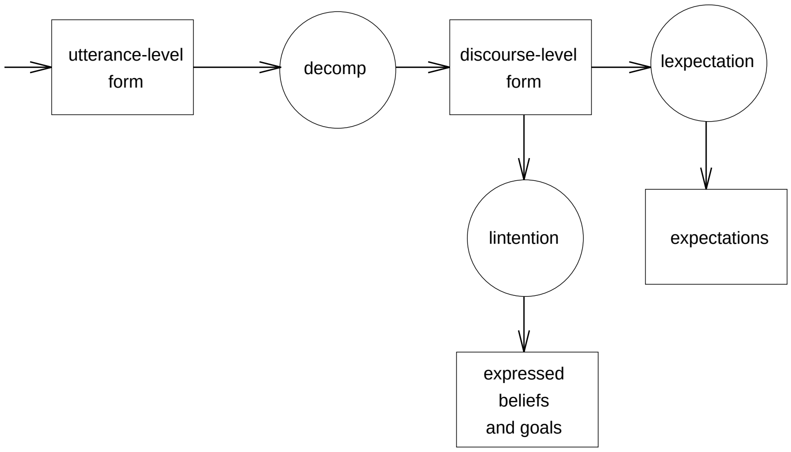

Figure 5
How the knowledge relations fit together

captures the notion of linguistic expectation discussed in Section 2.1, relating each act to the acts that might be expected to follow. Where there is more than one expected act, a condition is used to distinguish them. For example, the axioms representing the linguistic expectations of **askref** are shown below.[29]

FACT $lexpectation(\mathbf{do}(s_1, \mathbf{askref}(s_1, s_2, d)),$
$\mathbf{knowref}(s_2, d),$
$\mathbf{do}(s_2, \mathbf{informref}(s_2, s_1, d)))$.
"A speaker s_1 can expect that making an **askref** of d to s_2 will result in s_2 telling s_1 the referent of d, if s_2 knows it.

FACT $lexpectation(\mathbf{do}(s_1, \mathbf{askref}(s_1, s_2, d)),$
$\mathbf{not\ knowref}(s_2, d),$
$\mathbf{do}(s_2, \mathbf{inform}(s_2, s_1, \mathbf{not\ knowref}(s_2, d))))$.
"A speaker s_1 can expect that making an **askref** of d to s_2 will result in s_2 telling s_1 that s_2 does not know the referent of d, if s_2 does not know it."

Beliefs and goals. In the model, participants' actual beliefs and goals are distinguished from those that they express through their utterances. For the examples considered here, any model of belief would suffice; for simplicity we chose to include beliefs and goals explicitly in the initial background theory and allow agents to make assumptions about each other's beliefs and goals by default.[30]

Expectation. In addition to the notion of linguistic expectations, which exist in any situation, the model incorporates a cognitive, "belief-about-the-future" notion of expectation. These expectations depend on a speaker's knowledge of social norms, her understanding

29 It is actually controversial whether an **askref** followed by an **inform-not-knowref** is a valid adjacency pair. If such questions are taken to presuppose that the hearer knows the answer, a response to the contrary could also be considered a *challenge* of this presupposition (Tsui, 1991).
30 It would have been possible to characterize actual belief using an appropriate set of axioms, such as those defining a weak S4 modal logic. However, current formalizations do not seem to account for the context-sensitivity of speakers' beliefs. See McRoy (1993b) for a discussion.

of the discourse so far, and her beliefs about the world at a particular time. They are captured by the following Theorist rules:

DEFAULT $(2, expectedReply(p_{do}, p_{condition}, \mathbf{do}(s_1, a_{reply}), ts))$:
$active(p_{do}, ts)$
$\wedge\ lexpectation(p_{do}, p_{condition}, \mathbf{do}(s_1, a_{reply}))$
$\wedge\ believe(s_1, p_{condition})$
$\supset expected(s_1, a_{reply}, ts)$.

FACT $\neg lintentionsOk(a, ts)$
$\supset \neg expectedReply(p_{do}, p_{condition}, \mathbf{do}(s, a), ts)$.

The second rule says that one would not expect the action a_{reply} if the linguistic intentions associated with it are incompatible with the context ts.[31] Normally, as the discourse progresses, expectations for action that held in previous states of the context eventually cease to hold in the current context, because after the action occurs, it would be incompatible for an agent to say that he intends to achieve something that is already true. The compatibility between each of the linguistic intentions of a proposed action and each of the active suppositions in a context is captured by the predicate *lintentionsOk*, which is true if and only if none of the incompatibilities described in Section 3.2.1 hold.

For convenience, we also define a subjunctive form of expectation to reason about expectations that would arise as a result of future actions (*e.g.,* plan adoption) or that must be considered when evaluating a potential repair. This type of expectation differs from the type defined above in that it depends on the real beliefs of the agent performing the first (rather than the second) part of an adjacency pair and it does *not* depend on the activity of any suppositions or actions.

FACT $lexpectation(\mathbf{do}(s_1, a_1), p, \mathbf{do}(s_2, a_2))$
$\wedge\ believe(s_1, p)$
$\equiv wouldExpect(s_1, a_1, a_2)$.

Metaplans and misunderstandings. The metaplans encode strategies for selecting an appropriate act. The antecedents of these axioms refer to expectations. In addition, in order to preserve discourse coherence, they require that either the linguistic intentions of suggested actions be compatible with the context or that there be some overt acknowledgement of the discrepancy. (The theory presented here addresses only the former case; the latter one might be handled by adding an extra default with a stronger priority level.) Tables 1–6 give each of these axioms in detail.

Along with these metaplans, a speaker's linguistic theory includes two diagnostic axioms that characterize the speech act misunderstandings: self-misunderstanding and other-misunderstanding. The antecedents of these axioms refer to ambiguities and inconsistencies with expressed linguistic intentions, as well as expectations. For example, Table 5 describes how an observed inconsistency of s_1 performing a_{new} might be a symptom of s_2's misinterpretation of an earlier act by s_1. Such mistakes are possible when the surface form of the earlier act might be used to accomplish either $a_{observed}$ or $a_{intended}$.[32]

31 Although, like *expectedReply*, *active* is a default, *active* will take precedence over *expectedReply*, because it has been given a higher priority on the assumption that memory for suppositions is stronger than expectation.
32 It is possible that the same surface form might accomplish several different discourse acts, in which case it might be desirable to evaluate the likelihood of alternative choices. The work discussed by Reithinger and Maier (1995), for example, found statistical regularities in the misinterpretations that occurred in their corpus of appointment-scheduling dialogues.

The defaults that characterize misunderstandings have a lower priority than the metaplans, because speakers consider misunderstandings only when no coherent interpretation is possible. The preference for coherent interpretations is especially important when there is more than one discourse-level act for which the utterance is a possible decomposition.

Table 1

Name	Plan adoption
Purpose	Introducing a new goal
Axiom	DEFAULT $(3, \mathit{adoptPlan}(s_1, s_2, a_1, a_2, ts))$: $\mathit{hasGoal}(s_1, \mathbf{do}(s_2, a_2), ts)$ $\wedge\ \mathit{wouldExpect}(s_1, \mathbf{do}(s_1, a_1), \mathbf{do}(s_2, a_2))$ $\supset \mathit{shouldTry}(s_1, s_2, a_1, ts)$. FACT $\neg \mathit{lintentionsOk}(a_1, ts)$ $\supset \neg \mathit{adoptPlan}(s_1, s_2, a_1, a_2, ts)$.
Summary	Speaker s_1 should do action a_1 in discourse ts when: 1. s_1 wants speaker s_2 to do action a_2; 2. s_1 would expect a_2 to follow an action a_1; and 3. s_1 may adopt the plan of performing a_1 to trigger a_2 (*i.e.,* the linguistic intentions of a_1 are compatible with ts).

Table 2

Name	Acceptance
Purpose	Producing an expected reply
Axiom	DEFAULT $(2, \mathit{acceptance}(s_1, a_{reply}, ts))$: $\mathit{expected}(s_1, a_{reply}, ts)$ $\supset \mathit{shouldTry}(s_1, s_2, a_{reply}, ts)$. FACT $\mathit{active}(\mathbf{do}(s_1,\ a),\ ts)$ $\supset \neg \mathit{acceptance}(s_1,\ a,\ ts)$.
Summary	Speaker s_1 should do action a_{reply} in discourse ts when: 1. s_1 expects a_{reply} to occur next; and 2. s_1 may accept the interpretation corresponding to ts.

Table 3

Name	Fourth-turn repair
Purpose	Recovering from one's own misunderstanding
Axiom	DEFAULT $(2, makeFourthTurnRepair(s_1, s_2, a_{reply}, ts, ts_{reconstructed}))$: $active(\mathbf{mistake}(s_1, a_{intended}, a_{observed}), ts)$ $\wedge\ reconstruction(ts, ts_{reconstructed})$ $\wedge\ expected(s_1, a_{reply}, ts_{reconstructed})$ $\supset shouldTry(s_1, s_2, a_{reply}, ts)$. FACT $active(\mathbf{do}(s_1,\ a),\ ts)$ $\supset \neg makeFourthTurnRepair(s_1, s_2, a, ts, ts_{reconstructed})$.
Summary	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; and 4.s may perform a fourth-turn repair.

Table 4

Name	Third-turn repair
Purpose	Recovering from another speaker's misunderstanding
Axiom	DEFAULT $(2, makeThirdTurnRepair(s_1, s_2, a_{reply}, ts))$: $active(\mathbf{mistake}(s_2, a_{intended}, a_{observed}), ts)$ $\wedge\ a = \mathbf{inform}(s_1, s_2, \mathbf{do}(s_1, a_{intended}))$ $\wedge\ wouldExpect(s_1, do(s_1, a_{intended}), do(s_2, a_{reply}))$ $\supset shouldTry(s_1, s_2, a, ts)$. FACT $\neg lintentionsOk(a_{reply}), ts)$ $\supset \neg makeThirdTurnRepair(s_1, s_2, a_{reply}, ts)$
Summary	Speaker s_1 should initiate a repair in discourse ts (that speaker s_2 will later complete) by s_1 telling s_2 that she had performed the action $a_{intended}$ if: 1.s_2 has apparently mistaken an instance of act $a_{intended}$ for act $a_{observed}$; 2.s_1 would expect a_{reply} to follow $a_{intended}$; and 3.s_1 may perform a third-turn repair (*i.e.,* it would be reasonable and compatible for s_2 to perform a_{reply}).

Table 5

Name	Self-misunderstanding
Purpose	Detecting one's own misunderstanding
Axiom	DEFAULT $(3, selfMisunderstanding(s_1, s_2, p_{mistake}, a_{new}, ts))$: $active(\mathbf{do}(s_1, a_{observed}), ts)$ $\wedge\ lintention(a_{new}, p_l)$ $\wedge\ lintention(a_{observed}, p_{l_2})$ $\wedge\ inconsistentLI(p_l, p_{l_2})$ $\wedge\ ambiguous(a_{observed}, a_{intended})$ $\wedge\ p_{mistake} = \mathbf{mistake}(s_2, a_{intended}, a_{observed})$ $\supset try(s_1, s_2, a_{new}, ts)$. FACT $\neg(selfMisunderstanding(s_1, s_2, p_{mistake}, a_1, ts)$ $\wedge\ shouldTry(s_1, s_2, a_1, ts))$. FACT $\neg(selfMisunderstanding(s_1, s_2, p_{mistake}, a_1, ts)$ $\wedge\ ambiguous(a_1, a_2)$ $\wedge\ shouldTry(s_1, s_2, a_2, ts))$.
Summary	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 $a_{observed}$.

Table 6

Name	Other-misunderstanding
Purpose	Detecting another's misunderstanding
Axiom	DEFAULT $(3, otherMisunderstanding(s_1, s_2, p_{mistake}, a_{new}, ts))$: $active(\mathbf{do}(s_2, a_{intended}), ts)$ $\wedge\ ambiguous(a_{intended}, a_{similar})$ $\wedge\ wouldExpect(s_1, \mathbf{do}(s_2, a_{similar}), \mathbf{do}(s_1, a_{new}))$ $\wedge\ p_{mistake} = \mathbf{mistake}(s_1, a_{intended}, a_{similar})$ $\supset try(s_1, s_2, a_{new}, ts)$. FACT $otherMisunderstanding(s_1, s_2, p_{mistake}, a_1, ts)$ $\wedge\ ambiguous(a_1, a_2)$ $\supset \neg shouldTry(s_1, s_2, a_2, ts)$.
Summary	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 $a_{similar}$.

4. A detailed example

To show how our abductive account of repair works, we offer two examples that show repair of self-misunderstanding and other-misunderstanding, respectively. Here we will consider Example 2 from Russ's perspective, considering in detail Russ's reasoning about each turn and show an output trace from our implemented system. From Russ's perspective, this example demonstrates the detection of a self-misunderstanding and the production of a fourth-turn repair. In Appendix A we show the system's output for a third-turn repair, interleaving the perspectives of its two participants.

4.1 Overview

We now repeat Example 2:

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 the input we represent this dialogue as the following sequence.

T1	**m:**	**surface-request(m, r, informif(r, m, knowref(r, whoIsGoing)))**
T2	**r:**	**surface-request(r, m, informref(m, r, whoIsGoing))**
T3	**m:**	**surface-inform(m, r, not knowref(m, whoIsGoing))**
T4	**r:**	**surface-informref(r, m, whoIsGoing)**

From Russ's perspective, these utterances had the following discourse-level interpretations at the time each was produced:

T1	**m:**	**pretell(m, r, whoIsGoing)**
T2	**r:**	**askref(r, m, whoIsGoing)**
T3	**m:**	**inform(m, r, not knowref(m, whoIsGoing))**
T4	**r:**	**informref(r, m, whoIsGoing)**

After Russ hears T3, he decides that his interpretation of Mother's first turn as a **pretelling** is incorrect. This revision then leads him to reinterpret it as an **askref** and to provide a new response.

We will now show how Russ's beliefs might progress this way. In particular, we shall address the following questions:

- How Russ decides, after first concluding that T1 was a **pretelling**, that he will respond with an **askref**.
- How Russ decides, after hearing mother's response T3, that his earlier decision was incorrect.
- How Russ decides to produce an **informref** in T4.

Figures 6, 7, 9, and 10 will show the output of the system for each of the four turns of this dialogue, from Russ's perspective.

4.2 Initial assumptions

For this example, we shall assume that Russ believes that he knows who is going to the meeting (but also allows that Mother's knowledge about the meeting would be more accurate than his own). For simplicity, we represent these beliefs as facts.[33]

FACT *believe*(**r, knowref(r, whoIsGoing)).**
FACT *believe*(**r, knowsBetterRef(m,r,whoIsGoing)).**

We also assume that Russ believes that he knows whether (or not) he knows.

FACT *believe*(**r, knowif(r,knowref(r,whoIsGoing))).**

Lastly, we assume that he has linguistic expectations regarding **pretell**, **askref**, and **askif** as in Section 2.1.[34]

4.3 Turn 1: Russ decides that Mother is pretelling

According to the model, after Russ hears Mother's **surface-request**, "Do you know who is going to that meeting?", he interprets it by attempting to construct a plausible explanation of it. This requires tentatively choosing a discourse-level act on the basis of the decomposition relation and then attempting to abduce either that it is an intentional display of understanding or that it is a symptom of misunderstanding. Theorist is called to explain the utterance and returns with a list of assumptions that were made to complete the explanation. (The portion of the output from the update describes Russ's interpretation of this explanation; see Figure 6.)

In this simulation, T1 was explained as an intentional pretelling. The explanation contains the metaplanning assumption that Mother was pretelling as part of a plan to get Russ to ask a question. The reasoner also attributed to her the linguistic intentions of pretelling. We will now consider the complete explanation in detail.

Inference begins with a call to Theorist to explain the input,

utter(**m, r, surface-request(m,r,informif(r,m,knowref(r,whoIsGoing))),ts(0))**

This utterance must be explained by finding a discourse-level speech act that it might accomplish and a metaplan or misunderstanding that would explain this act. This makes use of the following default:

$$\begin{aligned}\text{DEFAULT } & (3, pickForm(s_1, s_2, a_{surfaceForm}, a, ts)) : \\ & decomp(a_{surfaceForm}, a) \\ & \wedge\ try(s_1, s_2, a, ts) \\ & \supset\ utter(s_1, s_2, a_{surfaceForm}, ts).\end{aligned}$$

To satisfy the first premise, the reasoner would need to find a speech act that is related to the surface form by the *decomp* relation, for example, either an **askif**, **askref**, or a **pretelling**:

decomp(**surface-request(m, r, informif(r, m, knowref(r, whoIsGoing))),**
pretell(m, r, whoIsGoing))

33 We might have used priorities to express different degrees of belief.
34 To keep this example of manageable size, we will not assume that he has any expectations regarding **testif** or **testref**, although in life he would.

```
| ?- startDialogue2.

      >>>surface-request(m,r,informif(r,m,knowref(r,whoIsGoing)))

***Interpreting Utterance***

Explaining
  utter(m,r,surface-request(m,r,informif(r,m,knowref(r,whoIsGoing))),ts(0))

Is formula
  pickForm(m,r,surface-request(m,r,informif(r,m,knowref(r,whoIsGoing))),
           pretell(m,r,whoIsGoing),ts(0)) ok (y/n)?y.

Explanation:
  intentionalAct(m,r,pretell(m,r,whoIsGoing),ts(0))
  adoptPlan(m,r,pretell(m,r,whoIsGoing),askref(r,m,whoIsGoing),ts(0))
  credulousB(m,knowsBetterRef(m,r,whoIsGoing))
  credulousI(m,ts(0))
  pickForm(m,r,surface-request(m,r,informif(r,m,knowref(r,whoIsGoing))),
           pretell(m,r,whoIsGoing),ts(0))

***Updating Discourse Model***

Interpretation: pretell(m, r, whoIsGoing) (turn number 1)
  expressed(do(m, pretell(m, r, whoIsGoing)), 1)

Linguistic Intentions of pretell(m,r,whoIsGoing):
  knowref(m,whoIsGoing)
  knowsBetterRef(m,r,whoIsGoing)
  intend(m,do(m,informref(m,r,whoIsGoing)))
  intend(m,knowref(r,whoIsGoing))

Suppositions Added:
  expressed(knowref(m, whoIsGoing), 1)
  expressed(knowsBetterRef(m, r, whoIsGoing), 1)
  expressed(intend(m, do(m, informref(m, r, whoIsGoing))), 1)
  expressed(intend(m, knowref(r, whoIsGoing)), 1)
Agent m adopted plan to achieve: askref(r,m,whoIsGoing)
```

Figure 6
The output for turn 1 from Russ's perspective

decomp(**surface-request(m, r, informif(r, m, knowref(r, whoIsGoing))),**
askref(m, r, whoIsGoing))

decomp(**surface-request(m, r, informif(r, m, knowref(r, whoIsGoing))),**
askif(m, r, knowref(r, whoIsGoing)))

In this case, the possibility that Mother is attempting a pretelling was considered. (The system uses an oracle, represented by the default *pickForm*, to simulate this choice.[35]) It is important to note that this is just one of the possible explanations available to Russ. Nothing in his beliefs rules out abducing explanations from either the **askif** or the **askref** interpretation.

To satisfy the second premise of the rule, the reasoner must explain:

try(**m, r, pretell(m, r, whoIsGoing), ts(0)**)

Two kinds of explanation might be possible: a hearer might assume that the act fulfills the speaker's intention to coherently extend the discourse as he has understood it or he might assume that one of the two types of misunderstanding has occurred.[36] If a discourse has just begun, then any utterance that starts an adjacency pair will be coherent. In this case, Russ finds that the former type of explanation is possible using the metaplan (for plan adoption) to explain *shouldTry*(**m, r, pretell(m, r, whoIsGoing), ts(0)**). The relevant defaults are repeated here:

DEFAULT $(1, intentionalAct(s_1, s_2, a, ts))$:
$shouldTry(s_1, s_2, a, ts)$
$\supset try(s_1, s_2, a, ts)$.

DEFAULT $(3, adoptPlan(s_1, s_2, a_1, a_2, ts))$:
$hasGoal(s_1, \mathbf{do}(s_2, a_2), ts)$
$\wedge\ wouldExpect(s_1, \mathbf{do}(s_1, a_1), \mathbf{do}(s_2, a_2))$
$\supset shouldTry(s_1, s_2, a_1, ts)$.

The conditions of the metaplan are satisfiable because there is a plausible goal act that a pretelling would help Mother to achieve and it is consistent for Russ to assume that achieving this act was, in fact, her goal.[37] Also, when we consider possible evidence against Mother adopting this plan, namely whether the linguistic intentions of pretelling were incompatible with those that have been expressed, it would be consistent to assume that Mother is intending this plan.

Russ infers

wouldExpect(r, pretell(m, r, whoIsGoing), askref(r, m, whoIsGoing))

because he has a linguistic expectation to that effect:

35 This oracle thus allows the analyst to test different interpretations.
36 The former possibility admits that an utterance that displays a misconception, such as a mistaken belief about initial knowledge, might still be coherent, unless such knowledge has been introduced into the discourse explicitly. Misconceptions are addressed by second-turn repairs, which are not considered here.
37 Because Russ's previous utterance had not been the first part of an adjacency pair, he cannot explain her utterance as acceptance or challenge.

FACT *lexpectation*(**do(m, pretell(m, r, whoIsGoing)),
knowsBetterRef(m, r, whoIsGoing),
do(r, askref(r, m, whoIsGoing))).

```
Explaining shouldTry(r,m,A,ts(1)),decomp(A2,A)

Answer: shouldTry(r,m,askref(r,m,whoIsGoing),ts(1)),
        decomp(surface-request(r,m,informref(m,r,whoIsGoing)),
               askref(r,m,whoIsGoing))

Explanation:
  intentionalAct(r,m,askref(r,m,whoIsGoing),ts(1))
  acceptance(r,askref(r,m,whoIsGoing),ts(1))
  expectedReply(do(m,pretell(m,r,whoIsGoing)),
                knowsBetterRef(m,r,whoIsGoing),
                do(r,askref(r,m,whoIsGoing)),ts(1))

***Updating Discourse Model***

Interpretation: askref(r,m,whoIsGoing) (turn number 2)
  expressed(do(r,askref(r,m,whoIsGoing)),2)

Linguistic Intentions of askref(r,m,whoIsGoing):
  not knowref(r,whoIsGoing)
  and intend(r,knowref(r,whoIsGoing))
  and intend(r,do(m,informref(m,r,whoIsGoing)))

Suppositions Added:
  expressedNot(knowref(r,whoIsGoing),2)
  expressed(intend(r,knowref(r,whoIsGoing)),2)
  expressed(intend(r,do(m,informref(m,r,whoIsGoing))),2)

Agent r performed expected act: askref(r,m,whoIsGoing)
***Generating Utterance***

<<<surface-request(r, m, informref(m, r, whoIsGoing))
```

Figure 7
The output for turn 2 from Russ's perspective

4.4 Turn 2: Russ decides to respond with an askref

In turn 2, Russ produces a **surface-request**. This utterance is appropriate, independent of whether or not Russ actually wants to know who is going to the meeting, because it displays acceptance of Mother's pretelling. From Russ's perspective it displays acceptance, because a **surface-request** is one way to perform an **askref**, an act that is expected according to Russ's model of the discourse after the first turn.[38]

As shown in Figure 7, Theorist finds that if Russ accepts Mother's **pretelling**, he should perform an **askref**. An **askref** would demonstrate acceptance because it is the expected next act. The derivation of this act relies on the rule for intentional action shown earlier in Section 4.3, along with the metaplan for acceptance repeated here:

38 If, for some reason, Russ did *not* want to know the information, he might decide not to produce an **askref**. However, he would then be accountable for justifying his action as well as for displaying his acceptance of Mother's displayed understanding (*e.g.,* by including an explicit rejection of her offer); otherwise she might think that one of them has misunderstood.

If we assume that Mother produced the first turn as an **askif**, she might also hear T2 as an intentional **askref**, but for a reason different than Russ would. Her explanation would include the metaplanning assumption that he was doing so as part of an adopted plan to get her to produce an **informref**. Although T2 might also be explained by abducing that Russ misunderstood T1 as an attempted **pretelling**, we see that she considers this explanation to be less likely because otherwise she would have been more inclined to make T3 a third-turn repair ("No, I'm asking you").[40]
Plan adoption (see Table 1) provides Mother a plausible explanation for T2 because:

1. **wouldExpect(r, askref(r, m, whoIsGoing), informref(m, r, whoIsGoing))** is explained because Mother has a linguistic expectation that says that an **askref** normally creates an expectation for the listener to tell the speaker the answer:

   FACT *lexpectation*(**do(r, askref(r, m, whoIsGoing)),**
   knowref(m, whoIsGoing),
   do(m, informref(m, r, whoIsGoing))).

2. Mother's credulousness about Russ's goals explains her belief that he wants her to perform the expected **informref**.
3. The linguistic intentions of **askref** are compatible with those that have been expressed, so it is consistent to assume that Russ is intending to use it as part of a plan. (They are consistent with the context because T1 expresses only that Mother does not know *whether* Russ knows and not that she does not herself know.)
4. Thus, by 1–3 and the metaplan for plan adoption, *shouldTry*(**r, m, askref(r, m, whoIsGoing), ts(0)**) is explainable.

Assuming this interpretation, Mother can then demonstrate acceptance using an **inform-not-knowref**.

Figure 8
How Mother interprets T2

DEFAULT $(2, acceptance(s_1, a_{reply}, ts))$:
$expected(s_1, a_{reply}, ts)$
$\supset shouldTry(s_1, s_2, a_{reply}, ts)$.

The **askref** would be expected (see Section 3.3.3) because:[39]

- According to the discourse model, it is true that *active*(**do(m pretell(m, r, whoIsGoing)), ts(1)**).
- There is a linguistic expectation that **askref** follow **pretell**.
- Russ believes the conditions of this relation: **knowsBetterRef(m, r, whoIsGoing)**.
- The linguistic intentions of **askref** are compatible with those already expressed.

4.5 Turn 3: Russ decides that his interpretation of Turn 1 was wrong

Mother replies with a **surface-inform**. This is interpreted as a discourse-level **inform-not-knowref**. This act signals a misunderstanding, because the linguistic intentions

39 See Figure 8 for how Mother might interpret this turn.

```
    >>>surface-inform(m, r, not knowref(m, whoIsGoing))

***Interpreting Utterance***

Explaining utter(m,r,inform(m,r,not knowref(m,whoIsGoing)),ts(2))

Is formula
  pickForm(m,r,surface-inform(m,r,not knowref(m,whoIsGoing)),
           inform(m,r,not knowref(m,whoIsGoing)),ts(2)) ok (y/n)?y.

Explanation:
  selfMisunderstanding(m,r,mistake(r,askref(m,r,whoIsGoing),
                                     pretell(m,r,whoIsGoing)),
                       inform(m,r,not knowref(m,whoIsGoing)),ts(2))
  persists(do(m,pretell(m,r,whoIsGoing)),2)
  pickForm(m,r,surface-inform(m,r,not knowref(m,whoIsGoing)),
           inform(m,r,not knowref(m,whoIsGoing)),ts(2))

***Updating Discourse Model***

Interpretation:
  inform(m, r, not knowref(m, whoIsGoing)) (turn number 3)
  expressed(do(m, inform(m, r, not knowref(m, whoIsGoing))), 3)

Linguistic Intentions of inform(m,r,not knowref(m,whoIsGoing)):
  not knowref(m,whoIsGoing)
  intend(m,knowif(r,not knowref(m,whoIsGoing)))

Suppositions Added:
  expressed(mistake(r, askref(m, r, whoIsGoing),
                       pretell(m, r, whoIsGoing)),3)
  expressedNot(knowref(m, whoIsGoing), 3)
  expressed(intend(m, knowif(r, not knowref(m, whoIsGoing))), 3)
Agent r misunderstood act do(m, askref(m, r, whoIsGoing))
as do(m, pretell(m, r, whoIsGoing))
```

Figure 9
The output for turn 3 from Russ's perspective

associated with it are incompatible with those previously assumed, ruling out an explanation that uses the default for intentional acts.[41]

Figure 9 shows that Theorist abduces that T3 is attributable to a misunderstanding on Russ's part, in particular, to his having incorrectly interpreted one of Mother's utterances as a **pretelling**, rather than as an **askref**. This explanation succeeded because each of the conditions of the default for self-misunderstanding were explainable. Below we will repeat this rule and then sketch the proof, considering each of the premises in the default.

DEFAULT $(3, selfMisunderstanding(s_1, s_2, p_{mistake}, a_{new}, ts))$:
$active(\mathbf{do}(s_1, a_{observed}), ts)$
$\wedge\ lintention(a_{new}, p_l)$
$\wedge\ lintention(a_{observed}, p_{l_2})$
$\wedge\ inconsistentLI(p_l, p_{l_2})$
$\wedge\ ambiguous(a_{observed}, a_{intended})$
$\wedge\ p_{mistake} = \mathbf{mistake}(s_2, a_{intended}, a_{observed})$
$\supset try(s_1, s_2, a_{new}, ts)$.

FACT $\neg(selfMisunderstanding(s_1, s_2, p_{mistake}, a_1, ts)$
$\wedge\ shouldTry(s_1, s_2, a_1, ts))$.

FACT $\neg(selfMisunderstanding(s_1, s_2, p_{mistake}, a_1, ts)$
$\wedge\ ambiguous(a_1, a_2)$
$\wedge\ shouldTry(s_1, s_2, a_2, ts))$.

Premise 1: A pretelling was active in **ts(2)**, because of Russ's interpretation of T1.[42]

Premises: 2–4: A pretelling would be incompatible with an **inform-not-knowref** happening now. The linguistic intentions of the pretelling are:

and(knowref(m, whoIsGoing),
and(knowsBetterRef(m, r, whoIsGoing),
and(intend(m, do(m, informref(m, r, whoIsGoing))),
intend(m, knowref(r, whoIsGoing)))))

The linguistic intentions of inform-not-knowref are:

and(not knowref(m, whoIsGoing),
intend(m, knowif(r, not knowref(m, whoIsGoing)))).

But these intentions are inconsistent, because **knowref(m, whoIsGoing)** and **not knowref(m, whoIsGoing)** are incompatible. As a result, *inconsistentLI* holds for these linguistic intentions.

41 For T3 to have demonstrated acceptance, the linguistic intentions of **inform(m, r, not knowref(m, whoIsGoing))** would need to have been compatible with this interpretation of the discourse. However, **not knowref(m, whoIsGoing)** is among these intentions, while *active*(**knowref(m, whoIsGoing),ts(2)**). As a result, T3 cannot be attributed to any expected act, and must be attributed to a misunderstanding either by Russ or by Mother.

42 In the discourse model, this was expressed as *expressed*(**do(m, pretell(m, r, whoIsGoing)), 0**), from which one can assume *persists*(**do(m, pretell(m, r, whoIsGoing)), 2**) by default.

Premise 5: This is a plausible mistake because the acts **pretell** and **askref** both have the same surface form:

surface-request(m, r, informif(r, m, knowref(r, whoIsGoing)))

So, *ambiguous*(**pretell(m, r, whoIsGoing), askref(m, r, whoIsGoing)**).

The constraints: There is no other coherent interpretation, so it is consistent to assume that a misunderstanding occurred:

selfMisunderstanding(m,r,
mistake(r,askref(m, r, whoIsGoing),
pretell(m, r, whoIsGoing)),
inform(m, r, not knowref(m, whoIsGoing)),
ts(2)).

Thus, *try*(**m, r, inform(m, r,not knowref(m, whoIsGoing)),ts(2)**) is explained. As a result of this interpretation, **not knowref(m, whoIsGoing)** is added to the discourse model as the fact *expressedNot*(**knowref(m, whoIsGoing)**). This addition terminates the activation of **knowref(m, whoIsGoing)** from the first turn. (At the same time, if Russ had revised any of his real beliefs on the basis of the first turn, he might now reconsider those revisions; however, our theory does not account for this.)

4.6 Turn 4: Russ performs a repair

After revising his understanding of Turn 1, Russ performs a **surface-informref** that displays his acceptance of the revised interpretation. When Theorist is called to find a coherent discourse-level act (*i.e.,* by using the default for intentional acts) it finds that Russ can perform a fourth-turn repair. The metaplan for this repair, repeated below, is similar to that for acceptance, but involves the reconstruction of the discourse model.

DEFAULT $(2, makeFourthTurnRepair(s_1, s_2, a_{reply}, ts))$:
$active(\mathbf{mistake}(s_1, a_{intended}, a_{observed}), ts)$
$\wedge\ reconstruction(ts, ts_{reconstructed})$
$\wedge\ expected(s_1, a_{reply}, ts_{reconstructed})$
$\supset\ shouldTry(s_1, s_2, a_{reply}, ts)$.

This metaplan applies because Russ had misunderstood a prior utterance by Mother, a reconstruction of the discourse is possible, and, within the reconstructed discourse, an **informref** is expected (as a reply to the misunderstood **askref**).[43]

An **informref** by Russ is expected (see Section 3.3.3) in the reconstructed dialogue because:

- There is a linguistic expectation corresponding to the adjacency pair *askref–informref*;
- Russ believes its conditions; and

43 From Mother's perspective, if indeed she did make an **askif** in T1, T4 can be seen as a display of acceptance of it, because a **surface-informref** is one way to do an **informif**. Thus, from her perspective, she need never recognize that Russ has misunderstood.

- The linguistic intentions of **informref** are compatible with the reconstruction.

```
Explaining shouldTry(r,m,A,ts(3)), decomp(As,A)

***Reconstructing Turn Number 1***

Suppositions Added:
  expressed(do(m, askref(m, r, whoIsGoing)), alt(1))
  expressedNot(knowref(m, whoIsGoing), alt(1))
  expressed(intend(m, knowref(m, whoIsGoing)), alt(1))
  expressed(intend(m, do(r, informref(r, m, whoIsGoing))), alt(1))

*********************************

Answer: shouldTry(r,m,informref(r,m,whoIsGoing),ts(3)),
        decomp(surface-informref(r,m,whoIsGoing),
               informref(r,m,whoIsGoing)

Explanation:
  intentionalAct(r,m,informref(r,m,whoIsGoing),ts(3))
  makeFourthTurnRepair(r,m,informref(r,m,whoIsGoing),ts(3),ts(1))
  reconstruction(ts(3),ts(alt(1)))

***Updating Discourse Model***

Interpretation: informref(r,m,whoIsGoing) (turn number 4)
  expressed(do(r,informref(r,m,whoIsGoing)),4)

Linguistic Intentions of informref(r,m,whoIsGoing):
  knowref(r,whoIsGoing)and intend(r,knowref(m,whoIsGoing))

Suppositions Added:
  expressed(knowref(r,whoIsGoing),4)
  expressed(intend(r,knowref(m,whoIsGoing)),4)

r performed fourth turn repair
***Generating Utterance***

<<<surface-informref(r,m,whoIsGoing)
```

Figure 10
The output for turn 4 from Russ's perspective

5. Related work

5.1 Accounts based on plan recognition

Plan-based accounts interpret speech acts by chaining from subaction to action, from actions to effects of other actions, and from preconditions to actions to identify a plan (*i.e.,* a set of actions) that includes the observed act. Heuristics are applied to discriminate among alternatives.

5.1.1 Allen and Perrault. Allen and Perrault (1979), Perrault and Allen (1980) show how plan recognition can be used to understand indirect speech acts (such as the use of "Can you pass the salt?" as a polite request to pass the salt). To interpret an utterance, the approach applies a set of context-independent inference rules to identify all plausible plans. For example, one rule says that if a speaker wants to know the truth value of some proposition then she might want the proposition to be made true. The final interpretation is then determined by a set of rating heuristics, such as "Decrease the rating of a path if it contains an action whose effects are already true at the time the action starts." These rating heuristics are problematic because they conflate linguistic and pragmatic knowledge with knowledge about the search mechanism itself. This approach also depends on there being only a few relationships between utterances and plans and is unable to handle any utterances that do not directly relate to the domain plan.

Although we have not yet considered the problem of indirect utterances in detail, we anticipate that such explanations might include as a subtask the kind of plan-based inference that has been proposed, but this inference would be limited by the hearer's own goals and expectations. However, many common uses of indirectness can be explained by the existence of a well-accepted social convention that makes them expected.

5.1.2 Litman. Work by Litman (1986) attempts to overcome some of the limitations of Allen and Perrault's approach by extending the plan hierarchy to include discourse-level *metaplans*, in addition to domain-level plans. Metaplans include actions, such as *introduce*, *continue*, or *clarify* and are recognized in part, by identifying cue phrases. Although the metaplans add flexibility by increasing the number of possible paths, they also add to the problem of pruning and ordering the paths, requiring additional heuristics. For example, there are specific rules for choosing among alternative metaplans on the basis of clue words, implicit expectations, or default preferences. Litman also adds a new general heuristic: stop chaining if an ambiguity cannot be resolved.

5.1.3 Carberry and Lambert. Carberry (1985), (1987), (1990) uses a similar approach. Her model introduces a new set of discourse-level goals such as *seek-confirmation* that are recognized on the basis of the current properties of the dialogue model and the mutual beliefs of the participants. Once a discourse-level goal is selected, a set of candidate plans is identified, and Allen-style heuristics are applied to choose one of them. Subsequent work by Lambert and Carberry (1991), (1992) introduces an intermediate, problem-solving level of plans that link the discourse-level acts to domain plans. The processing rules, by their specificity, eliminate the need for many of the heuristics. The sacrifice here is a loss of generality; the mechanisms for recognizing goals are specific to Carberry's implementation.

5.1.4 Cawsey. Cawsey (1991) proposes a method of extending Perrault and Allen's (1980) inference rule approach to produce repairs. She also suggests including some of the information captured by various rating heuristics as premises in the rules, allowing that these new premises may be assumed by default. For example, the following rule is proposed for capturing pretellings:

```
if request(S1, S2, informif(S2, S1, knowref(S2, D)))
   and know(S2, knowref(S1, D))
then know(S2, wants(S1, knowref(S2, D)))
```

To handle misunderstandings, she suggests that these defaults can be retracted if necessary and any utterances whose interpretations depend on that belief are simply reinterpreted. Allowing defeasible beliefs is a step in the right direction; however, the approach still misses the point that participants are able to *negotiate* meanings. Preconditions as `know(S2, not knowref(S1, D))` influence interpretations only to the extent that they provide support for or evidence against a particular (abductive) explanation. In Example 2, even if Mother knew who was going, she could still be asking Russ a question, albeit insincerely. Similarly, even if Russ suspected that Mother did not know who was going, he might still have chosen to treat her utterance as a pretelling, perhaps to confirm his suspicions or to delay answering.

5.1.5 Traum and Hinkelman. Hinkelman's (1990) work incorporates some abductive reasoning in her model of utterance interpretation. The model treats different features in the input, such as the mood of a sentence or the presence of a particular lexical item, as manifestations of different speech acts. During interpretation, procedures that test for particular features of the input suggest candidates. The system then removes any candidates whose implicatures are inconsistent with prior beliefs.

Traum and Hinkelman (1992) extend this work by generalizing the notion of speech act to *conversation* act. Conversation acts are not assumed to be understood without some positive evidence by the hearer and are formalized in a multi-agent logic. In addition, the theory now includes grounding acts such as for taking and releasing turns, and initiating, clarifying, or acknowledging an utterance; however, unlike the speech acts, grounding acts are organized into a finite-state grammar. In conversation, grounding acts that violate the grammar are not recognized. Traum and Hinkelman suggest that such violations should be used to trigger a repair, but admit that, except when a repair has been requested explicitly, the model itself says nothing about when a repair should be uttered (p. 593).[44]

Traum and Allen (1994) extend the work to include a notion of social obligation, which serves much the same purpose as expectations in our model.

5.2 Other expectation-driven accounts

Within the speech understanding community, the word "expectation" has been used differently from our use here. Expectation in the speech context refers to what the next word or utterance is likely to be *about*.[45] For example, after the computer asks the user to perform some action A, it might expect any of the following types of responses:

1. A statement about background knowledge that might be needed.
2. A statement about the underlying purpose of A.
3. A statement about related task steps (*i.e.,* subgoals of A, tasks that contain A as a step, or tasks that might follow A).
4. A statement about the accomplishment of A.

These expectations are independent of the belief state of an agent and are specified down to the semantic (and sometimes even lexical) level. This information has long been used to discriminate between ambiguous interpretations and correct mistakes made by the speech recognizer (Fink and Biermann 1986; Smith 1992). Typically, an utterance will be interpreted according to which expectation matches it most closely. By contrast, our approach and that of the plan-based accounts use "expectation" to refer to agents' beliefs about how future utterances might *relate* to prior ones. These expectations are determined both by an agent's understanding of typical behavior and by his or her mental state. These two notions of expectation are complimentary and any dialogue model that uses speech as input must be able to represent and reason with both.

44 Interpretations that have the right pragmatic force but inconsistent implicatures are still ruled out as in Hinkelman (1990); however, as with grounding acts, presumably the elimination of all possible interpretations could cue some type of repair mechanism, if they chose to incorporate one.

45 This is the same sense of "expectation" as used by Riesbeck (1974).

5.3 Approaches to misconception

Misconceptions are a deficit in an agent's knowledge of the world; they can become a barrier to understanding if they cause an agent to unintentionally evoke a concept or relation. To prevent misconceptions from triggering a misunderstanding, agents can check for evidence of misconception and try to resolve apparent errors. The symptoms of misconception include references to entities that do not map to previously known objects or operations (Webber and Mays 1983) or requests for clarification (Moore, 1989). Errors are corrected by replacing or deleting parts of the problematic utterance so that it makes sense. Several correction strategies have been suggested:

- Generalize a description by selectively ignoring some constraints (see Goodman 1985; McCoy 1985, 1986, 1988; Carberry 1988; Calistri-Yeh 1991; Eller and Carberry 1992),
- Make a description more specific by adding extra constraints (see Eller and Carberry 1992), and
- Choose a conceptual "sibling", by combining generalization and constraint operations. For example, if there is more than one strategy for achieving a goal, then an entity that corresponds to a step from one strategy might be replaced by one corresponding to a step of one of the other strategies (see Carberry 1985, 1987; Eller and Carberry 1992; Moore 1989).

Although these approaches do quite well at preventing certain classes of misunderstandings, they cannot prevent them all. Moreover, these approaches may actually trigger misunderstandings because they always find *some* substitution, and yet they lack any mechanisms for detecting when one of their own previous repairs was inappropriate. Thus, a conversational participant will still need to be able to address actual misunderstandings.

5.4 Collaboration in the resolution of nonunderstanding

In this paper, we have concentrated on the repair of *mis*-understanding. Our colleagues Heeman and Edmonds have looked at the repair of *non*-understanding. The difference between the two situations is that in the former, the agent derives exactly one interpretation of an utterance and hence is initially unaware of any problem; in the latter, the agent derives either more than one interpretation, with no way to choose between them, or no interpretation at all, and so the problem is immediately apparent. Heeman and Edmonds looked in particular at cases in which a referring expression uttered by one conversant was not understood by the other (Heeman and Hirst 1995; Edmonds 1994; Hirst et. al. 1994). Clark and his colleagues (Clark and Wilkes-Gibbs 1986; Clark 1993) have shown that in such situations, conversants will collaborate on repairing the problem by, in effect, negotiating a reconstruction or elaboration of the referring expression. Heeman and Edmonds model this with a plan recognition and generation system that can recognize faulty plans and try to repair them. Thus (as in our own model) two copies of the system can converse with each other, negotiating referents of referring expressions that are not understood by trying to recognize the referring plans of the other, repairing them where necessary, and presenting the new referring plan to the other for approval.

6. Conclusions

In human dialogues, both the producer and the recipient of an utterance have a say in determining its interpretation. Moreover, they may both change their minds in the face of new information. Dialogue participants are able to *negotiate* the meaning of utterances because in responding to what the hearer decides are the speaker's goals and expectations regarding an utterance, the hearer also provides evidence of that decision and hence constraints on what the speaker may do next. If the speaker disagrees with a displayed interpretation, she can challenge it directly or decide to respond in such a way that the hearer must infer a misunderstanding.

The long-term goal of our work is to construct a model of communicative interaction that will be able to support the negotiation of meaning. We have considered the information sources and reasoning processes that agents need to determine their beliefs about the goals and expectations associated with each other's utterances. Whereas previous models of dialogue tend to represent discourse meaning from some global perspective, make use of either purely structural or purely intentional information, and give minimal attention to repair, in our model:

- Each agent has his or her own model of the discourse.
- Agents rely on both structural and intentional information in the discourse.
- Agents distinguish between intended actions and misunderstandings.
- Agents interpret utterances on the basis of expectations derived from previous utterances as well as expectations for future actions that are predicted by the utterance under interpretation.
- Agents are able to detect and repair their own misunderstandings as well as those of others.

We see this work as providing some of the first steps toward a unified account of interpretation, generation, and repair.

The primary contributions of this work have been to treat misunderstanding and repair as intrinsic to conversants' core language abilities and to account for them with the same processing mechanisms that account for normal speech. In particular, both interpretation and repair are treated as explanation problems, modeled as abduction. In order to account for the repair of misunderstandings, we have proposed a representation of the discourse that captures the agent's interpretation of the conversation both before and after a repair and which is independent of the actual beliefs of the participants—a dynamic mental artifact that is the *object* of belief and repair. With such a record of the discourse, agents can refer to alternative interpretations or to the repair process itself, potentially enabling them to recover from rejected repairs. By addressing the problem of repair, this work should facilitate efforts to build natural language interfaces that can better recover from their own mistakes as well as those of their users.

Acknowledgements

This work was supported by the Natural Sciences and Engineering Research Council of Canada. We thank Ray Reiter for his suggestion that we use abduction and James Allen for his advice about temporal logics. We thank Hector Levesque, Mike Gruninger, Sheila McIlraith, Javier Pinto, and Stephen Shapiro, for their comments on many of the formal aspects of this work. We also thank David Chapman, Susan Haller, Diane Horton, C. Raymond Perrault, Mark Steedman, Evan Steeg, and the reviewers for their comments on drafts of this paper or the thesis on which it is based (McRoy, 1993a).

References

Allen, James. 1983. Recognizing intentions from natural language utterances. In Michael Brady, Robert C. Berwick, and James Allen, editors, *Computational Models of Discourse.* The MIT Press, pages 107–166.

Allen, James and Raymond Perrault. 1979. Plans, inference, and indirect speech acts. In *17th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, pages 85–87.

Bach, Kent and Robert M. Harnish. 1979. *Linguistic Communication and Speech Acts.* The MIT Press.

Beun, Robbert-Jan. 1990. Speech acts and mental states. In *Proceedings of the Fifth Rocky Mountain Conference on Artificial Intelligence, Pragmatics in Artificial Intelligence*, pages 75–80, Las Cruces, New Mexico.

Brewka, Gerhard. 1989. Preferred subtheories: An extended logical framework for default reasoning. In *Proceedings of the 11th International Joint Conference on Artificial Intelligence*, pages 1043–1048, Detroit, MI.

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

Carberry, Sandra. 1985. *Pragmatics Modeling in Information Systems Interfaces.* Ph.D. thesis, University of Delaware, Newark, Delaware.

Carberry, Sandra. 1987. Pragmatic modeling: Toward a robust natural language interface. *Computational Intelligence*, 3(3):117–136.

Carberry, Sandra. 1988. Modeling the user's plans and goals. *Computational Linguistics*, 14(3):23–37.

Carberry, Sandra. 1990. *Plan Recognition in Natural Language Dialogue.* The MIT Press, Cambridge, MA.

Cawsey, Alison J. 1991. A belief revision model of repair sequences in dialogue. In Ernesto Costa, editor, *New Directions in Intelligent Tutoring Systems.* Springer Verlag.

Charniak, Eugene and Robert Goldman. 1988. A logic for semantic interpretation. In *26th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, pages 87–94, Buffalo, NY.

Clark, Herbert H. 1993. *Arenas of Language Use.* The University of Chicago Press, and Stanford: Center for the Study of Language and Information, Chicago.

Clark, Herbert H. and Deanna Wilkes-Gibbs. 1986. Referring as a collaborative process. *Cognition*, 22:1–39. (Reprinted in Cohen, Philip R.; Morgan, Jerry; and Pollack, Martha, (eds), *Intentions in Communication.* The MIT Press, pages 463–493.).

Cohen, Philip R. and Hector Levesque. 1985. Speech acts and rationality. In *The 23rd Annual Meeting of the Association for Computational Linguistics: Proceedings of the Conference*, pages 49–60, Chicago.

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

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

Fink, P. E. and A. W. Biermann. 1986. The correction of ill-formed input using history-based expectation with applications to speech understanding. *Computational Linguistics*, 12(1):13–36.

Fox, Barbara. 1987. Interactional reconstruction in real-time language processing. *Cognitive Science*, 11:365–387.

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

Grice, H. P. 1957. Meaning. *The Philosophical Review*, 66:377–388.

Heeman, Peter and Graeme Hirst. 1995. Collaborating on referring expressions. *Computational Linguistics*, 21(3).

Hinkelman, Elizabeth Ann. 1990. *Linguistic and Pragmatic Constraints on Utterance Interpretation.* Ph.D. thesis, Department of Computer Science, University of Rochester, Rochester, New York. Published as University of Rochester Computer Science Technical Report 288, May 1990.

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

Hobbs, Jerry R., Mark Stickel, Paul Martin, and Douglas Edwards. 1988. Interpretation as abduction. In *26th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, pages 95–103.

Hobbs, Jerry R., Mark E. Stickel, Douglas E. Appelt, and Paul Martin. 1993. Interpretation as abduction. *Artificial Intelligence*, 63:69–142.

Jose, Paul E. 1988. Sequentiality of speech acts in conversational structure. *Journal of Psycholinguistic Research*, 17(1):65–88.

Lambert, Lynn and Sandra Carberry. 1991. A tri-partite plan-based model of dialogue. In *29th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, pages 47–54, Berkeley, CA.

Lambert, Lynn and Sandra Carberry. 1992. Modeling negotiation dialogues. In *30th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, pages 193–200, Newark, Delaware.

Litman, Diane J. 1986. Linguistic coherence: A plan-based alternative. In *24th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, pages 215–223, New York.

Loveland, D. W. 1978. *Automated Theorem Proving: A Logical Basis.* North-Holland, Amsterdam, The Netherlands.

Luperfoy, Susann. 1992. The representation of multimodal user interface dialogues using discourse pegs. In *30th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, pages 22–31, Newark, Delaware.

McCoy, Kathleen F. 1985. 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.

McCoy, Kathleen F. 1986. The ROMPER system: responding to object-related misconceptions using perspective. In *24th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, pages 97–105.

McCoy, Kathleen F. 1988. Reasoning on a highlighted user model to respond to misconceptions. *Computational Linguistics*, 14(3):52–63.

McCoy, Kathleen F. 1989. Generating context-sensitive responses to object misconceptions. *Artificial Intelligence*, 41(2):157–195.

McLaughlin, Margaret L. 1984. *Conversation: How Talk is Organized.* Sage Publications, Beverly Hills.

McRoy, Susan W. 1993a. *Abductive Interpretation and Reinterpretation of Natural Language Utterances.* Ph.D. thesis, Department of Computer Science, University of Toronto, Toronto, Canada. Published as CSRI Technical Report No. 288, University of Toronto, Department of Computer Science.

McRoy, Susan W. 1993b. Belief as an effect of an act of introspection. In *1993 AAAI Spring Symposium on Reasoning about Mental States: Formal Theories and Applications*, pages 86–89, Stanford University, March.

Moore, Johanna D. 1989. Responding to "Huh?": Answering vaguely articulated follow-up questions. In *Conference on Human Factors in Computing Systems (CHI'89)*, pages 91–96, Austin, TX. ACM Press / Addison-Wesley. Also published as a special issue of *SIGCHI Bulletin* (unnumbered).

Nagata, Masaaki and Tsuyoshi Morimoto. 1993. An experimental statistical dialogue model to predict the speech act type of the next utterance. In *International Symposium on Spoken Dialogue*, Tokyo, Japan.

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

Perrault, C. Raymond and James Allen. 1980. A plan-based analysis of indirect speech acts. *Computational Linguistics*, 6:167–183.

Pollack, Martha E. 1986a. Inferring domain plans in question-answering. Technical Report TR 403, Artificial Intelligence Center, SRI International, Menlo Park, CA.

Pollack, Martha E. 1986b. A model of plan inference that distinguishes between the beliefs of actors and observers. In *Proceedings of the 24th annual meeting of the Association for Computational Linguistics*, pages 207–214, New York.

Pollack, Martha E. 1990. Plans as complex mental attitudes. In P. Cohen, J Morgan, and M. Pollack, editors, *Intentions in Communication.* MIT Press, pages 77–103.

Poole, David, Randy Goebel, and Romas Aleliunas. 1987. 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.* Springer-Verlag, New York, pages 331–352. Also published as Research Report CS-86-06, Faculty of Mathematics, University of Waterloo, February, 1986.

Reithinger, Norbert and Elisabeth Maier. 1995. Utilizing statistical dialogue act processing in verbmobil. In *Proceedings, 33rd annual meeting of the Association for Computational Linguistics*, pages 116–121, Cambridge, MA, June.

Riesbeck, C. 1974. Computational understanding: Analysis of sentences and context. Technical Report STAN-CS-74-437, Computer Science Department, Stanford University, Durham, NC.

Schegloff, Emanuel A. 1988. Presequences and indirection: Applying speech act theory to ordinary conversation. *Journal of Pragmatics*, 12:55–62.

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

Schegloff, Emanuel A. and Harvey Sacks. 1973. Opening up closings. *Semiotica*, 7:289–327.

Smith, Ronnie. 1992. *A Computational Model of Expectation-Driven Mixed-Initiative Dialog Processng.* Ph.D. thesis, Department of Computer Science, Duke University, Durham, NC.

Spencer, Bruce. 1990. *Assimilation in plan recognition via truth maintenance with reduced redundancy.* Ph.D. thesis, Department of Computer Science, University of Waterloo.

Stickel, M. E. 1989. A Prolog technology theorem prover. *Journal of Automated Reasoning*, 4:353–360.

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

Traum, David and James Allen. 1994. Discourse obligations in dialogue processing. In *32nd Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, pages 1–8, Las Cruces, NM.

Traum, David and Elizabeth Hinkelman. 1992. Conversation acts in task-oriented spoken dialogue. *Computational Intelligence*, 8(3). Special Issue: Computational Approaches to Non-Literal Language.

Tsui, Amy B. M. 1991. Sequencing rules and coherence in discourse. *Journal of Pragmatics*, 15:111–129.

Umrigar, Zerksis D. and Vijay Pitchumani. 1985. An experiment in programming with full first-order logic. In *Symposium of Logic Programming*, Boston, MA. IEEE Computer Society Press.

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

Walker, Marilyn A. 1991. Redundancy in collaborative dialogue. In *1991 AAAI Fall Symposium on Discourse Structure in Natural Language Understanding and Generation*, pages 124–129, Pacific Grove, Monterey, CA, November.

Webber, Bonnie Lynn and Eric Mays. 1983. Varieties of user misconceptions: Detection and correction. In *Proceedings of the Eighth International Joint Conference on Artificial Intelligence*, pages 650–652, Karlsruhe.

A. A third-turn repair

This appendix gives an annotated version of the output for the following example of third-turn repair (from McLaughlin (1984)). We will consider the results of a simulation in which two copies of our system, each with its own beliefs and goals, converse with each other.

For this example, we assume that A (`a`) wants B (`b`) to tell her the time of the dinner for Alfred, that she believes that she does not already know, that he knows when it is, and that he

Example 4

T1 **A:** When is the dinner for Alfred?

T2 **B:** Is it at seven-thirty?

T3 **A:** No, I'm asking you.

T4 **B:** Oh, I don't know.

believes that she does not know, of any given time (including seven-thirty), whether it is the time of the dinner. Notice that the last belief of `a`, is her belief about `b` believes about what she knows.

```
fact hasGoal(a, do(b, informref(b, a, whenD)),ts(0)).
fact believe(a, not knowref(a, whenD)).
fact believe(a, knowref(b,  whenD)).
fact believe(b, not knowif(a, dinnerAtSevenThirty)).
```

We also assume that speaker `b`, believes that he does not know the time of the dinner, but suspects it is at seven-thirty and believes that `a` does know when it is.

```
fact believe(b, not knowref(b, whenD)).
fact believe(b, suspectRef(b, whenD, dinnerAtSevenThirty)).
fact believe(b, knowref(a, whenD)).
```

We will show the output for both sides of the conversation, showing the perspective of speaker `b` as boxed text.

First, speaker `a` generates a request to try to satisfy her goal:

When is the dinner for Alfred?

```
| ?- startDialogue1.

Explaining shouldTry(a,b,A,ts(0)),decomp(U,A)

Answer:
  shouldTry(a,b,askref(a,b,whenD),ts(0)),
  decomp(surface-request(a,b,informref(b,a,whenD)),
         askref(a,b,whenD))

Explanation:
  intentionalAct(a,b,askref(a,b,whenD),ts(0))
  adoptPlan(a,b,askref(a,b,whenD),
            informref(b,a,whenD),ts(0))

***Updating Discourse Model***

Interpretation: askref(a,b,whenD) (turn number 1)

Suppositions Added:
  expressed(do(a,askref(a,b,whenD)),1)
  expressedNot(knowref(a,whenD),1)
  expressed(intend(a,knowref(a,whenD)),1)
  expressed(intend(a,do(b,informref(b,a,whenD))),1)

Agent a adopted plan to achieve: informref(b,a,whenD)

***Generating Utterance***

<<<surface-request(a,b,informref(b,a,whenD))
```

Speaker b interprets this utterance as a test:

```
| ?- startDialogue2.

      >>>surface-request(a,b,informref(b,a,whenD))

***Interpreting Utterance***

Explaining utter(a,b,request(a,b,informref(b,a,whenD)),ts(0))

Is formula pickForm(a,b,surface-request(a,b,informref(b,a,whenD)),
                    askref(a,b,whenD),ts(0)) ok (y/n)?n.

Is formula pickForm(a,b,request(a,b,surface-informref(b,a,whenD)),
                    testref(a,b,whenD),ts(0)) ok (y/n)?y.

Explanation:
  intentionalAct(a,b,testref(a,b,whenD),ts(0))
  adoptPlan(a,b,testref(a,b,whenD),assertref(b,a,whenD,ts(0))
  credulousB(a,knowref(b,whenD))
  credulousI(a,ts(0))
  pickForm(a,b,surface-request(a,b,assertref(b,a,whenD)),
           testref(a,b,whenD),ts(0))

***Updating Discourse Model***

Interpretation: testref(a,b,whenD) (turn number 1)

Suppositions Added:
  expressed(do(a,testref(a,b,whenD)),1)
  expressed(knowref(a,whenD),1)
  expressed(intend(a,do(b, assertref(b,a,whenD))),1)

Agent a adopted plan to achieve: assertref(b,a,whenD)
```

Speaker b now replies: *Is it at seven-thirty?*

He produces this reply because he has a linguistic expectation that says that "if someone is testing you and you suspect that X is the answer to their test, the appropriate thing to do is to (tentatively) propose X as the answer".[46] (In the explanation below, this expectation is indicated by the default "`expectedReply`").

46 **suspectRef**, like **knowsBetterRef**, is a supposition of a less than certain belief that an agent knows the answer. Except for allowing an agent the (extra) possibility of responding with **askif**, it expresses the same information as **not knowref**. Under our initial assumptions, speaker b might also have answered "I don't know."

```
Explaining shouldTry(b,a,A,ts(1)), decomp(U,A)

Answer:
  shouldTry(b,a,askif(b,a,dinnerAtSevenThirty),ts(1)),
  decomp(surface-request(b,a,informif(a,b,dinnerAtSevenThirty)),
         askif(b,a,dinnerAtSevenThirty))

Explanation:
  intentionalAct(b,a,askif(b,a,dinnerAtSevenThirty),ts(1))
  acceptance(b,askif(b,a,dinnerAtSevenThirty),ts(1))
  expectedReply(do(a,testref(a,b,whenD)),
                suspectRef(b,whenD,dinnerAtSevenThirty),
                do(b,askif(b,a,dinnerAtSevenThirty)),ts(1))

***Updating Discourse Model***

Interpretation: askif(b,a,dinnerAtSevenThirty) (turn number 2)

Suppositions Added:
  expressed(do(b,askif(b,a,dinnerAtSevenThirty)),2)
  expressedNot(knowif(b,dinnerAtSevenThirty),2)
  expressed(intend(b,knowif(b,dinnerAtSevenThirty)),2)
  expressed(intend(b,do(a,informif(a,b,dinnerAtSevenThirty))),2)

Agent b performed expected act: askif(b,a,dinnerAtSevenThirty)

***Generating Utterance***

<<<surface-request(b,a,informif(a,b,dinnerAtSevenThirty))
```

Speaker a recognizes that Speaker b has produced a yes-no question, misunderstanding her request as a test:

```
     >>>surface-request(b,a,informif(a,b,dinnerAtSevenThirty))

***Interpreting Utterance***

Explaining utter(b,a,surface-request(b,a,informif(a,b,dinnerAtSevenThirty)),ts(1))

Is formula pickForm(b,a,surface-request(b,a,informif(a,b,dinnerAtSevenThirty)),
                    askif(b,a,dinnerAtSevenThirty),ts(1)) ok (y/n)?y.

Explanation:
  otherMisunderstanding(b,a,mistake(b,askref(a,b,whenD),
                                    testref(a,b,whenD)),
                        askif(b,a,dinnerAtSevenThirty),ts(1))
  credulousB(b,suspectRef(b,whenD,dinnerAtSevenThirty))
  pickForm(b,a,surface-request(b,a,informif(a,b,dinnerAtSevenThirty)),
           askif(b,a,dinnerAtSevenThirty),ts(1))

***Updating Discourse Model***

Interpretation: askif(b,a,dinnerAtSevenThirty) (turn number 2)

Suppositions Added:
```

```
  expressed(do(b,askif(b,a,dinnerAtSevenThirty)),2)
  expressedNot(knowif(b,dinnerAtSevenThirty),2)
  expressed(intend(b,knowif(b,dinnerAtSevenThirty)),2)
  expressed(intend(b,do(a,informif(a,b,dinnerAtSevenThirty))),2)

Agent b mistook askref(a,b,whenD) for testref(a,b,whenD):
  expressed(mistake(b,askref(a,b,whenD),testref(a,b,whenD)),2)
```

Speaker a then produces a third-turn repair:
No, I'm *asking* you.

```
Explaining shouldTry(a,b,A,ts(2)), decomp(U,A)

Answer:
  shouldTry(a,b,inform(a,b,do(a,askref(a,b,whenD))),ts(2)),
  decomp(surface-inform(a,b,do(a,askref(a,b,whenD))),
         inform(a,b,do(a,askref(a,b,whenD))))

Explanation:
  intentionalAct(a,b,inform(a,b,do(a,askref(a,b,whenD))),ts(2))
  makeThirdTurnRepair(a,b,informref(b,a,whenD),ts(2))

***Updating Discourse Model***

Interpretation:
      inform(a,b,do(a,askref(a,b,whenD))) (turn number 3)

Suppositions Added:
  expressed(do(a,inform(a,b,do(a,askref(a,b,whenD)))),3)
  expressed(do(a,askref(a,b,whenD)),3)
  expressed(intend(a,knowif(b,do(a,askref(a,b,whenD)))),3)

m performed third turn repair

***Generating Utterance***

<<<surface-inform(a,b,do(a,askref(a,b,whenD)))
```

Speaker b recognizes his misunderstanding:

```
      >>>surface-inform(a,b,do(a,askref(a,b,whenD)))

***Interpreting Utterance***

Explaining utter(a,b,inform(a,b,do(a,askref(a,b,whenD))), ts(2))

Is formula
    pickForm(a,b,
          surface-inform(a,b,do(a,askref(a,b,whenD))),
          inform(a,b,do(a,askref(a,b,whenD))), ts(2))
ok (y/n)?y.

Explanation:
  persists(do(a,testref(a,b,whenD)),2)
  selfMisunderstanding(a,b,mistake(b,askref(a,b,whenD),
                                   testref(a,b,whenD)),
                       inform(a,b,
                              do(a,askref(a,b,whenD))), ts(2))
  pickForm(a,b,
           surface-inform(a,b,do(a,askref(a,b,whenD))),
           inform(a,b,do(a,askref(a,b,whenD))),ts(2))

***Updating Discourse Model***

Interpretation:
     inform(a,b,do(a,askref(a,b,whenD))) (turn number 3)

Suppositions Added:
  expressed(do(a,inform(a,b,do(a,askref(a,b,whenD)))),3)
  expressed(do(a,askref(a,b,whenD)),3)
  expressed(intend(a, knowif(b,do(a,askref(a,b,whenD)))), 3)

Agent b mistook askref(a,b,whenD) for testref(a,b,whenD):
  expressed(mistake(b,askref(a,b,whenD),testref(a,b,whenD)),3)
```

Speaker b produces a fourth-turn repair: *Oh, I don't know.*

```
Explaining shouldTry(b,a,A,ts(3)), decomp(U,A)

***Reconstructing Turn Number 1***

  expressed(do(a,askref(a,b,whenD)),alt(1))
  expressedNot(knowref(a,whenD),alt(1))
  expressed(intend(a,knowref(a,whenD)),alt(1))
  expressed(intend(a,do(b,informref(b,a,whenD))),alt(1))

Answer:
 shouldTry(b,a,inform(b,a,not knowref(b,whenD)),ts(3)),
 decomp(surface-inform(b,a,not knowref(b,whenD)),
        inform(b,a,not knowref(b,whenD)))

Explanation:
  intentionalAct(b,a,inform(b,a,not knowref(b,whenD)),ts(3))
  makeFourthTurnRepair(b,a,inform(b,a,not knowref(b,whenD)),
                       ts(3),ts(alt(1)))
  expectedReply(do(a,askref(a,b,whenD)),
                not knowref(b,whenD),
                do(b,inform(b,a,not knowref(b,whenD))),
                ts(alt(1)))
  reconstruction(ts(3),ts(alt(1)))


***Updating Discourse Model***

Interpretation: inform(b,a,not knowref(b,whenD)) (turn number 4)

Suppositions Added:
  expressed(do(b,inform(b,a,not knowref(b,whenD))),4)
  expressedNot(knowref(b,whenD),4)
  expressed(intend(b,knowif(a,not knowref(b,whenD))),4)

r performed fourth turn repair

***Generating Utterance***

<<<surface-inform(b,a,not knowref(b,whenD))
```

Speaker a takes this utterance as an acceptance of her initial request.

```
      >>>surface-inform(b,a,not knowref(b,whenD))

***Interpreting Utterance***

Explaining utter(b,a,inform(b,a,not knowref(b,whenD)),ts(3))

Is formula pickForm(b,a,surface-inform(b,a,not knowref(b,whenD)),
                    inform(b,a,not knowref(b,whenD)),ts(3)) ok (y/n)?y.

Explanation:
  persists(do(a,askref(a,b,whenD)),3)
  persists(do(a,askref(a,b,whenD)),2)
  intentionalAct(b,a,inform(b,a,not knowref(b,whenD)),ts(3))
  acceptance(b,inform(b,a,not knowref(b,whenD)),ts(3))
  expectedReply(do(a,askref(a,b,whenD)),
```

```
                not knowref(b,whenD),
                do(b,inform(b,a,not knowref(b,whenD))),ts(3))
  credulousB(b,not knowref(b,whenD))
  pickForm(b,a,surface-inform(b,a,not knowref(b,whenD)),
           inform(b,a,not knowref(b,whenD)),ts(3))


***Updating Discourse Model***

Interpretation: inform(b,a,not knowref(b,whenD)) (turn number 4)

Suppositions Added:
  expressed(do(b,inform(b,a,not knowref(b,whenD))),4)
  expressedNot(knowref(b,whenD),4)
  expressed(intend(b,knowif(a,not knowref(b,whenD))),4)

Agent b performed expected act: inform(b,a,not knowref(b,whenD))
```