# Towards a Computational Theory of Grounding in Natural Language Conversation

David R. Traum

Technical Report 401
October 1991

UNIVERSITY OF
ROCHESTER
COMPUTER SCIENCE

# Towards A Computational Theory of Grounding in Natural Language Conversation

David R. Traum

The University of Rochester
Computer Science Department
Rochester, New York 14627

Technical Report 401

October 1991



**Abstract**

Theories of speech acts view utterances as *actions* which attempt to change the mental states of agents participating in a conversation. Recent work in computer science has tried to formalize speech acts in terms of the logic of action in AI planning systems. However most of these planning systems make simplifying assumptions about the world which are too strong to capture many features of conversation.

One of these assumptions has been that the intent of an utterance is mutually understood by participants in a conversation, merely in virtue of its having been uttered in their presence. [Clark and Marshall, 1981] have assumptions of *attention, rationality*, and *understandability* to accomplish this. [Perrault, 1990] uses an assumption of *observability.* While these assumptions may be acceptable for processing written discourse without time constraints, they are not able to handle a large class of natural language utterances, including acknowledgements, and repairs. These phenomena have been studied in a descriptive fashion by sociologists and psychologists.

I present ideas leading to a computational processing model of how agents come to reach a state of mutual understanding about intentions behind utterances. This involves a richer, hierarchical notion of speech acts, and models for tracking the state of knowledge in the conversation.

This material is based upon work supported in part by the NSF under research grant no. IRI-9003841, by ONR under research grant no. N00014-90-J-1811, and Air Force - Rome Air Development Center under research contract no. F30602-91-C-0010

# Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

Austin observed that utterances in conversation are *speech acts*, and as such should be treated as part of a theory of action [Austin, 1962]. This observation and the subsequent research program on speech acts within the philosophy of language has been followed up on by researchers in AI, treating speech acts like other actions which an agent can perform, recognize another agent as performing, and reason about. The traditional way of doing this has been to see speech acts as attempts to change the *cognitive state* of another agent, analogous to the way physical actions change the state of the physical world. Speech act operators have been devised using the formalisms from AI planning systems, so that deciding what to say can be seen as *utterance planning*, and interpretation of the intention behind an utterance can be viewed as *plan recognition*.

One difficulty with formalizing speech acts in this way is that all speech acts are collaborative acts: they require both a speaker to make an utterance and a listener to understand and accept it in some way. The result of a speech act is in some way negotiated by the conversational participants.

The collaboration process is made more complicated by the fact that participants cannot infallibly recognize the intent of the other participants: the hearer of an utterance cannot know for sure that he has understood the intent of the speaker, and knowing this fact about the hearer, the speaker cannot be sure that he has been understood.

Since most previous NLP systems have been built as part of a question-answer system based on single database retrieval, where complicated discourse interactions aren't possible, or story understanding, where there is no facility for interaction and the off-line processing can be performed at leisure, they have largely ignored this problem, and have assumed that the intent of the utterance can be recognized merely by being present when the utterance occurs, using only the form of the utterance itself plus background context including the knowledge of the other participant.

In contrast, the study of conversation shows that there is quite a rich system for coordinating understanding. For example, studies of conversations in the TRAINS domain show that about half of the utterances in a conversation are related to keeping the conversation on track rather than being domain level utterances on the topic of the conversation [Allen and Schubert, 1991]. There is an acknowledgement system to make sure that utterances are heard and understood. There is an acceptance system so that requests and information can be agreed to. There are facilities for clarification and repair of potential or actual misunderstandings. These facilities have been the object of extensive study in the field of Conversation Analysis, but have only just started being adopted in computer science systems (see Section 2.6).

A difficulty in describing actions (such as speech acts) in a multi-agent setting is what point of view is being described. Three common points of view are:

1. The objective (what "really" is)

2. The point of view of the performing agent

3. The point of view of the observing agent

Ideally, one would like all three to coincide, i.e. The actor decides what it wants to do, performs an action which accomplishes this intention, and the intention and action are correctly recognized by the observer. Unfortunately, this kind of situation is not guaranteed. The actor may have incorrect beliefs, or may fail in his action, so that what it believes it did is not what it "really" did. The observer also has limited knowledge, and may misinterpret an action. It also has no access to and only limited evidence about the mental state of the actor, and may not recognize what is intended.

In speech acts, what "really" happened is of less importance than that the conversing agents reach some sort of understanding. The question of whether the meaning of an utterance has some objective status aside from what is intended and recognized by the agents is controversial, and not really relevant here. All that is important for communication is that one agent used a particular locution to convey something to another agent, and this intention becomes mutually understood by both agents (*grounded*), regardless of any objective meaning of the utterance.

These distinctions have not generally been made in most speech act work, so it is often difficult to tell the ontological status of many proposed acts: are they objective phenomena, observable and testable by an (ideal?) observer? Are they part of the mental states of the agents, consciously used and necessary for getting at what was intended? Is a set of speech acts to be interpreted as an objective description of the conversation process, or a psychological model of communicating agents (or both)?

## 1.1 Thesis Statement

I propose providing a computational model for how conversants reach a state of mutual understanding of what was intended by the speaker of an utterance. Most previous computational systems have ignored the problem, assuming that this happens more or less automatically as a result of the speaker and hearer being together, and instead have concentrated on the problem of correctly interpreting the meaning of the utterance. Instead I want to say that it is not so important to come up with the "correct" interpretation, but to get an approximately correct interpretation, and then use repairs to fix things up when the interpretation is not close enough for current purposes (the *grounding criterion* of [Clark and Wilkes-Gibbs, 1986]). This approach seems increasingly necessary as researchers are finding many important problems in Natural Language Processing and Planning which are intractable for the optimal case (e.g. [Perrault, 1984; Chapman, 1987; Reiter, 1990])

Several researchers in other fields have come up with similar schemes for presenting a post-hoc analysis of conversation, but have not presented formal models which would show how an agent could do something like this on-line.

## 1.2 Outline of Proposal

Chapter 2 gives an overview of some of the previous research programs which bear on this problem. Section 2.1 relates some of the previous work in formalizing speech acts as planning operators in a computational system. Section 2.2 discusses the problem of representation and acquisition of mutual belief between agents, which is generally taken as the aim of speech acts. Section 2.3 goes over some of the work on formalizing shared intentions and plans. Section 2.4 relates some of the most important insights from the subfield of Sociology known as *Conversation Analysis.* Section 2.5 examines the proposals put forth by Clark and his colleagues for a descriptive model of grounding. Finally, Section 2.6 describes previous attempts to incorporate ideas from Conversation Analysis into Natural Language Processing systems.

Chapter 3 describes work that has already been done towards the aims of the thesis. Section 3.1 describes a preliminary model for on-line reasoning about conversation that was implemented as part of the TRAINS-90 system. Section 3.2 describes a simple extension to this model which allows for acknowledgements and the distinction between private and mutual knowledge of an intention. Section 3.3 describes a classification scheme for speech acts. This classification is meant as both a guide for describing utterances in a conversation and as a resource for planning and plan recognition. Key points are the notion of a *Discourse Unit* which corresponds to a single intention being mutually understood, and *Grounding Acts* which comprise the Discourse Unit and lead to this mutual understanding. Section 3.4 presents a sort of grammar for Discourse Units, showing which combinations of acts are deemed possible, and which combinations form a completed Discourse Unit. Section 3.5 describes a processing model based on the beliefs and intentions of conversing agents for how utterances change the mental state, and how an agent can plan to use utterance acts to accomplish its purposes. Section 3.6 shows how this model can be used to explain the distribution of utterance acts described in Section 3.4.

Chapter 4 describes some natural extensions to the work described in Chapters 2 and 3, a subset of which will be carried out as part of the thesis.

# Chapter 2

# Related Work

## 2.1 Previous NLP Speech Act Work

### 2.1.1 Bruce

Bruce was the first one to try to account for Speech Act theory in terms of AI work on actions and plans [Bruce, 1975]. He defined natural language generation as *social action*, where a *social action* is one which is defined in terms of beliefs, wants, and intentions. He also presented *Social Action Paradigms* which showed how speech acts could be combined to form larger discourse goals. He showed how acts such as **Inform** or **Request** could be used in achieving intentions to change states of belief.

### 2.1.2 Cohen, Allen, and Perrault

Cohen and Perrault [Cohen and Perrault, 1979] tried to define speech acts as plan operators which affect the beliefs of the speaker and hearer. They write that any account of speech acts should answer the following questions:

- Under what circumstances can an observer believe that a speaker has sincerely and successfully performed a particular speech act in producing an utterance for a hearer?
- What changes does the successful performance of a speech act make to the speaker's model of the hearer, and to the hearer's model of the speaker?
- How is the meaning (sense/reference) of an utterance $x$ related to the acts that can be performed in uttering $x$?

They continue that a theory of speech acts based on plans should achieve the following:

- A planning system: a formal language for describing states of the world, a language for describing operators, a set of plan construction inferences, a specification of legal plan structures. Semantics for the formal languages should also be given.

- Definitions of speech acts as operators in the planning system. What are their effects? When are they applicable? How can they be realized in words?

These issues are still central to the work going on in discourse planning.

Cohen and Perrault's models of mental states consist of two types of structures: beliefs and wants. Beliefs are modal operators which take two arguments: an agent who is the believer, and a proposition which is believed. They also follow [Hintikka, 1962], augmenting the belief structure to include quantified propositions. Thus an agent can believe that something has a value without knowing what that value is, or an agent can believe another agent knows whether a proposition is true, without the first agent knowing if it's true or not. Wants are different modal operators which can nest with beliefs. Wants model the goals of agents.

Perrault and Cohen then proceeded to make a first stab at satisfying these issues. The planning system they use is a modified version of STRIPS. They maintain STRIPS's way of dealing with the frame problem, by assuming nothing can change the world except the explicit changes mentioned by the effects of an operator. They describe two different types of preconditions, both of which must hold for the action to succeed. *cando* preconditions indicate propositions which must be true for the operator to be applicable. *want* preconditions are meant to cover sincerity conditions. In order to successfully perform an action, the agent (speaker) must want to do that action. They model the speech acts REQUEST and INFORM, using their planning system.

Allen and Perrault [Allen and Perrault, 1980] use essentially the same formalism as Cohen and Perrault, but for a slightly different purpose. They investigate the role of plan inference and recognition in a cooperative setting. They show how the techniques of recognizing another agent's plans can allow one to recognize an indirect speech act, and provide more information than was requested, in a coherent and relevant manner.

The planning system is again, basically a STRIPS system. There are preconditions and effects, and a body, which is a specification of the operator at a more detailed level.

### 2.1.3 Litman & Allen

Litman and Allen [Litman, 1985; Litman and Allen, 1990] extend Allen and Perraults's work to include dialogues rather than just single utterances, and to have a hierarchy of plans rather than just a single plan [Litman, 1985; Litman and Allen, 1990]. They describe two different types of plans: domain plans and discourse plans. Domain plans are those used to perform a cooperative task, while discourse plans, such as clarification and correction, are task independent plans which are concerned with using the discourse to further the goals of plans higher up in the intentional structure. They also use a notion of ***meta-plan*** to describe plans (including discourse plans) which have other plans as parameters. Using these notions, Litman and Allen are able to account for a larger range of utterances than previous plan-based approaches, including subdialogues to clarify or correct deficiencies in a plan under discussion. There is still no facility for explaining acknowledgement, as the assumption of perfect understanding is maintained.

### 2.1.4 Nonmonotonic Theories of Speech Acts

[Perrault, 1990] takes as a starting point the problem that the utterance itself is insufficient to determine the effects of a speech act. All effects are going to be based in part on the prior mental states of the agents as well as what was actually uttered. However, formalizing the precise conditions which must hold is a tricky endeavor, because of the many possible contingencies. Thus an axiom stating the effects of an utterance in declarative mood must take account of the possibilities of lies, failed lies, and irony as well as standard information-giving acts. Perrault's approach is to state the effects in terms of Default Logic [Reiter, 1980], so that the simple, most common effects can be derived directly, unless there is some defeater. He has a simple axiomatization of belief, intention and action, along with some normal default rules, including a *Belief Transfer* rule which says that if one agent believes that another agent believes something the first agent will come to believe it too, and a *Declarative* rule, which states that if an agent said a declarative utterance, then it believes the propositional content of that utterance. This simple schema allows Perrault to derive expected consequences for the performance of a declarative utterance in different contexts.

Although the formalization is simple and elegant, it still contains a number of serious difficulties. Foremost is the lack of a serious treatment of belief revision. Although intuitively, speech acts are used to *change* beliefs, Perrault's framework can only handle the case of new beliefs being added. As well as not allowing the kind of discourses in which one agent would try to change the beliefs of another, it also has the strange property that one agent can convince itself of anything it has no prior beliefs about merely by making an utterance to that effect in the presence of another agent! It also does not lend itself to a *computational implementation*, since one would need a complete, inductive proof scheme to make all of the necessary deductions.

[Appelt and Konolige, 1988] reformulate Perrault's theory in terms of Hierarchic Autoepistemic Logic [Konolige, 1988]. It has the advantages of implementability and the ability to order the defaults to overcome the problems that Perrault had with normal default logic, but it also loses the simplicity of Perrault's framework. It is also hard to see whether Appelt and Konolige are trying to describe something from the point of view of an ideal observer or from a participant in the conversation. They also resort to unintuitive devices such as the beliefs of an utterance to formulate their theory.

### 2.1.5 Cohen & Levesque

Cohen and Levesque have been attempting to solve a number of problems relating to formal characterizations of Speech Acts, through the use of a logic of action and mental attitudes. [Cohen and Levesque, 1990b] lays out the framework of the basic theory of rational action. It is based on a dynamic modal logic with a possible worlds semantics. They give axiomatizations for modal operators of beliefs and goals, and then derive intentions as persistent goals, those to which an agent is committed to either bring about or realize that they are unachievable.

[Cohen and Levesque, 1990c] attempts to use this logic to show how the effects of illocutionary acts can be derived from general principles of rational cooperative interaction.

They claim, contrary to [Searle and Vanderveken, 1985], that communicative acts are not primitive. They define what it means for an agent to be sincere and helpful, and give characterizations of imperatives and requests. They claim that recognizing the illocutionary force of an utterance is not necessary, that all that is important is that the hearer do what the speaker want, not that he recognize which act the speaker performed as a part of this process. They thus appear to be claiming that illocutionary acts should be seen as descriptive models of action, not as resources for agents. They conclude with a description of how Searle and Vanderveken's conditions on acts can be derived from their rational agent logic.

[Cohen and Levesque, 1990a] extends the framework to handle Performatives. They define all illocutionary acts as attempts. Performatives are acts which have a request component and an assertion component, and the assertion component is made true merely by the *attempt*, not the success of the action. Thus *request* is a performative verb, while *frighten* is not (because it requires a successful attempt and the success is beyond the control of the speaker), and *lie* is paradoxical when uses performatively, because the explicit mention defeats the aim.

[Cohen and Levesque, 1991a] trys to provide an explanation of why confirmations appear in task-oriented dialogue. Using their theory of joint intentions developed in [Levesque *et al.*, 1990] (described below in Section 2.3), they state that the participants in one of these task oriented dialogues have a joint intention that the task be completed. As part of the definition of joint intention, if one party believes the object of intention to be already achieved or to be unachievable, he must strive to make it mutually believed, and this drives the agent to communicate a confirmation. Although this is perhaps the first attempt in the computational literature to explicitly concern itself with the generation of confirmations through plans, it is noticeably lacking in several respects. It has no mention of how the intention to make something mutually believed turns into an intention to perform a confirmation. There is also some distance still from the logic to actual utterances. It is not explained just what would count as a confirmation, and how one might recognize one.

Cohen and Levesque have provided a nice formal logic with which to precisely state and analyze problems of multiagent coordination and communication, but it is difficult to see how it could be used by a resource bounded agent in planning it's actions or recognizing the intentions of others.

### 2.1.6 Other Recent Work

Moore has been working in the area of natural language explanation in expert and advice-giving systems. In her dissertation [Moore, 1989] she presents a system which can respond to user follow-up questions. It maintains the dialogue history as well as the plan used to form the initial explanation, in order to provide useful responses. It can repair a variety of problems in which the user signals lack of comprehension. This represents an improvement over earlier systems by allowing the assumption that the system has been understood by the user to be relaxed when the system is presented by evidence to the contrary. However, it still maintains the assumption in the first place, and does not expect acknowledgements or complain about their absences.

[Turner, 1989] has a conversation system which integrates intention and convention in a natural way. It starts with a case-based memory of conversation plans which represent both the conventions of language and ways of achieving particular conversational goals. The plan is flexible however, and can adapt to changing goals and unanticipated utterances by the user.

[Galliers, 1989] uses the framework of Cohen & Perrault to model cooperative dialogue. She relaxes the typical assumptions of cooperativeness, and shows how conflict and conflict resolution plays an important role in dialogue.

### 2.1.7 Multi-Agent Planning

A speech act theory which can account for conversations must include at least the following extensions to classical planning:

- temporal reasoning, including reasoning about overlapping and simultaneous actions
- uncertainty: attempted actions may fail to achieve their desired results, unexpected results may follow.
- multiple agents, each with individual knowledge, goals, etc.
- cooperation among agents
- real-time resource bounded reasoning
- integration of planning and acting

There is a large amount of research dedicated to addressing these problems, much more than can be summarized here. [Traum and Allen, 1991] explores some of the complexities involved in reasoning and acting in a richer environment. The annual European workshops on Modeling Autonomous Agents in a Multi Agent World (MAAMAW) (reprinted in [Demazeau and Muller, 1990; Demazeau and Muller, 1991]) contain a variety of approaches to these problems.

The next two sections will describe work on capturing the kinds of shared attitudes which seem central to multiagent cooperation.

## 2.2 Mutual Belief

Most of the theories of speech acts as plans reported in Section 2.1 have some of the main effects of speech acts be some new *Mutual Beliefs*. Mutual beliefs are also taken to be some of the prerequisites for felicitous utterance of speech acts. But just what are Mutual beliefs? This section reviews some of the proposals for how to represent the properties of mutual beliefs in terms of simpler beliefs, and how one could acquire new mutual beliefs.

### 2.2.1 Formulations of Mutual Belief

While people agree for the most part about the intuitions underlying the phenomenon of mutual belief, there have been a variety of different ways proposed of modeling it. [Barwise, 1989] compares model theories for three different formulations.

Schiffer uses what Barwise calls "the *iterate* approach" ([Barwise, 1989] p. 202). He defines mutual knowledge between two agents $A$ and $S$ of a proposition $p$, $K^*_{SA}p$ as ([Schiffer, 1972] p. 30):
$K_Sp \wedge K_Ap \wedge K_SK_Ap \wedge K_AK_Sp \wedge K_SK_AK_Sp \wedge K_AK_SK_Ap \wedge \cdots$
It is thus an infinite conjunction of nested beliefs. This approach has since been adopted by many others, including [Allen, 1983] and [Perrault, 1990], who provides an elegant default logic theory of how to obtain each of these beliefs given prior knowledge and a conversational setting. [Grosz and Sidner, 1990] use Perrault's theory for deriving some of the mutual beliefs they take as necessary for forming shared plans.

Barwise credits Harman with the *fixed-point approach*. Harman formulates mutual knowledge as "knowledge of a self-referential fact: A group of people have mutual knowledge of $p$ if each knows *p and we know this*, where *this* refers to the whole fact known" ([Harman, 1977] p. 422). As Barwise point out, the fixed point approach is strictly stronger than the iterate approach, because it includes as well the information that the common knowledge is itself common knowledge. It also replaces an infinite conjunction with a self-referential one.

The final approach discussed by Barwise is the *shared-situation approach*. He credits it to Lewis. Lewis formulates rules for common knowledge as follows ([Lewis, 1969] p. 56):

> Let us say that it is *common knowledge* in a population **P** that **X** if and only if some state of affairs **A** holds such that:
>
> 1. Everyone in **P** has reason to believe that **A** holds.
> 2. **A** indicates to everyone in **P** that everyone in **P** has reason to believe that **A** holds.
> 3. **A** indicates to everyone in **P** that **X**.

This schema is also used by Clark and Marshall, and is apparently the one which Barwise himself endorses.

[Cohen, 1978] uses a *belief spaces* approach to model belief. Each space contains a set of propositions believed by an agent. Nested belief is represented by nested spaces. There is a space for the systems beliefs (SB) which can contain a space for the systems beliefs about the user's beliefs (SBUB) which in turn can contain a space for the systems beliefs about the user's beliefs about the system's beliefs (SBUBSB). If Cohen were to adopt the iterated approach directly, it would require an infinity of belief spaces. Instead, he takes the space one deeper than the deepest which contains any non-mutual beliefs, and points it to its parent space, thus creating a loop, where each even nesting is the same as every other even nesting. Now each of the nested beliefs in the iterated approach can be generated or seen to be present in his belief spaces, by iterating through the loop. This approach shares some features with the fixed-point approach (the self-referentiality) and it allows quick determination of whether mutual belief exists (by searching for a loop) unlike the iterated approach, but it is in fact not as strong as the fixed point approach because the higher-order implications of the fixed-point approach, such as mutual belief about the mutual belief, can not be represented.

A slight modification is to add a separate kind of space, a *mutual belief* space to represent mutual beliefs. This is the approach taken by [Bruce and Newman, 1978]. The Rhetorical knowledge representation system [Allen and Miller, 1989] also uses a Mutual belief space, but disallows nested beliefs within a mutual belief space, giving essentially the power of Cohen's system. This also seems to be the approach used by [Maida, 1984].

### 2.2.2 How can Mutual Belief be Achieved?

If Mutual Belief includes at least the infinite conjunction of nested beliefs, there is a problem as to how to achieve mutual belief, or to recognize when it has been achieved. Several Researches have put forth proposals, yet none seem completely satisfactory.

Perrault uses an extremely strong set of assumptions to drive his default theory [Perrault, 1990]. He has an axiom of observability which states that if an agent is "observing" another agent, then he will recognize all actions (such as declaring a certain proposition) performed by that agent. Agents also have complete memory of prior beliefs, and persist their beliefs into the future (Perrault can not handle belief revision). He also has two default rules, a *belief transfer* rule which states that if one agent believes that a second agent believes something, then the first agent should come to believe it (assuming it doesn't conflict with his prior beliefs), and a *declarative rule* which states that if an agent declares a proposition, then he believes it to be true. With Perrault's set-up, one can derive all the nested beliefs of the iterated approach, assuming there were no prior contradictory beliefs. In the case of some prior inconsistent beliefs, such as in the case of a lie or ironic assertion it also derives the correct set of beliefs. From a computational paradigm, however, it is difficult to see how an agent using Perrault's framework could recognize mutual belief without an infinite amount of computation (or at least some kind of inductive proof procedure for default logic). This would seem to pose a problem for Grosz and Sidner, who would like to use Perrault's system for recognizing mutual belief as the result of a declarative utterance in a task based dialogue ([Grosz and Sidner, 1990] p. 433). Perrault also doesn't mention what might happen in the case where his assumptions are too strong.

Clark and Marshall describe two kinds of heuristics to get at mutual knowledge in a finite amount of time. *Truncation heuristics* look at just a few of the nested beliefs, and then infer mutual belief if all of those check out. *Copresence heuristics* involve the agents recognizing that they and the object of mutual knowledge are jointly present. Clark and Marshall discount the truncation heuristics as implausible, since it is hard for people to reason overtly about nested beliefs. Also, the situation that usually provides evidence for the beliefs checked by the truncation heuristic is usually what would be used directly by the copresence heuristics.

They list four main ways of achieving the copresence necessary for mutual belief, with several subdivisions of these. Their table with the auxiliary assumptions ([Clark and Marshall, 1981] p. 43) is reproduced in Table 2.1:

| Basis for mutual knowledge | Auxiliary assumptions |
|---|---|
| 1. Community membership | Community comembership, universality of knowledge |
| 2. Physical copresence | |
| a. Immediate | Simultaneity, attention, rationality |
| b. Potential | Simultaneity, attention, rationality, locatability |
| c. Prior | Simultaneity, attention, rationality, recallability |
| 3. Linguistic copresence | |
| a. Potential | Simultaneity, attention, rationality, locatability, understandability |
| b. Prior | Simultaneity, attention, rationality, recallability understandability |
| 4. Indirect copresence | |
| a. Physical | Simultaneity, attention, rationality (locatability or recallability), associativity |
| b. Linguistic | Simultaneity, attention, rationality, (locatability or recallability), associativity understandability |

Table 2.1: Methods of Achieving Copresence for Mutual Knowledge

*Community co-membership* is achieved when two agents mutually know that they are part of some community (e.g. people, squash players, computer scientists, etc.). *Universality of knowledge* refers to the assumption that certain things will be mutually known by everyone in a community. These two assumptions together: that two agents **A** and **B** are part of a community and that everyone in this community mutually knows **x**, is sufficient to conclude that **A** and **B** mutually know **x**.

The *simultaneity* assumption is that the agents are simultaneously in the same situation. The *attention* assumption is that the agents are paying attention to the shared situation. The *rationality* assumption is that the agents are rational, and can draw normal inferences. If the situation is a case of physical co-presence, then if it is a case of immediate co-presence these three assumptions are sufficient, (e.g. Ann and Bob are looking at a candle, and looking at each other looking at the candle, so a definite reference of *the candle* is felicitous). If the situation is in the past, then an additional assumption of *recallability* is necessary: it's not enough that the situation occurred, they have to remember it. If the situation hasn't happened, but very easily could, then you need *locatability*. For example, say Ann and Bob are in a room with the candle, but not looking at it; then a reference is felicitous, assuming that the candle is locatable: the reference itself would provide the impetus for achieving the shared situation.

For linguistic copresence (reference to an object mentioned in prior discourse) an additional assumption is required, *understandability*. This is that the utterance which introduces the object can be understood as having done so. The final type of mutual knowledge is a mixture of common knowledge and one of the other two. An example of this is when a candle has been introduced, and then a definite reference to the price, or the wick is made. An additional assumption of *associativity* is needed to be sure that the hearer can make the connection.

There is still a problem with their characterization, in fact, the same problem which motivated them to take up mutual knowledge in the first place. Their conditions for potential copresence are not sufficient. Taking the example they use to show the insufficiency of any finite set of nested beliefs for definite reference, we can see it is also insufficient for potential coreference. Assuming a prior episode of Ann and Bob looking in the morning newspaper and seeing that *A Day at the Races* is playing at the Roxy theater, if Ann, later sees a correction in the evening paper that the movie will be *Monkey Business*, it would not be a felicitous reference to say "the movie at the Roxy" to mean *Monkey Business*. This is true even if Bob has seen the correction, and Ann knows Bob has seen the correction, and she knows he knows she knows he has seen the correction. As long as the sequence is finite, the chain always bottoms out, and we are left with *A Day at the Races* being the more felicitous. But this is precisely the situation with potential coreference. In normal circumstances, Ann can ask if Bob has seen the movie even if she doesn't know if he knows what it is, as long as he can locate what the movie is – perhaps the paper is in front of him. But if we have the prior circumstance of joint knowledge of another referent, we have a problem, no matter how locatable the intended referent is. There are several difficulties in using Clark and Marshall's account: we must not only pick out the unique object in the situation which the definite description refers to, we must also pick out the (unique?) situation in which we can find such an object. This suggests first of all that Clark and Marshall's assumptions for potential copresence are insufficient, and secondly, that perhaps, as Johnson-Laird suggests ([Johnson-Laird, 1982] p.41), their examples do not show that mutual knowledge is necessary. Clark and Carlson ([Clark and Carlson, 1982] p. 56) counter that they are talking about mutual expectation and belief as much as knowledge, and thus Johnson-Laird's proposed counterexamples are not problems for their account. There is still the following potential difficulty: as shown above, the assumptions for potential copresence are not sufficient; therefore, something else is needed. Perhaps this something else can also get them out of the original problem without recourse to mutual knowledge. They at least need to work out the relationships between different basis situations.

Clark and Marshall recognize that reference can fail and can be repaired. They distinguish two types of repair, which they term *horizontal repair* and *vertical repair*. *Horizontal repair* refers to giving more information about the item, but keeping the basis (the type of copresence) the same, where as *vertical repair* is giving a new basis (with presumably fewer assumptions) such as pointing out an item to change physical copresence from potential or prior to immediate.

While the above assumptions may be sufficient for an expectation of mutual belief and felicitous use of a definite referring expression, they are not sufficient to provide actual mutual belief because of the possibility of error (and possible repair). If **A** makes a reference, she can not be sure that it will be understood by **B**. Because of this, even if **B** believes he understands the reference (he still might be mistaken) he can not be sure that **A** believes he does. Each further nested statement introduces more and more uncertainty, and after a while, one of them must be certain to be disbelieved. There is a wealth of linguistic evidence that understandability and attention (or "observability" in Perrault's scheme) are not just mutually assumed. Statements in discourse are often acknowledged by the listener to provide the speaker with evidence that he has been heard and understood. Utterances like "okay", "unhuh", "mmh" are often used to acknowledge the previous utterance. With observability assumed, there would be no need to ever make such utterances.

[Perner and Garnham, 1988] show some additional problems with Clark and Marshall's copresence heuristics. They end up proposing something very much like the shared situation approach from [Lewis, 1969], with the additional restriction that the indications to the population that the situation holds be based on mutual beliefs.

[Halpern and Moses, 1990] present several notions of group knowledge, ranging from implicit group knowledge to full mutual knowledge. They also offer a proof that mutual knowledge is unachievable in an unsynchronized noisy environment, where communication is not guaranteed. They also investigate weaker notions of common knowledge that are achievable.

## 2.3 Shared Plans

A big conceptual difficulty in formalizing cooperative activity is just what is collective intentional behavior and what is it that separates shared plans and intentions from individual intentions? How do shared intentions guide individual actions, and how can individual beliefs and intentions come together to form shared intentions?

Lewis studied several of these problems in [Lewis, 1969]. He defined a *Convention* as a situation in which there is some regularity R in behavior in a population, and everyone conforms to R, everyone expects everyone else to conform to R, and everyone prefers to conform to R, given that everyone else will. A typical example is which side of the road to drive on. In England it is the left side, in America, the right. It doesn't really matter to the drivers which side to drive on, as long as everyone agrees. Coordinated activity is thus seen as individual intention in a state of mutual knowledge about norms. Knowledge of conventions serve to make it in the mutual self interest of each of the members of the population to follow along.

[Grosz and Sidner, 1990] take basically the same viewpoint. They formalize a notion of *SharedPlan* as a set of mutual beliefs about the executability of actions and the intentions of particular agents to perform parts of that action, based on Pollack's definition of a Simple Plan [Pollack, 1990]. They also present some conversational default rules based on cooperativeness to use communication to add to the shared beliefs. Although their framework seems to have many difficulties for implementation, for one thing it is often difficult to figure out exactly what their formalism is really trying to model, some of the extensions [Lochbaum *et al.*, 1990; Balkanski, 1990] may prove to be viable.

[Searle, 1990] starts with the intuition that collective intention is not just a summation of individual intentions. He wants to distinguish between just following a convention and actual cooperative activity. He postulates that *we-intentions* are a primitive form of intentionality, not reducible to individual intentions. There is still a problem of how *we-intentions* can produce the individual intentions necessary for an individual to act.

Cohen and Levesque [Levesque *et al.*, 1990; Cohen and Levesque, 1991b] present their own theory, not in terms of individual intentions (which also aren't primitive in their theory) but in terms of mutual belief and weak mutual goals. Their formulation says that the individuals each have the goals to perform the action until they believe that either it has been accomplished or becomes impossible. Also in the event of it becoming completed or impossible, the agents must strive to make this belief mutual. This framework is also used to explain certain types of communicative behavior such as confirmations as described above in section 2.1.

## 2.4 Previous work in Conversation Analysis

The primary aim of the subfield of sociology known as *Conversation Analysis*[1] (henceforth CA) has been to study actual conversations and inductively discover recurring patterns found in the data. Although the professed aims seem to be to steer away from intuitions or prior formalization, CA has produced a number of useful insights for how Natural Language conversation is organized, and which features of conversation a conversant should orient to. Although the conversation analysts do not formulate it in this way, they examine some of the properties of conversation which show it to be the results of interactions among multiple autonomous agents. The rest of this section is devoted to a brief overview of some of the most relevant findings for designing a computational system to converse in Natural Language.

### 2.4.1 Turn-taking

[Sacks *et al.*, 1974] present several observations about the distribution of speakers over time in a conversation. Although there are frequently periods of overlap in which more than one conversant is speaking, these periods are usually brief (accounting for no more than and often considerably less than 5% of the speech stream [Levinson, 1983] p. 296). Conversation can thus be seen as divided into *turns*, where the conversants alternate at performing the role of speaker. The "floor" can be seen as an economic resource whose control must be divided among the conversants. Although in general conversation (as opposed to more formal communicative settings such as debates, court trials, or classes) there is no predetermined structure for how long a particular turn will last, there are locally organized principles for shifting turns from conversant to conversant. Turns are built out of *Turn Constructional Units*, which correspond to sentential, clausal, phrasal or lexical syntactic constructions ([Sacks *et al.*, 1974] p. 702). Following a Turn constructional unit is a *Transition relevance place*, which is an appropriate moment for a change of turn. Subsequent turns can be

[1]This gloss of some of the findings of Conversation Analysis comes mainly from [Levinson, 1983]

allocated by one of two methods, either the current speaker can select the next one (as in a question directed to a particular individual), or the next speaker can self select, as in an interruption or restarting after an unmarked pause.

One important observation about turn-taking is that it is locally managed. The length and structure of a turn is an emergent property of interaction rather than a predetermined structure. The length of one speaker's turn will be determined by the speaker and other conversants who might end things at different times by taking over. A speaker can direct another to speak, but this does not by itself effect a transfer of the turn, the other must pick it up as well. Keeping the stream of talk to mostly be used by a single speaker at a given time is a coordination problem similar to that of two motorists crossing each other's path (though with less drastic consequences for failure). [Schegloff, 1987] presents an explanation of how conversants re-utter overlapped talk at the beginning of turn-transitions.

### 2.4.2 Adjacency Pairs

Adjacency pairs are pairs of utterances that are ([Levinson, 1983] p. 303):

1. adjacent
2. produced by different speakers
3. ordered into a *first part* and a *second part*
4. typed so that a particular first requires a particular (range of) second(s)

Typical examples of adjacency pairs are question-answer, greeting-greeting, offer-acceptance, and assessment-agreement.

The way that first parts and second parts are connected is not by some sort of grammar rule for legal conversations, but in that the first will make the second *conditionally relevant.* The following utterance by the speaker after the utterer of the first should be either a second, an explanation that the second is not forthcoming, or something preparatory to a second, e.g. a clarification question. Utterances that come between a first and it's second are called *insertion sequences.*

There are two types of seconds that can follow a first. These are known as *preferred* and *dispreferred* responses. Preferred responses are generally direct follow-ups and are unmarked. Dispreferred responses are generally marked with one or more of the following: pauses, prefaces (such as "uhh" or "well"), insertion sequences, apologies, qualifiers (e.g. "I'm not sure but ..."), explanations ([Levinson, 1983] p. 334). Table 2.2 (from [Levinson, 1983] p. 336) shows some common adjacency pairs with preferred and dispreferred seconds:

Adjacency pairs can thus serve as contextual resources for interpreting utterances. If a first part has been made, it makes a second conditionally relevant. The next utterance can be checked to see if it forms a plausible second. Markedness or its absence can be seen as pointing to the preferred or dispreferred second.

| | | | | | |
|---|---|---|---|---|---|
| **First Parts:** | | | | | |
| | Request | Offer/Invite | Assessment | Question | Blame |
| **Second Parts:** | | | | | |
| **Preferred:** | acceptance | acceptance | agreement | expected answer | denial |
| **Dispreferred** | refusal | refusal | disagreement | unexpected answer or non-answer | admission |

Table 2.2: Adjacency Pairs

### 2.4.3 Repairs

*Repairs* can be characterized as attempts to fix previous utterances that are perceived to be (possibly) insufficient for conveying what was intended. Repairs include both *Clarifications*, in which new information is added, and *Corrections*, in which changes are made. Repairs are classified as to who they are made by (self or other), who they are initiated by (self or other), and how many utterances they are removed from the utterance that they are repairing. In the first (same) turn we can have only self-initiated self-repair. In the second turn, we can have other repair or other-initiated self-repair. There is also third-turn repair (when the Initiator subsequently determines, in virtue of the other's previous utterance, that he has been misunderstood), and fourth turn repair, (when the other later realizes that his own interpretation was in error). One can initiate a repair by the other conversant with a *Next Turn Repair Initiator* (or NTRI), which seems to be basically the same as a clarification question. [Schegloff *et al.*, 1977] shows that a preference scheme exists for when to perform a repair. The highest preference is to perform self-initiated self-repair in the same turn. The next most preferred is to perform self-initiated self-repair in the transition space between turns. Then other initiated self-repair in the next turn, via an NTRI. The least preferred is other initiated other repair.

## 2.5 Grounding in Conversation and the Contribution Model

Clark and several of his colleagues have been looking at coordination and collaborative activity in conversation, making explicit reference to both the traditions of Conversation Analysis and Speech Act Theory [Clark and Wilkes-Gibbs, 1986; Clark and Schaefer, 1989; Brennan, 1990; Clark and Brennan, 1990]. They try to identify several principles serving to guide collaborative behavior to account for the kinds of things observed by the Conversation Analysts.

One of the points that they make is that conversants need to bring a certain amount of common ground to a conversation, in order to understand each other. They call the process of adding to this common ground *Grounding*. Grounding can be seen as adding to the mutual beliefs of the conversants (in fact [Clark and Schaefer, 1989] gloss it this way), but it seems reasonable to make a distinction. Though mutual belief, as defined by any of the proposals described in Section 2.2, is probably sufficient for common ground, it may be that only some weaker notion is actually necessary, and that we can have some sort of common ground without full mutual belief. This question is taken up further in Section 4.6.

[Clark and Schaefer, 1989] present a model for representing grounding in conversation via *contributions*. Contributions are composed of two parts: first the contributor specifies the content of his contribution and the partners try to register that content, second the contributor and partners try to reach the *Grounding criterion*, which Clark and Schaefer state as follows, "The contributor and the partners mutually believe that the partners have understood what the contributor meant to a criterion sufficient for the current purpose" ([Clark and Schaefer, 1989] p. 262). Clark and Schaefer divide the contribution into two phases as follows (for two participants, **A** and **B**) ([Clark and Schaefer, 1989] p. 265):

**Presentation Phase:** **A** presents utterance **u** for **B** to consider. He does so on the assumption that, if **B** gives evidence **e** or stronger, he can believe that **B** understands what **A** means by **u**.

**Acceptance Phase:** **B** accepts utterance **u** by giving evidence **e'** that he believes he understands what **A** means by **u**. He does so on the assumption that, once **A** registers evidence **e'**, he will also believe that **B** understands.

Once both phases have been completed, Clark and Schaefer claim that it will be common ground between **A** and **B** that **B** understands what **A** meant. Each element of the contribution may take multiple conversational turns. Rather than a straightforward acceptance, **B** can instead pursue a repair of **A**'s presentation, or ignore it altogether. **B**'s next turn, whether it be an acceptance, or some other kind of utterance, is itself the presentation phase of another contribution. Thus **A** must accept **B**'s acceptance, and so on.

Although the contribution model is perhaps the first explicit model of how grounding takes place, and why acknowledgements occur, it still is lacking in a number of particulars. For one thing, it is often hard to tell whether a particular utterance is part of the presentation phase or the acceptance phase. Self-Initiated Self-repair is considered part of the presentation phase, though other repair seems to be part of the acceptance phase. Either one can have embedded contributions, in the form of insertion sequences or clarification subdialogues, so in the case of an other initiated self-repair, it's hard to tell whether it is part of the presentation phase or the acceptance phase. We often need to look at large segments of the conversation, both before and afterwards before deciding how a particular utterance fits in. The model also seems insufficient to use as a guide for an agent in a conversation deciding what to do next based on what has happened before. Realizing that a presentation has been made but has not yet been accepted can lead one to initiate the acceptance phase, but it's not clear when a presentation or acceptance is complete, or whether the knowledge of being in the presentation phase or acceptance phase has any consequences for what should be uttered.

There are different types of evidence which can be given to show understanding. The main types considered by Clark and Schaefer are shown in Table 2.3, in order from strongest to weakest (from [Clark and Schaefer, 1989] p. 267):

| | | |
|---|---|---|
| 1 | **Display** | B displays verbatim all or part of A's presentation. |
| 2 | **Demonstration** | B demonstrates all or part of what he has understood A to mean. |
| 3 | **Acknowledgement** | B nods or says "uh huh", "yeah" or the like. |
| 4 | **Initiation of relevant next contribution** | B starts in on the next contribution that would be relevant at a level as high as the current one. |
| 5 | **Continued Attention** | B shows that he is continuing to attend and therefore remains satisfied with A's presentation. |

Table 2.3: Types of Evidence of Understanding

The strength of evidence needed for grounding depends on several factors, including the complexity of the presentation, how important recognition is, and how close the interpretation has to be. They try to avoid infinite recursion in accepting acceptances by invoking the following **Strength of Evidence Principle:** The participants expect that, if evidence $e_0$ is needed for accepting presentation $u_0$, and $e_1$ for accepting presentation of $e_0$, then $e_1$ will be weaker than $e_0$.

[Clark and Wilkes-Gibbs, 1986] present a **Principle of Least Collaborative Effort** *which states that "In conversation the participants try to minimize* their collaborative effort – the work that both do from the initiation of each contribution to its mutual acceptance." This principle is contrasted with Grice's maxims of quantity and manner which concern themselves more with the least effort for the speaker. [Clark and Brennan, 1990] show how the Principle of least collaborative effort can help in explicating the preferences for self repair shown by [Schegloff *et al.*, 1977]. They also show how this principle predicts different types of grounding mechanisms for different conversational media, based on the resources available and their costs in those different media.

[Brennan, 1990] provides experimental evidence for how grounding takes place in conversational tasks, and the principles described above. She has a computer based location task, where one party must describe where on a map the other is to point his cursor. The experiment is broken down along two dimensions: familiar vs. unfamiliar maps, to change the grounding criterion, and trials where the director can see where the matcher is vs. trials where the director cannot, and must rely on verbal descriptions from the matcher, to change the strength and type of evidence available for accepting presentations. As might be expected, participants took longer to describe and find locations on the unfamiliar map, and the grounding process was shorter where more direct evidence was available.

## 2.6 Previous attempts to incorporate CA in NLP systems

Although there has been an awareness of the work from Conversation Analysis among some of the AI researchers for some time [Hobbs and Evans, 1979], it is only recently that several researchers have begun attempting to incorporate the findings from Conversational Analysis into computational systems for understanding natural language or human computer interaction.

Suchman [Suchman, 1987] contrasts the classical planning framework, characterized by complete knowledge and forming fully specified plans, with the situated nature of real world action, in which too much is changeable and unknown to plan in complete detail far in advance. In the situated view, "plans are best viewed as a weak resource for what is primarily *ad hoc* activity ([Suchman, 1987] p. ix). She also presents some of the observations and methods of Conversation Analysis, and uses them to analyze the behavior of a computer program to communicate instructions to users of a photocopier, based on attributing a state of certain sensors on the machine to a step in one of the possible plans for making different kinds of copies. She finds that many of the problems the users have in understanding the instructions of the system come about as a result of the system not conforming to typical patterns of conversation usage. The system would mean one thing which would be understood as another by the users. Suchman calls for system designers and researchers in conversation planning to use the rules of conversation as resources to orient on.

This call to design interfaces which are based on the observations of conversation analysis has been taken up by several of the researchers whose work appears in [Luff *et al.*, 1990]. [Frohlich and Luff, 1990] have tried to use the principles of CA in building *The Advice System*, a natural language expert system front end. Although the system uses mouse controlled menu-based input, and fairly authoritarian control over what can be said, it at least pays lip service to the findings of CA, including adjacency pairs, turn constructional units, repairs, including next turn repair initiators (clarification questions), standard openings and closings (including pre-closings), and preferred and dispreferred responses. They have "a declarative definition of the interaction between user and system" ([Frohlich and Luff, 1990] p. 201) composed of elaborate logical grammar rules which specify legal conversations down to low level details. These rules serve both to update the context as the conversation progresses and to help the system choose what to do next.

Although the *Advice System* seems to be a step in the right direction, there are several problems with it in practice. First, it is much too restrictive in its input to be called real conversation. Its notion of utterance types is restricted to **Questions**, **Answers**, and **Statements**. Although the designers consider all possible combinations of any of these by speaker and hearer, they reject far too many as being impossible. Though something can only be an answer if there is an outstanding question, and the existence of an outstanding question will tend to make a next utterance be seen as an answer, there seems to be no reason to outlaw statements immediately following questions uttered by the same agent. This is a very common pattern for repairs (e.g. "Where is the Engine? That's Engine E3."). The *only* point at which a user can interrupt is at a Transition Relevance Place, whereas in real conversation, that is merely the most common and expected place. The menu-based input also trivializes the interpretation problem, and it's unclear why the user should ever have to make a repair. Empirical testing will show if users find the *Advice System* usable or not, but it may well suffer from the same problems that Suchman's copier system suffers from: using familiar patterns in unfamiliar ways, ending up misleading the user.

[Raudaskoski, 1990] describes an attempt to study local repair mechanisms in a telephone message-leaving system. She allowed five different kinds of repair initiators, and ran a simulated experiment where a person acted as intermediary, typing input which was spoken by the user, and reading the responses of the system over the phone. The experiment didn't work very well, mainly because the system could only interpret a very small set of inputs, which the users generally went beyond. Also, the repair mechanisms didn't work very well: the full variety was not used, and those that were used often led to misunderstanding. One of the system's repair initiators seemed too much like a confirmation, so the user thought she had succeeded in leaving the message and went on to the next message but the system was still hoping to get the user to start over.

[Cawsey, 1990] describes the *EDGE* system, which is an expert/advice giver, that combines AI planning with CA local interactions. It has a model of discourse structure based on [Grosz and Sidner, 1986], and plan schemas which it uses to construct explanations. It also allows local interactions, including forcing the user to mouse-click acknowledgement after every utterance, and allowing the user to break in with repair initiators. Planning is done when required (e.g. to fill in the content for a user requested repair) not in advance.

[Cawsey, 1991] uses the endorsement based ATMS model for belief revision presented by a belief revision scheme presented in [Galliers, 1990] to model Third and Fourth turn repair. The belief revision scheme keeps a set of endorsements with each assumption and when conflict occurs, throws out the assumption set with the weakest endorsement. Cawsey uses a Speech act plan recognition system based on [Perrault and Allen, 1980], but makes the interpretations *assumptions*, subject to change if conflict occurs. Thus one can change previous interpretations to bring them in line with new evidence.

# Chapter 3

# Completed Research

The work related in Chapter 2 points out many interesting open avenues for research. The computational speech act/planning approach summarized in Section 2.1 seems to be a very promising way to attack the problem of formal description of language use, in a manner suitable for computational implementation. There are still many open issues, including devising suitable planning models (as noted in Section 2.1.7), and covering an adequate range of language use, such as the phenomena described in Sections 2.4 and 2.5. Sections 2.2 and 2.3 described some of the important effects of speech acts (mutual belief and shared plans, respectively) and some approaches to modelling them.

This chapter presents some first steps towards achieving some of these goals. Section 3.1 presents a preliminary conversation model which has been implemented in the TRAINS system, a system which converses with a user to come up with a shared domain plan and then sends orders to situated agents to implement that plan. This model will serve as a skeleton for adding the advanced coverage, and provides a concrete basis for examining the effects of actions. Section 3.2 provides a simple extension to the model which adds the ability to handle acknowledgements. Section 3.3 presents a hierarchical classification scheme for *Conversation Acts*, generalized actions which are performed using both smaller and larger amounts of language than are associated with traditional speech acts, and can be used to cover some of the phenomena described in Sections 2.4 and 2.5. Sections 3.4 through 3.6 give further ideas on how to recognize and produce these acts in an on-line computational system, presenting an architecture which extends the one in Section 3.1.

## 3.1 TRAINS-90 Model for Discourse

During the Summer of 1990, a preliminary conversation model was designed and built as part of the TRAINS Project. The TRAINS system must cooperatively construct a plan with a human manager to meet some domain goals of the manager. Details of the model can be found in [Traum, 1991], while an overview of the aims of the TRAINS project can be found in [Allen and Schubert, 1991]. The model includes the following components:

### 3.1.1 The Speech Act Analyzer

The **Speech Act Analyzer** takes semantic interpretations of utterances and trys to recognize which acts have been performed by the speaker in making the utterance. The method is based on the one proposed by [Hinkelman, 1990], using linguistic information as a filter to decide which acts are possible interpretations, and using plan based information to choose the most likely among the possible interpretations. One complication is that there is not always a one-to-one correspondence between utterances and speech acts. One utterance may be the realization of several acts (e.g. a follow-up request which implicitly acknowledges the previous utterance and also releases the turn) and some acts may take several utterances before they are completed (e.g. a complex suggestion).

### 3.1.2 The Discourse Context

The **Discourse Context** contains the following kinds of information which must be maintained during the conversation:

- **Turn-taking**: the notion of who has the turn is important in deciding whether to wait for the other agent to speak, or whether to formulate an utterance. It will also shape the type of utterance that will be made, e.g. whether to use some kind of interrupting form or not. The *turn* is represented by a variable which indicates the current holder. The turn is changed by means of **turn-taking acts** which are realized in particular utterances. **turn-taking acts** are described in more detail in Section 3.3.

- **Discourse Segmentation** information is kept for a variety of reasons. Some of these have to do with linguistic interpretation and generation, such as the ability to determine the possible referents for a referring expression. Others have more to do with the relations between utterances, things like adjacency pairs or clarification subdialogues. The currently open segment structure will signal how certain utterances will be interpreted. Utterances like "yes" can be seen as an acceptance of the last question asked but unanswered, if one exists in an open segment. Certain utterances like "by the way", or "anyway", or "let's go back to .." or "let's talk about .." will signal a shift in segments, while other phenomena such as clarifications will signal their changes in structure just by the information content. Arguments for the importance of discourse segmentation structure can be found in [Grosz and Sidner, 1986].

- A record of the system's **Discourse Obligations** is maintained so that the obligations can be discharged appropriately. An accepted offer or a promise will incur an obligation. Also a request or command by the other party will bring an obligation to perform or address the requested action. If these requests are that the system say something (as in a **release-turn** action) or to inform (as in a question), then a discourse obligation is incurred. Rather than going through an elaborate planning procedure starting from the fact that the question being asked means that the speaker wants to know something (e.g. [Allen and Perrault, 1980; Litman and Allen, 1990]), which should then cause the system to adopt a goal to answer, meeting the request is registered directly as an obligation, regardless of the intent of the questioner or the other goal structure of the system. If the system doesn't address the obligation, then it must deal with the usual social problems of obligations which have not been met. This should help distinguish things expected by convention (e.g. that a question be answered) from simple cooperative behavior (e.g. doing what another agent wants). Other parts of the system might also bring about discourse obligations. For example, in some circumstances if the execution of the plan goes wrong, this would bring an obligation to inform the user. [Dipert, 1989] presents some ideas on how different types of obligations can be represented and used in a planning system.

- The system maintains **Discourse goals** in order to use the conversation to satisfy its own goals. The over-riding goal for the TRAINS domain is to work out an executable plan that is shared between the two participants. This leads to other goals such as accepting things that the other agent has suggested, doing domain plan synthesis, or proposing plans to the other agent that the domain planner has constructed. Another top level goal is to fulfill all discourse obligations.

### 3.1.3 Domain Plan Contexts

From the point of view of the Discourse Reasoner, Domain Plans are abstract entities which contain a number of parts. These include: the goals of the plan, the actions which are to be performed in executing the plan, objects used in the plan, and constraints on the execution of the plan. The composition of plans are negotiated by the conversational participants to come up with an agreement on an executable plan, which can then be carried out. Seen this way, the conversational participants can have different ideas about the composition of a particular plan, even though they are both talking about the "same" plan. TRAINS-90 domain plans are described in detail in [Ferguson, 1991].

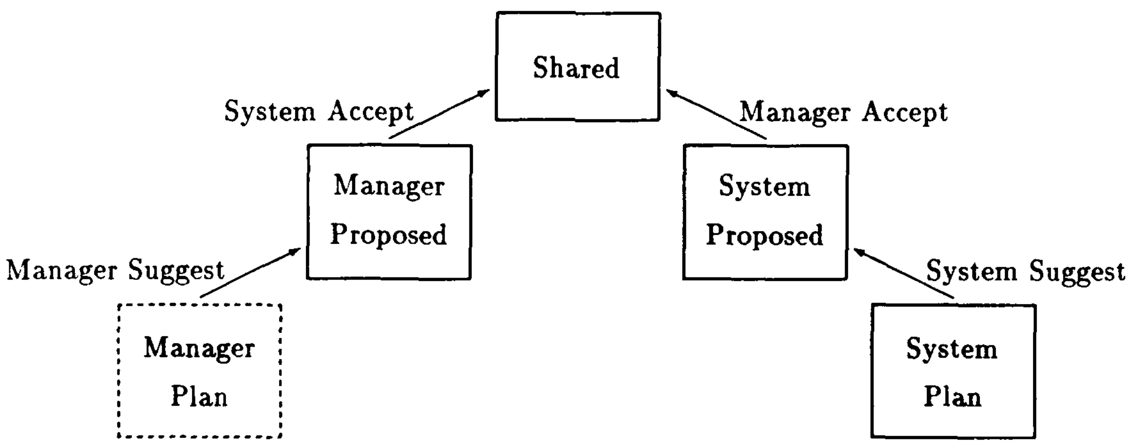


Figure 3.1: TRAINS-90 Domain Plan Contexts

In order to keep track of the negotiation of the composition of a plan during a conversation, a number of plan contexts are used. These are shown in Figure 3.1. The system's private knowledge about a plan is kept in the **System Plan** context. Items which have been suggested by the system but not yet accepted by the manager are in the **System Proposed** context. Similarly, items which have been suggested by the manager but not accepted by the System are in the **Manager Proposed** context. Items which have been proposed by one party and accepted by another are in the **Shared** context. The **Manager Plan** context is shown in dashed lines, because the system has no direct knowledge of and does not represent the private reasoning of the manager. Spaces inherit from the spaces shown above them in the diagram. That is, everything in **Shared** will be in both **System Proposed** and **Manager Proposed**. Also, everything in **System Proposed** will be in **System Plan**.

### 3.1.4 The Discourse Actor

The **Discourse Actor** is the central *agent* of the Discourse Reasoner. It decides what to do next, given the current state of the conversation and plan. It can perform speech acts, by sending directives to the **NL Generator**, make calls to the domain plan reasoner to do plan recognition or plan construction in one of the domain plan contexts, or it can manipulate the state of the plan contexts when appropriate.

### 3.1.5 Capabilities of The Model

The TRAINS-90 discourse model can handle a fairly complex range of task oriented conversations along the lines of the one in Figure 3.2. It can process indirect speech acts, and infer plans which are never explicitly stated. It can carry on a fairly sophisticated negotiation of the content of plans, until an executable plan is shared. It has a rudimentary way of dealing with turn taking, and handles obligations incurred in conversation more straightforwardly than previous systems.

| | | |
|---|---|---|
| MANAGER: | (1) | We have to make OJ. |
| | (2) | There are oranges at I |
| | (3) | and an OJ Factory at B. |
| | (4) | Engine E3 is scheduled to arrive at I at 3PM |
| | (5) | Shall we ship the oranges? |
| SYSTEM: | (6) | Yes, |
| | (7) | shall I start loading the oranges in the empty car at I? |
| MANAGER: | (8) | Yes, and we'll have E3 pick it up. |
| | (9) | OK? |
| SYSTEM: | (10) | OK |

Figure 3.2: Sample TRAINS Conversation

As an example of the model, consider how it handles the conversation in Figure 3.2, with the relevant portion of the trains world shown in Figure 3.3. (1) introduces the current plan and outlines its goal, to make OJ. The rest of this fragment is devoted to working out an implementable plan to fulfill this goal. Utterances 2-4, while they have the surface form of inform acts, are be interpreted in the context of building the plan as suggestions. Thus

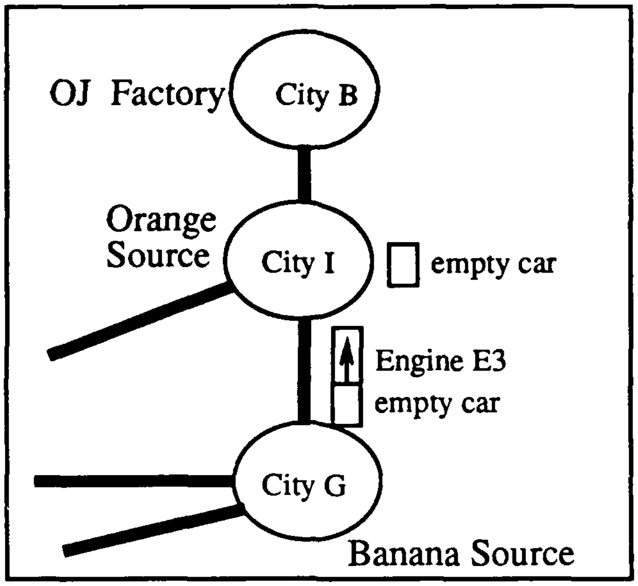


Figure 3.3: Trains World Scenario for Figure 3.2

the manager is not merely informing the system of the locations of various objects in the TRAINS world (the system already knows these facts), but is suggesting that they are somehow relevant to the plan. In performing plan recognition, the system discovers that the manager is suggesting using the OJ Factory at City B to make OJ from the oranges at City I, using Engine E3 to transport them. *This also fills in the missing context for* utterance 5: we want to ship the oranges at I to B using engine E3, as part of out plan to make OJ. Utterance 5 is also seen as a *release-turn* action, in virtue of its question form.

The first thing the system does after receiving the turn is to accept the previous suggestions. While the previous plan recognition and inference had all been going on within the **Manager Proposed** context, this acceptance moves the entire plan, as so far constructed, to the **Shared** context. Now the discourse reasoner calls the domain plan reasoner to do further *plan construction* on this plan to fill in any missing pieces. It comes back with the information (in the **System Plan** context) that in order to transport the oranges, a car is necessary to put them in. There are two likely candidates, as shown in Figure 3.3, one being C1, the empty car already at City I, the other being C2, the car already attached to e3. The system arbitrarily decides to pick C1, and suggests this to the manager in utterance (7), moving the new plan to **System Proposed**. This also releases the turn back to the manager. The manager accepts this suggestion with utterance (8) (moving this part to **Shared**), and also adds the the item that E3 will couple to C1 and take it to B. Utterance (9) requests acceptance, and releases the turn to the system. Everything now seems complete (the unexpressed actions of unloading the oranges and starting the factory having been assumed recognized by plan recognition), so the system accepts the plan (utterance (10)), and sends commands off to the executer to put the plan into effect.

There are still many things that this architecture cannot deal with, one of the most important being acknowledgement and repairs. It is also using an oversimplified model of speech acts. The turn-taking mechanism, is particularly impoverished, treating questions and requests as being always and the only indicators of releasing the turn. Still the framework here, splitting up acts from particular inputs, and basing the functioning of the system on acts, will allow easy integration of a more sophisticated analysis.

## 3.2 Adding Acknowledgements to the TRAINS-90 Model

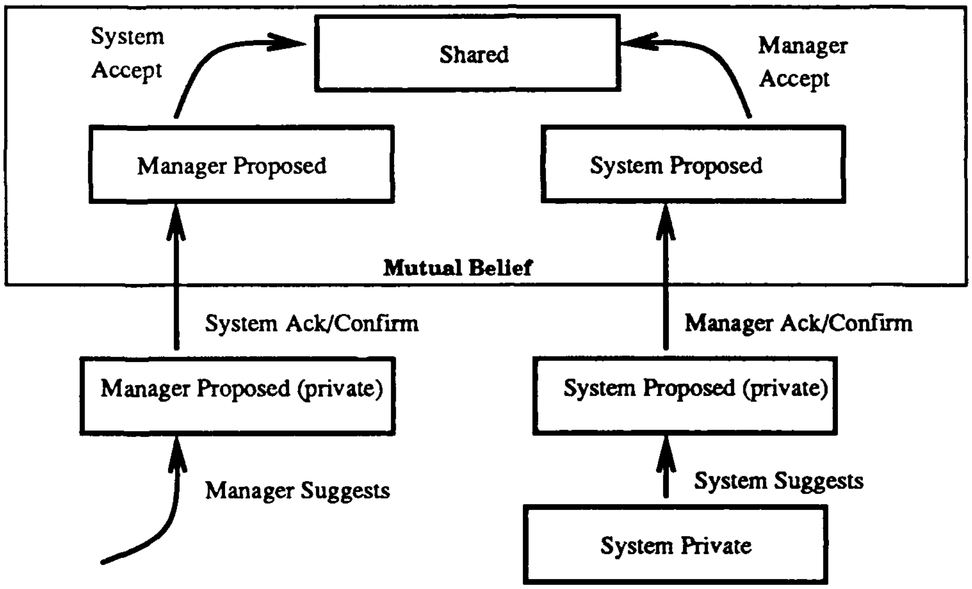


Figure 3.4: Adding Simple Acknowledgements TRAINS-90 Plan Spaces

The TRAINS-90 model maintained the standard assumption that Speech acts were understood as they were uttered. An agent could refuse to accept a proposal, but there was no distinction made between a proposal that was understood but rejected and one which was simply not understood. A simple fix to this is to add two more plan spaces, as shown in Figure 3.4. Here we have a proposed (private) space for when an item is merely proposed but not responded to. When the proposal has been acknowledged by the other agent, we move it to the Mutual Believed proposed space. It still requires an acceptance in order to be part of a shared plan. Now we can negotiate back and forth between agents about which plan to accept, without there being confusion over whether the suggestion is understood. In actual circumstances, one utterance may serve to both acknowledge and accept a prior suggestion, but this will not always be the case.

## 3.3 Categorizing Speech Acts

Since the aim of the TRAINS project is to understand and converse in natural language, conversations between humans have been studied. As part of an experiment to study the role prosody plays in interpreting intention, a series of spoken conversations in the TRAINS domain has been collected [Nakajima and Allen, 1991]. This corpus has been the object of analysis, in order to develop a speech act classification scheme based on intentions of the speaker, which could be used in processing a conversation.

Most prior speech act work has worked with the following assumptions:

1. Utterances are heard and understood correctly by the listener as they are uttered, and it is expected that they will be so understood.

2. Speech acts are single agent plans executed by the speaker. The listener is only passively present.

3. Each utterance encodes a single speech act.

In fact each of these assumptions are too strong to be able to handle many of the types of conversations people actually have:

1. Not only are utterances often misunderstood, conversation is structured in such a way as to take account of this phenomenon. Rather than just assuming that an utterance has been understood as soon as it has been said, this assumption is not made until some positive evidence is given by the listener (an acknowledgement) that he has understood. Some acknowledgements are made with explicit utterances (so called *backchannel responses* such as "okay", "right", "uh huh"), some by continuing with a next relevant response (e.g. a second part of an adjacency pair such as an answer to a question), and some by visual cues, such as head nodding, or continued eye contact. If some sort of evidence is not given, however, the speaker will assume that he has not made himself clear, and either try to repair, or request some kind of acknowledgement (e.g. "did you get that?")

2. Since the traditional speech acts require at least an initial presentation by one agent and an acknowledgement of some form by another agent, they are inherently multi-agent actions. Rather than being formalized in a single agent logic, they must be part of a framework which includes multiple agents.

3. Each utterance can encode parts of several different acts. It can be a presentation part of one act as well as the acknowledgement part of another act. It can also contain turn-taking acts, and be a part of other relationships relating to larger scale discourse structures. It is not surprising that an utterance can encode several acts, since an utterance itself is not an atomic action, but can be broken down into a series of phonetic and intonational articulations.

We have tentatively identified the following different hierarchical levels of conversation acts, summarized in Table 3.1. Action attempts at each of these levels may be signaled by direct surface cues in the discourse.

| Discourse Level | Act Type | Sample Acts |
|---|---|---|
| Sub UU | **Turn-taking** | `release-turn keep-turn assign-turn take-turn` |
| UU | **Grounding** | `Initiate Continue Ack Repair ReqRepair ReqAck` |
| DU | **Core Speech Acts** | `Inform WHQ YNQ Acc Req Den Sug Eval ReqPerm Offer Promise` |
| Multiple DUs | **Argumentation** | `Convince Summarize Find-Plan Elaborate` |

Table 3.1: Conversation Act Types

### 3.3.1 The Core Speech Acts: DU Acts

We would like to keep as much of the previous analysis and work on Speech acts as possible, while still relaxing the overly strong assumptions described above. We maintain most of the traditional speech acts, such as **Inform**, **Request** and **Promise**, calling them *Core Speech Acts.* Instead of the traditional, indefensible assumption that these acts correspond to a single utterance, instead we posit a level of structure which we call a **Discourse Unit** (DU), which is composed of the initial presentation and as many subsequent utterances by each party as are needed to make the act mutually understood. Typically, a DU will contain an initial presentation and an acknowledgement (which may be implicit in the next presentation), but it may also include any repairs that are needed. A discourse unit corresponds more or less to a top level *Contribution*, in the terminology of [Clark and Schaefer, 1989].

### 3.3.2 Argumentation Acts

We may build higher level discourse acts out of combinations of DU acts. We may, for instance, use an **inform** act in order to summarize, clarify, or elaborate prior conversation. A very common Argumentation action is the Q&A pair, used for gaining information. We may use a combination of informs, and questions to convince another agent of something. We may even use a whole series of acts in order to build a plan, such as the top-level goal for the conversations in the TRAINS domain [Allen and Schubert, 1991]. The kinds of actions generally referred to as *Rhetorical Relations* take place at this level, as do many of the actions signalled by cue phrases.

### 3.3.3 Grounding Acts: UU Acts

An *Utterance Unit* (UU) is defined as more or less continuous speech by the same speaker, punctuated by prosodic boundaries. Principles for segmentation into utterance units can be found in [Nakajima and Allen, 1991]. Each utterance corresponds to one *Grounding act* for each DU it is a part of. An Utterance Unit may also contain one or more turn-taking acts. Grounding Acts include

**Initiate(DU-type)** An initial utterance component of a Discourse unit - traditionally this utterance alone has been considered sufficient to accomplish the core speech act.

**Repair** Changes the content of the current DU. This may be either a correction of previously uttered material, or the addition of omitted material which will aid in understanding the speaker's intention. A `repair` can change either the content or Core Speech Act type of the current DU. `repair` actions should not be confused with domain clarifications, e.g. CORRECT-PLAN and other members of the *Clarification Class* of Discourse Plans from [Litman and Allen, 1990]. `repairs` are concerned merely with the grounding of content. Domain clarifications would be argumentation acts.

**Continue** A continuation of a previous act performed by the same speaker. Part of a separate phonetic phrase, but syntactically and conceptually part of the same act. This category also includes `restart-continue`, which is where some part of the previous utterance is repeated before continuing on.

**Acknowledge** Shows understanding of a previous utterance. It may be either a repetition or paraphrase of all or part of the utterance, a *backchannel response* (e.g. "okay", "right"), or implicit signalling of understanding, such as by proceeding with the initiation of a new DU which would logically follow the current one in the lowest level argumentation act. Typical cases of implicit acknowledgement are answers to questions or acceptances of suggestions or requests. Acknowledgements are also referred to by some as *confirmations* (e.g. [Cohen and Levesque, 1991a]) or *acceptances* (e.g. [Clark and Schaefer, 1989]). We prefer the term *acknowledgement* as unambiguously signalling understanding, reserving the term *acceptance* for a DU level action signalling agreement with a proposed domain plan.

**ReqRepair** A request for repair. Asks for a repair by the other party. This is roughly equivalent to a *Next Turn Repair Initiator* [Schegloff *et al.*, 1977]. Often a ReqRepair can be distinguished from a repair or acknowledgement only by intonation.

**ReqAck** Attempt to get the other agent to acknowledge the previous utterance.

### 3.3.4 Turn-taking Acts: Sub UU Acts

We hypothesize a series of low level acts to model the turn taking process. The basic acts are **Keep-turn**, **release-turn** (with a subvariant, **assign-turn**) and **take-turn**. Conversants can attempt these acts by any of several common speech patterns, but it will be a matter of negotiation as to whether the attempt succeeds. Other participants may also use plan recognition on seeing certain kinds of behavior to determine that the other party is attempting to perform a particular act, and may then facilitate it. For example, in utterance 102 in Figure 3.5 the manager is speaking, and hears the system interrupt. The manager can deduce that the system is attempting a **take-turn** action, and stops talking, handing over the turn to the system.

Any instance of starting to talk can be seen as a take-turn attempt. We say that this attempt has succeeded when no one else talks at the same time (and attention is given to the speaker). It may be the case that someone else has the turn when the take turn attempt is made. In this case, if the other party stops speaking, the attempt has been successful. If the new speaker stops shortly after starting, while the other party continues, we say that the take-turn action has failed, and a keep-turn action by the other party has succeeded. If both parties continue to talk, then neither has the turn, and both actions fail.

Similarly, any instance of continuing to talk can be seen as a keep-turn action. Certain sound patterns, such as "uhh", seem to carry no semantic content beyond keeping the turn (e.g. 087, 091).

Pauses generally release the turn. Certain pauses (for example the one between utterances 86 and 87 which begins the dialogue fragment in Figure 3.5) are marked by context as to who has the turn. Even here, an excessive pause can open up the possibility of a take-turn action by another conversant. Other release turn actions can be signaled by intonation. Assign-turn actions are a subclass of release-turn in which a particular other agent is directed to speak next. A common form of this is a question directed at a particular individual.

### 3.3.5 Examples of Conversation Acts

Figure 3.5 presents a small conversation fragment from the TRAINS domain, annotated with examples of conversation acts. The goal of the TRAINS Project [Allen and Schubert, 1991] is to build an intelligent planning assistant that can communicate with a human manager in natural language to cooperatively construct and execute a plan to meet the manager's goal. The domain is transportation and manufacturing, with the execution being carried out by remote agents such as train engineers and factory operators. As a guide to the types of interactions such a system should be able to handle, a corpus of (spoken) task oriented conversations in this domain has been collected with a person playing the part of the system. Figure 3.5 is a small excerpt taken from the TRAINS corpus (experiment 8, utterance units 87-104). This experiment requires the manager to get 100 tankerloads of beer to a particular city within three weeks time. The manager and the system are trying to form a plan to accomplish this. The transcription breaks the discourse into utterance units, numbered consecutively from the beginning of the dialogue. The entire problem takes 451

| Speaker | U# | Utterance | UU Act |
|---|---|---|---|
| | | <long pause> | |
| M | 087 | system, why don't we uhh take uhh engine E-two<br>TT KT KT | Initiate$_1$ |
| M | 088 | and go get tanker T-one<br>KT | continue$_1$(87) |
| M | 089 | and bring it back to city D<br>KT AT | continue$_1$(88) |
| S | 090 | okay<br>TT RT | Ack$_1$ |
| | | <short pause> | |
| M | 091 | and why don't we . use engine E-three .. to uhh<br>TT KT | Initiate$_2$ |
| M | 092 | go to city I to get..get boxcar B-eight, | Continue$_2$(91) |
| M | 093 | go to city B to get tanker T-two<br>KT | Continue$_2$(92) |
| M | 094 | go to city B to get tanker B-seven<br>KT | Continue$_2$(93) |
| S | 095 | sorry, those are boxcars, you mean<br>TT AT | ReqRepair$_2$(94) |
| M | 096 | aaah I'm sorry, yes<br>TT RT | Repair$_2$(94) |
| M | 097 | I wanna get boxcar seven and eight and tanker T-two<br>KT | Initiate$_3$ |
| | | <short pause><br>RT | |
| S | 098 | okay<br>TT | – |
| S | 099 | and tanker T-two at B<br>KT RT | Ack$_3$ |
| M | 100* | yes<br>TT | Ack$_3$ |
| S | 101 | yes<br>TT RT | Ack$_3$ |
| M | 102 | and I would like to . bring<br>TT RT | Initiate$_4$ |
| S | 103* | use E-three for that<br>TT RT | Ack$_2$ |
| M | 104 | yes<br>TT | Ack$_2$ |
| M | 105 | and then I would like to take those to uhhh city F<br>KT KT | Initiate$_5$ |
| | | <short pause><br>RT | |
| S | 106 | okay<br>TT | Ack$_5$ |

Figure 3.5: Dialogue Fragment with Conversation Acts

utterances (about 17 minutes), so this fragment is taken from near the beginning. After querying the system as to the available resources (beer already in warehouses, locations of beer factories, train cars, engines, and raw materials), the manager is now in the middle of formulating a plan to collect some of the train cars together.

The table shows the dialogue as well as some of the conversation acts which are performed. The table can be read as follows: the first column shows the speaker: *M* for manager, or *S* for system. The second column gives the number of the utterance, the third column the transcription of the utterance, and the last column the type of utterance act which is performed, subscripted with the number of the Discourse Unit of which it is a part (numbered in order of initiation from the beginning of this fragment). Utterance numbers appended with an asterisk indicate utterances which overlap temporally with the previous utterance, with the text lined up directly under the point in the previous utterance at which the overlap begins. Turn-taking acts are shown directly under the part of the utterance which signals this attempt. Turn-taking acts are labelled TT, for take-turn, KT for keep-turn, RT for release-turn, and AT for assign-turn. Table 3.2 shows the *core speech acts* which correspond to the DUs numbered in Figure 3.5.

DU#5 exemplifies the fewest possible number of Grounding acts to complete a Discourse Unit, an initiation followed by an acknowledgement. On the other hand, DU#2 shows a moderately complicated one, with several continues, a repair request, and even an embedded inform act which further serves an argumentation relation of clarifying the suggestion. DU#4 is interrupted and never acknowledged, it is as if the suggestion has never been made. This forces the manager to start a new suggestion with DU#5.

The DUs in this fragment are also part of higher level conversation acts, though they are not shown in the table. The whole thing is part of a large action of finding a plan to satisfy the domain goal. At a smaller level, all of these suggestions are part of an action of formulating a plan to put this large train together which will later be used to ferry beer along. On a still smaller scale, DU#3, an inform act, is used to summarize the intentions of the suggestion in DU#2. Topic switching markers, such as the name address "System" in utterance 087, signal the start of a higher level conversation act, in this case consisting of the suggestions shown in Figure 3.5 and rechecking and acceptance which immediately follows the presented fragment.

| DU# | DU Act | Initial U# | Final U# |
|---|---|---|---|
| 1 | Suggestion | 087 | 090 |
| 2 | Suggestion | 091 | 104 |
| 3 | Inform | 097 | 101 |
| 4 | Suggestion | 102 | – |
| 5 | Suggestion | 105 | 106 |

Table 3.2: DU Acts from Dialogue Fragment From Figure 3.5

Figure 3.6 is another fragment, taken from earlier in this same conversation. The manager is busy querying the system about available resources, having just finished finding out about available boxcars (UUs 42-53).

| Speaker | U# | Utterance | UU Act |
|---|---|---|---|
| | | <short pause> | |
| S | 054 | `let's see, so there're`<br>`TT RT` | Initiate$_6$ |
| M | 055* | `where where are my beer factories`<br>`TT TT AT` | Initiate$_7$ |
| S | 056 | `the beer factories are at city D and E`<br>`TT RT` | Ack$_7$ Initiate$_8$ |
| M | 057 | `I see`<br>`TT RT` | Ack$_8$ |

| DU# | DU Act | Initial U# | Final U# |
|---|---|---|---|
| 6 | - | 054 | 054 |
| 7 | WHQ | 055 | 056 |
| 8 | Inform | 056 | 057 |

Figure 3.6: Second Dialogue Fragment

In utterance #55, we can see that the first take-turn attempt is unsuccessful (the System does not stop speaking), though the second one is. Utterance #54 corresponds to an attempt to take the turn and start something new (perhaps a summary of part of the current plan), but it is broken off in the middle. Utterance #56 is both an implicit acknowledgement of the question initiated in utterance #55, and the initiation of an inform DU (which together with the question forms a higher level argumentation action). Utterance #57 completes the inform DU, and also the Q&A argumentation action, after this fragment a new higher level action (a system suggestion) begins.

## 3.4 A "Grammar" for DUs

A completed Discourse Unit is one in which the intent of the Initiator becomes mutually understood (or *grounded*) by the conversants. While there may be some confusion among the parties as to what role a particular utterance plays in a unit, whether a discourse unit has been finished, or just what it would take to finish one, only certain patterns of actions are allowed. For instance, a speaker cannot acknowledge his own immediately prior utterance. He may utter something which is often used to convey an acknowledgement, but this cannot be seen as an acknowledgement in this case. Often it will be seen as a request for acknowledgement by the other party.

We can identify at least six different possible states for a discourse unit to be in. These can be distinguished by their relevant context and what is preferred to follow, as shown in Table 3.3. The superscripts stand for the agent performing that action, *I* for the *Initiator*, the agent starting this DU, and *R* for the *Responder*, the other agent. State S represents a DU that has not been initiated yet, state F represents one that has been grounded, though we can still, for a time, add on more, as in a further acknowledgement or a repair or repair request. The other states represent DUs which still need one or more utterance acts to be

Meanings of States

| State | Relevant Context | Preferred Next |
|---|---|---|
| S | | Initiate$^{I}$ |
| 1 | Initiate$^{I}$ | Ack$^{R}$ |
| 2 | ReqRepair$^{R}$ | Repair$^{I}$ |
| 3 | Repair$^{R}$ | Ack$^{I}$ |
| 4 | ReqRepair$^{I}$ | Repair$^{R}$ |
| F | Done | next DU |

Table 3.3: Preferred Nexts of Discourse Unit States

grounded. State 1 represents the state in which all that is needed is an acknowledgement by the Responder, this is also the state that results immediately after an initiation. However, the Responder may also request a repair, in which case we need a repair by the Initiator before the Responder acknowledges, this is State 2. The Responder may also repair directly (state 3), in which case the Initiator needs to acknowledge this repair. Similarly the Initiator may have problems with the Responder's repair, and may request that the Responder repair further, this would be state 4.

| Next Act | In Transition | | | | | |
|---|---|---|---|---|---|---|
| | S | 1 | 2 | 3 | 4 | F |
| Initiate$^{I}$ | 1 | | | | | |
| Continue$^{I}$ | | 1 | | | 4 | |
| Continue$^{R}$ | | | 2 | 3 | | |
| Repair$^{I}$ | | 1 | 1 | 1 | 4 | 1 |
| Repair$^{R}$ | | 3 | 2 | 3 | 3 | 3 |
| ReqRepair$^{I}$ | | | 4 | 4 | 4 | 4 |
| ReqRepair$^{R}$ | | 2 | 2 | 2 | 2 | 2 |
| Ack$^{I}$ | | | | F | 1* | F |
| Ack$^{R}$ | | F | F* | | | F |
| ReqAck$^{I}$ | | 1 | | | | 1 |
| ReqAck$^{R}$ | | | | 3 | | 3 |

*repair request is ignored

Table 3.4: DU Transition Diagram

Although these states have acts which are in some sense *preferred*, any of a number of acts can follow at any given state. Table 3.4 shows a finite state machine which gives the possible transitions from state to state, and tracks the progress of Discourse Units. This finite state machine has been constructed by analyzing common sequences of utterances in the TRAINS corpus, guided by intuitions about possible continuers and what the current state of knowledge is. It can be seen as doing much the same kind of work as Clark & Schaefer's Contribution model, though it is more explicit, and therefore also more easily falsifiable.

The entries in the table signal which state to go into next given the current state and the utterance act. A Discourse Unit starts with the utterance of an initiator (state S), and is considered completed when it reaches the final state (state F). As can be seen, however, it may continue beyond this point, either because one partner is not sure that it has finished, or if it gets reopened with a further repair. At each state, there are only a limited number of possible next actions by either party. Impossible actions are represented in the table by blanks. If one is in a state and recognizes an impossible action by the other agent, there are two possibilities, the action interpretation is incorrect, or the other agent does not believe that the current DU is in the same state (through either not processing a previous utterance or interpreting its action type differently). Either way, this is a cue that repair is needed and should be initiated. One also always has the option of initiating a new DU, and it may be the case that more than one is open at a time. If a DU is left open (as in an abandoned act) then its contents should not be seen as grounded.

This network serves mainly as guide for interpretation, though it can also be an aid in utterance planning. It can be seen as part of the discourse segmentation structure described in Section 3.1.2. It can be a guide to recognizing which acts are possible or of highest probability, given the context of which state the conversation is currently in. It can also be a guide to production, channeling the possible next acts, and determining what more is needed to see things as grounded. It is still mainly a descriptive model; it says nothing about when a repair should be uttered, only what the state of the conversation is when one is uttered. We can evaluate this model on correctness by checking to see how it would divide up a conversation, and whether it seems to handle acknowledgements correctly. We can also evaluate it as to its utility for processing, whether it serves as a useful guide or not. The type of behavior it describes can be analyzed in terms of the sorts of considerations given in Section 3.5, below, but having an explicit model of this nature may serve to repair interactions, and make processing more efficient.

Figure 3.7 traces how the DUs from Figure 3.5 and Figure 3.6 proceed through this transition network, as the dialogue progresses. All of them begin at the start state (S) and move to state 1 as a result of the initiate act. All of the DUs from the first fragment are initiated by the manager, who has the initiative in this part of the dialogue. Immediately after this fragment, starting with Utterance # 107, the System takes the initiative and begins a series of DUs intended to check on the suggestions made in this fragment. The second fragment shows a position of mixed and transitional initiative.

DU#3 shows how the "final" state (F) is not necessarily final. After Utterance # 099, the DU seems complete, but we can still have further acknowledgements which do not change the state, though they probably make the participants more certain that this is indeed where they are. A more fine grained model would need a graded model of belief, so that we could talk about increasing the confidence that a DU is grounded. Such a model is needed to handle interactions with differences in the Grounding Criterion [Clark and Schaefer, 1989], but is beyond the scope of the current project. Utterance #56 is an example of an utterance which plays a grounding role in more than one DU. It plays an initiate role for DU#8, starting it off in state S, while moving DU #7 to state F, leaving it grounded.

**DU # 1**

| UU # | act | new state |
|---|---|---|
| 087 | Initiate$^I$ | 1 |
| 088 | Continue$^I$ | 1 |
| 089 | Continue$^I$ | 1 |
| 090 | Ack$^R$ | F |

**DU # 2**

| UU # | act | new state |
|---|---|---|
| 091 | Initiate$^I$ | 1 |
| 092 | Continue$^I$ | 1 |
| 093 | Continue$^I$ | 1 |
| 094 | Continue$^I$ | 1 |
| 095 | ReqRepair$^R$ | 2 |
| 096 | Repair$^I$ | 1 |
| 103 | Ack$^R$ | F |
| 104 | Ack$^I$ | F |

**DU # 3**

| UU # | act | new state |
|---|---|---|
| 097 | Initiate$^I$ | 1 |
| 099 | Ack$^R$ | F |
| 100 | Ack$^I$ | F |
| 101 | Ack$^R$ | F |

**DU # 4**

| UU # | act | new state |
|---|---|---|
| 102 | Initiate$^I$ | 1 |

**DU # 5**

| UU # | act | new state |
|---|---|---|
| 105 | Initiate$^I$ | 1 |
| 106 | Ack$^R$ | F |

**DU # 6**

| UU # | act | new state |
|---|---|---|
| 54 | Initiate$^I$ | 1 |

**DU # 7**

| UU # | act | new state |
|---|---|---|
| 55 | Initiate$^I$ | 1 |
| 56 | Ack$^R$ | F |

**DU # 8**

| UU # | act | new state |
|---|---|---|
| 56 | Initiate$^I$ | 1 |
| 57 | Ack$^R$ | F |

Figure 3.7: Traces of Transitions of DUs from Dialogue Fragments

DUs #2 and #4 make an interesting study. While DU #3 is related to the understanding of DU #2 (and might even be seen as a continued repair of it, and not a new DU at all), DU #4 seems to be starting something new, another suggestion at the same levels as DU #1 and #2. This behavior seems to indicate that the manager thought that DU #2 had already been closed, perhaps by utterance # 099 or # 101. The system did not seem to agree (or at least decided that the situation wasn't clear), and interrupted with an acknowledgement. After acknowledging this, DU #4 has been left open, so the manager starts a new DU with utterance #105.

## 3.5 A Model for Processing Grounding Acts

Figure 3.8 gives a schematic of the types of information and processes necessary for grounding using the actions described in Section 3.3.3. This figure shows the process from the point of view of an agent named X. The other agent that X is communicating with is named Y. The boxes can be seen as distinct knowledge bases, which are related by differ-

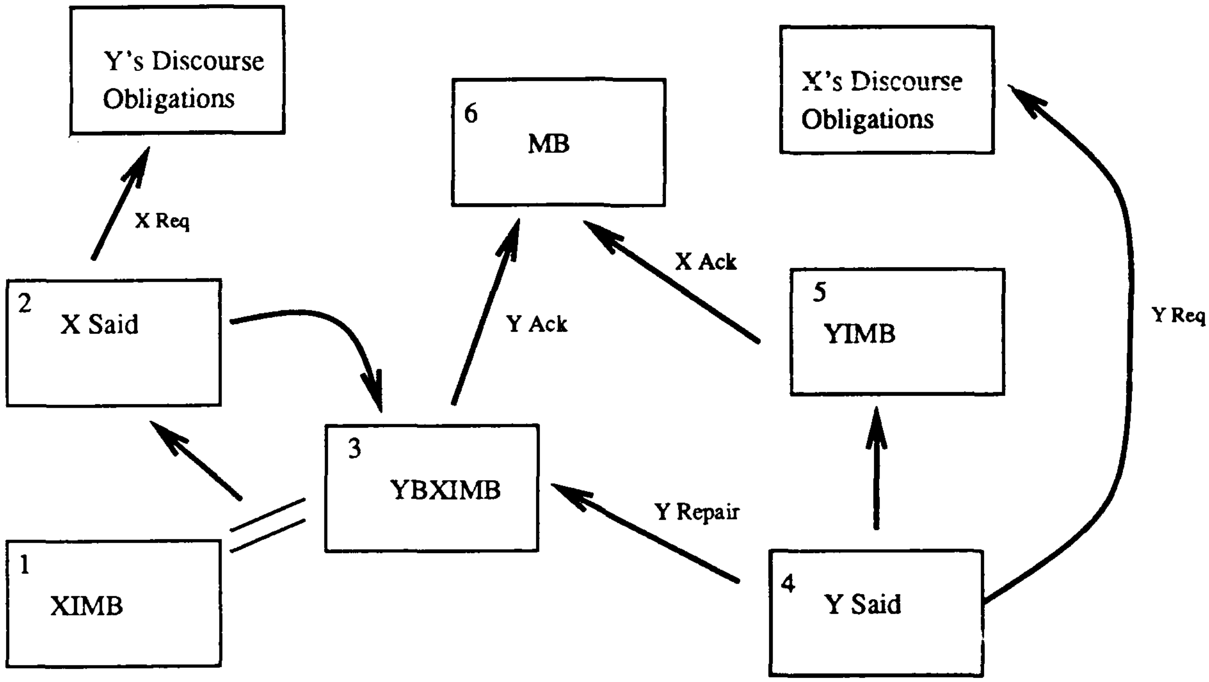


Figure 3.8: Architecture for X's Model of Conversation with Y

ent inference processes. Particular actions can move information from one box to another. Box 1 *represents items that* X *intends* be mutually believed. Box 3 represents (X's beliefs about) Y's beliefs about what X intends be mutually believed. Box 2 represents (X's interpretations of) utterances that X makes in order to change Y's beliefs, and thereby bring about mutual beliefs (Box 6). Box 4 represents (X's interpretations of) Y's utterances, and Box 5, (X's beliefs about) Y's intentions. In addition, two other boxes are shown, which represent (X's views of) the current ***discourse obligations*** of the two agents. Discourse obligations (described above in Section 3.1.2) result from the normative expectations of minimal cooperativeness from agents engaged in a conversation.

The grounding process is started when one party or the other makes an utterance which initiates a discourse unit. X will decide to initiate a DU if there is something in Box 1 which is not elsewhere in this diagram, and the proper contextual factors apply (X has the turn and there are no outstanding discourse obligations or goals to do something else). X will invoke the Utterance Planner to come up with an utterance that will convey X's intention to Y. When X actually performs the utterance, we have in box 2, the interpretation of that utterance. This will most likely be the same as was intended (if the utterance planner is good), but may not be, due to resource constraints on the utterance planning process. If somehow the interpretation of the utterance is different from what was intended, that will provide the basis for planning some sort of repair which can repair the previous utterance.

If the interpretation is a "content" act, such as **Initiate**, **Continue**, or **Repair**, then the next step is to do plan recognition and inference based on Y's beliefs, to see what Y will likely believe about X's intentions. The result of this plan recognition, including the act interpretation, and its implicatures, will be placed in Box 3. Now if Box 3 contains the same items as Box 1, X believes that his communication was adequate, and must wait for (or prompt with a `ReqAck`) an acknowledgement from Y that Y correctly understood. If, on the other hand, the contents of Box 3 are not the same as those of Box 1, that is further impetus for the utterance planner to remedy this with a further utterance. The subsequent utterance may come out as a repair, a continue, or even a new initiate, depending on the particular differences. These subsequent utterances would also be subject to the same processes of interpretation and inference leading back to Box 3.

When Y makes an Utterance, its interpretation goes into Box 4. If there is some problem with the interpretation, such as no plausible interpretation, or no evidence to choose between two or more possible interpretations, this will provide the basis for the utterance planner to come up with some sort of repair initiator, most probably a repair request, but perhaps (if contextual factors indicate what Y *should have* said) a direct repair.

Once X thinks that it understands Y's utterance, what happens next depends on the actions that X thinks Y has performed. If Y has performed an `acknowledgement`, then the items acknowledged move from Box 3 to Box 6 (MB). If the utterance is an `Initiate`, `Continue`, or `Repair`, then X will do plan recognition and inference to deduce Y's intentions and put the results in Box 5. X can make the contents of Box 5 grounded by uttering an acknowledgement, moving the contents on to Box 6. If Y's utterance is either a request for acknowledgement or a request for repair, this will give evidence for more inference to be performed on the contents of Boxes 3 and 5, as well as adding **Discourse Obligations** for X to perform or respond to the requested action.

| Action | Reason | Effects |
|---|---|---|
| `Initiate` | Item in (1), not elsewhere | Move item to (3) |
| `Continue` | Item in (1), part but not all in (3) | Move item to (3) |
| `Ack` | Item in (5), not in (6) | Move item from (5) to (6) |
| `Repair` | Either item in (2) or (3) doesn't match item in (1) or item in (4) is unclear (either no interpretation, no preferred interpretation, or interpretation doesn't match expectations) but there is enough context to say what it *should* be | Move item to (3) |
| `ReqRepair` | Item in (4) is unclear (either no interpretation, no preferred interpretation, or interpretation doesn't match expectations) | Add discourse obligation for Y to respond to this request |
| `ReqAck` | Item in (3) matches item in (1), Y has passed up a chance to acknowledge | Add discourse obligation for Y to respond to this request |

Table 3.5: X's actions

Table 3.5 summarizes the reasons for X to do each type of action and its effects. For each of X's actions, after coming up with the intention to perform the action, X will first plan the utterance and then perform it, then interpret the actual utterance and place the results in Box 2. Further effects depend on the type of action, and are described in the third column. Table 3.6 shows the effects of Y's actions. For all of these actions, the interpretation of the utterance will start in Box 4.

| Action | Effects |
|---|---|
| `Initiate` | Put item in (5) |
| `Continue` | Put item in (5) |
| `Acknowledge` | Move item from (3) to (6) |
| `Repair` | Move/Change item to/in (5) |
| `ReqRepair` | Change (3); Add Discourse Obligation to respond to this request |
| `ReqAck` | Add Discourse Obligation to respond to this request; **`ReqRepair`** if unsure what Y wants acknowledged |

Table 3.6: Y's actions

## 3.6 Explaining Grounding Act Distribution with the Processing Architecture

Table 3.7 shows constraints on performance by X or Y of the Grounding Acts. The Acts are appended with *I* or *R*, depending on whether the speaker is acting as the initiator or responder, as in Table 3.4. These constraints are all relative to the knowledge of X, as represented in Figure 3.8.

Using this information, we can now try to account for the constraints on the distribution of grounding acts in a discourse unit shown in Table 3.4. In state S, there is nothing of the current Discourse Unit in any of the Boxes (other than perhaps Box 1), so according to Table 3.7, the only act possible is an `Initiate`. Similarly, an Initiate act is not possible in any other state, because this DU has already been initiated (though, of course, either party may begin a new DU with a subsequent Initiate act).

State 1 corresponds to there being something in Box 3, if X is the initiator, or Box 5, if Y is the initiator. From State 1, $\texttt{Ack}^{I}$ is disallowed, because there is nothing in the appropriate box (Box 5 if X is initiator, Box 3 if Y is the initiator) for the act to acknowledge. Similarly, there is nothing for the initiator to request repair of. Continuations and Repairs by the initiator will just add more to Box 3 if X is initiator or Box 5 if Y is the initiator. An acknowledgement by the responder will move items into Box 6.

State 2 corresponds to a point after a repair request by the responder. If X is the initiator, then there is something in Box 3, and the repair request by Y in Box 4. Also, X has a discourse obligation to make a repair. A continuation is precluded because, given the

| Actions | Conditions for X | Conditions for Y |
|---|---|---|
| `Initiate`$^I$ | Item in (1), not elsewhere | none |
| `Continue`$^I$ | Part of item in (3), part in (1) | Item in (5) |
| `Repair`$^I$ | Item in (2) or (3) not equal to item in (1) | Item in (4) or (5) |
| `Repair`$^R$ | Item in (4) or (5) but not what it should be | Item in (3) |
| `ReqRepair`$^I$ | Item in (4) which is unclear | Item in (2) or (3) |
| `ReqRepair`$^R$ | Item in (4) which is unclear, or item in (5) which doesn't seem right | Item in (2) or (3) |
| `Ack`$^I$ `Ack`$^R$ | Item in (5) | Item in (3) |
| `ReqAck`$^I$ `ReqAck`$^R$ | Item in (3) | Item in (4) or (5) |

Table 3.7: Constraints on Grounding Acts

obligation, it would be seen as somehow addressing the request, and therefore a repair. If, somehow, the initiator's next utterance were seen as a continuation, it would be a signal that the initiator did not process the previous repair request. As in State 1, the initiator can not acknowledge, because there is nothing in the proper box. The expected operation from State 2 is that the initiator will perform the requested repair, but there are a few other possibilities. The initiator may not be able to interpret the request, and may request a repair of the **ReqRepair**, shifting the discourse obligation over to the responder, and putting us in State 4. The responder may realize that his request might not be interpreted correctly, and may repair it, remaining in State 2. The responder may also make a different repair request, also remaining in State 2. The final possibility, is that on further reflection, the responder realizes the answer without the initiator having to repair. In this case the responder may acknowledge the original contribution, just as though the request had never been made. This takes us directly to State F, and removes the obligation to repair.

State 3 is reached when the responder has directly repaired the initiator's utterance. Here the responder is shifting from what the initiator intended for the DU to what the responder intends[1]. In making the repair, an item has been placed in Box 3, when X is

[1]Or more precisely what it thinks that the initiator *should* intend. This distinction is not currently made

responder, or Box 5, when Y is the responder. In some ways, the responder can be seen as shifting the roles and seizing the initiative. This state is thus in some ways a mirror of State 1. The initiator can repair in return, seizing back the initiative and moving back to State 1. Also the responder can make a follow up repair, adding more items to the appropriate box, but remaining in State 3. The initiator might not understand the repair, and may make a repair request, moving to State 4. The responder may also have a problem with something else that the initiator said, and may "release the initiative" with a repair request, moving back to State 2. The initiator may also acknowledge this repair, moving the items to Box 6. The responder may no longer acknowledge its own repair, though it may request an acknowledgement, or even rescind the repair (e.g. "oh, sorry, you're right.").

State 4 is perhaps the most complicated state. It corresponds to the responder having a discourse obligation to repair. Tracing back the conditions, this can only happen after an original `Initiate` by the initiator, some response by the responder, and then a repair request by the initiator. Thus there is something in each of Boxes 2-5. Also the responder has an obligation to make a repair. From this state, the initiator may make a further repair request, or repair his previous request, remaining in State 4, the responder may repair, moving on to State 3, or the responder may request repair of the repair request, moving back to State 2.

State F occurs when items have moved on to Box 6. Ideally things are now grounded, and no further action is necessary. It is still, however, possible to reopen the discourse unit, as shown in the last column of table 3.4. A repair will put the new item in Box 3 if performed by X, and in Box 5 if performed by Y. A `ReqRepair` or `ReqAck` will produce the *appropriate discourse obligation*. In addition, a follow-up acknowledgement will keep the DU in State F.

in this system.

# Chapter 4

# Possible Directions for Future Research

## 4.1 Refining The Speech Act Classification

The models presented in Chapter 3 are still rather preliminary. Although the speech act classification presented in Section 3.3 was based on examination of a corpus of conversation, one thing that needs to be done is to go back and see how well this classification really covers the corpus. An attempt should be made to answer the following questions: what percentage of acts can be classified reliably? How many utterances don't seem to fit the scheme? Does classifying the acts lead to acceptable assumptions about the intentions of the participants? While it is hard to quantify the acceptability of a classification scheme since there is no direct access to the mental states of the participants, examinations can still lead to some basis for comparison with other classification proposals. Some constraints on an acceptable classification scheme are:

- Does it cover an acceptably large subset of the utterances to a sufficient degree?
- Are act types which seem to human analysts to be close to each other (e.g. hard for people to distinguish which type a particular utterance is) shown to be close in the classification? There should be some sort of hierarchical structure so that ambiguities can be concisely represented and reasoned about.
- Is it possible to use the classification of acts towards recognizing the intentions of the speaker, and determining what to do next?
- What are tests which can distinguish one act from another? These tests should be based only on information available to an observer of the act. It should include syntactic, prosodic, and contextual information, but not be based on some private knowledge of the speaker's mental state which is not deducible from prior context.

One particular change we are considering is adding a **cancel** Grounding Act, to handle utterances which close off and abandon the current DU, leaving it ungrounded. This would be used to handle sequences such as the following:

```
okay, now I have a good i
oh no, we can't use the same thing
```

Here the second utterance cancels whatever was started in the first.

## 4.2 Descriptions of the Intentions underlying Speech Acts

Once we have a classification scheme, we can take each act and deduce what effects such an act will make in the world (i.e. how it will affect the *Cognitive State* of the conversants), conditions for its use and likely inferences which can be drawn. Although these factors will play a large role in formulating an advantageous classification, one needs to look at the whole classification scheme to deduce the precise effects. A participant in a conversation has a choice of acts to perform (including just remaining silent) and inferences about the effects of one act can be made based on which other acts have not been done. For instance, following up an utterance by another speaker with a next relevant contribution can be seen as an acknowledgement because it is not a repair, but does give evidence that the the first speaker has been heard and understood [Clark and Schaefer, 1989].

## 4.3 Implementation

The model in Section 3.5 describes an architecture for how a system might hold a conversation. What still remains is to see how viable this architecture is in practice. It is fine to say that a responder has the options of acknowledging, repairing, or ignoring a previous utterance, but we cannot reliably say which will happen at any given time. There are too many variables at work in trying to predict whether an agent will understand a particular utterance. What we need to do is to see if two agents can manage to understand each other in practice. The TRAINS domain provides an ideal platform for testing these models. Once the whole system is in place, we can measure understanding by action. What will matter is not so much if the system understands every utterance the same way that the manager does, but whether it can do what the manager wants in the context of TRAINS world actions. If the manager cannot get his ideas across, or cannot understand what the system is saying, then there is a problem. The manager can determine whether he is getting his ideas across by whether the system does what he wants, and he can determine whether he understands the system correctly by seeing if the system's actions meet his expectations.

## 4.4 Modifications to the Grounding Model

As the models become more formalized and implemented, several changes may appear to be more fruitful. Some other possibilities which may be useful to try to implement and compare are:

- recursive repairs (either Clark & Schaefer's Contributions or some other model): In actual conversation, we can repair a repair. The model in Section 3.4 kept all repairs at the top level, which is clearly an oversimplification of sorts. A problem comes up when a repair is acknowledged, but the whole utterance is still unacknowledged. Moving to a recursive repair model will mean that an acknowledgement will be ambiguous as to how many levels it is acknowledging. Similarly, a second repair becomes ambiguous as to whether it is a repair of the repair or a second repair to the main utterance. Such a model, while it will have higher coverage probably vastly overgenerates the kind of behavior people actually exhibit, and may make interpretation harder by introducing a kind of spurious ambiguity. Some sort of happy medium between the two extremes would be nice, perhaps a two-level model might be sufficient.

- We can also look at allowing *conditional acknowledgements*. These would be the interpretation of certain types of confirmations, which might acknowledge understanding, given certain things being the case (e.g. the confirmation is accepted). This would be analogous to a kind of "tail-recursion", where a certain response might get us out of several levels. It might be that utterance # 095 from Figure 3.5 appears to be something like this to the manager.

- Another idea is to look more closely at the "core speech act" of each utterance act, and see the plans behind speech acts as *meta-plans*, which have other plans as arguments, along the lines of [Litman and Allen, 1990].

- We may also want to think about hierarchical notions of acknowledgement, where an utterance can selectively acknowledge part of the current context but not others. This would not be restricted to an utterance by utterance grounding, as in the Contribution model, but would be based more on content (e.g. we could recognize that we are requested to move an Engine somewhere, but we might not have recognized which engine or where it is to move). This approach may come out more or less naturally from an implementation of the architecture described in Section 3.5.

## 4.5 Formal Account of Speech Planning

Once we have the basic mechanisms, in the form of the model and working implementation, we can formalize what is going on in order to catch some of the subtleties and reason about the possibilities for extension. For instance, the propositional attitudes of intention and belief have here been described (and the implementation will most likely be built) using belief spaces. But this could easily be formally translated into a modal logic, which might have better descriptive capabilities.

## 4.6 Mutual belief

Along the same lines, once we have a working system, we can examine just what is really happening with respect to mutual belief or some other sort of mutuality. We can look at just what sort of mutuality appears to be necessary, and how it is that our system acquires (or assumes) it. We can give a formal account of what the system is doing, and see how it relates to the accounts of mutual belief described in Section 2.2.

[Sperber and Wilson, 1986] suggest that mutual knowledge is not a necessary precondition for communication, contrary to Clark & Marshall. Instead they use the term *mutual manifestness*, saying that something must be mutually available for a reference to it to be made. They reject Mutual Knowledge as lacking psychological plausibility ([Sperber and Wilson, 1986] p. 31). We can examine the question of whether *Grounding* really is the same as mutual belief acquisition, or whether something else seems to be going on.

## 4.7 Degrees of Belief

The model of belief assumed here is a straight forward all or nothing model, something is either believed or it isn't. This kind of model is easily represented using belief spaces, or modal logic, but is insufficent for some purposes. Certain phenomena involved with grounding (e.g. [Clark and Schaefer, 1989]'s *Grounding Criterion* and *Strength of Evidence Principle*) seem to require a graded model of belief, where different strengths of belief of understanding are needed for different purposes. If we adopted such a scheme, we might be able to separate out different types of acknowledgement and explain multiple acknowledgements.

## 4.8 Speech acts as part of general account of multi-agent interaction

While speech is its own modality, with several features which are distinctive from other types of action, there is still a significant overlap with other kinds of action. For instance, a request can be to perform another speech act, or to perform some physical action. Speech acts are often made to help satisfy domain goals in similar ways to the way physical actions are made. In order to capture some of the regularities of conversation planning, it is necessary to say how it fits in to a larger account of planning in a multi agent domain. Cohen & Levesque give the beginnings of such an account, but not in a form which is useful for an agent involved in planning and acting in the world.

Discovering how speech acts fit into a more general theory of multi-agent action will certainly help in the overall deliberation process of an agent as to "what to do next", and when to talk or do other things. It may also help with the speech act classification enterprise, by presenting regularities which would otherwise be missed.

## Acknowledgements

The Classification scheme for conversation acts in Section 3.3 was devised in Collaboration with my advisor, James Allen, and Shin'ya Nakajima. I am also greatful to Shin'ya and Derek Gross for collecting and transcribing the conversations which have been used as data for this analysis. The TRAINS discourse model in Sections 3.1 and 3.2 was also devised in collaboration with James. I would also like to thank the other members of the TRAINS System project, Len Schubert, George Ferguson, Janet Hitzeman, Chung Hee Hwang, Jonas Karlsson, Alice Kyburg, Marc Light, Nat Martin, Brad Miller, Massimo Poesio, and Josh Tenenberg, for providing a context within which to apply these ideas, as well as good early feedback on how well they seemed to work. I would also like to thank James, Len, Chung Hee, Alice, Massimo, Mike Tanenhaus, Graeme Hirst, Susan Brennan, Elizabeth Hinkelman, and Peter Heeman for making many helpful comments on earlier versions of this work.

# Bibliography


[Allen, 1983] James Allen, "Recognizing Intentions From Natural Language Utterances," In Michael Brady and Robert C. Berwick, editors, *Computational Models of Discourse*. MIT Press, 1983.

[Allen *et al.*, 1990] James Allen, James Hendler, and Austin Tate, editors, *Readings In Planning*, Morgan Kaufmann, 1990.

[Allen and Perrault, 1980] James Allen and C. Perrault, "Analyzing Intention in Utterances," *Artificial Intelligence*, 15(3):143–178, 1980.

[Allen and Miller, 1989] James F. Allen and Bradford W. Miller, "The Rhetorical Knowledge representation System: A User's Manual," Technical Report TR238R, Computer Science Dept. University of Rochester, March 1989.

[Allen and Schubert, 1991] James F. Allen and Lenhart K. Schubert, "The TRAINS Project," TRAINS Technical Note 91-1, Computer Science Dept. University of Rochester, 1991.

[Appelt and Konolige, 1988] Douglas Appelt and Kurt Konolige, "A Nonmonotonic Logic for Reasoning About Speech Acts And Belief Revision," In *Proceedings of Second International Workshop on Non-monotonic Reasoning*, pages 164–175, 1988.

[Austin, 1962] J. A. Austin, *How to Do Things with Words*, Harvard University Press, 1962.

[Balkanski, 1990] Cecile T. Balkanski, "Modelling Act-Type Relations In Collaborative Activity," Technical Report 23-90, Harvard University Center for Research in Computing Technology, 1990.

[Barwise, 1989] Jon Barwise, *The Situation in Logic*, chapter 9: On the Model Theory of Common Knowledge, CSLI Lecture Notes: Number 17. Center for The Study of Language and Information, 1989.

[Brennan, 1990] Susan Brennan, *Seeking and Providing Evidence for Mutual Understanding*, PhD thesis, Stanford University, 1990.

[Bruce and Newman, 1978] Bertram Bruce and Denis Newman, "Interacting Plans," *Cognitive Science*, 2:195–233, 1978.

[Bruce, 1975] Bertram C. Bruce, "Generation as a Social Action," In *Theoretical Issues in Natural Language Processing-1*, pages 64–67, 1975, Also appears in [Grosz *et al.*, 1986], pp. 419-422.

[Cawsey, 1990] Alison Cawsey, "A Computational Model of Explanitory Discourse," In Paul Luff, Nigel Gilbert, and David Frohlich, editors, *Computers and Conversation.* Academic Press, 1990.

[Cawsey, 1991] Alison Cawsey, "A Belief Revision Model of Repair Sequences in Dialogue," In Ernesto Costa, editor, *New Directions in Intelligent Tutoring Systems.* Springer-Verlag, forthcoming, 1991.

[Chapman, 1987] David Chapman, "Planning for Conjunctive Goals," *Artificial Intelligence*, 32:333–377, 1987, also appears in [Allen *et al.*, 1990].

[Clark and Brennan, 1990] Herbert H. Clark and Susan E. Brennan, "Grounding in Communication," In L. B. Resnick, J. Levine, and S. D. Behrend, editors, *Perspectives on Socially Shared Cognition.* APA, 1990.

[Clark and Carlson, 1982] Herbert H. Clark and Thomas B. Carlson, "Critic's Beliefs about Hearers' Beliefs: A rejoinder to Johnson-Laird, Sperber, and Wilks," In N. V. Smith, editor, *Mutual Knowledge.* Academic Press, 1982.

[Clark and Marshall, 1981] Herbert H. Clark and Catherine R. Marshall, "Definite Reference and Mutual Knowledge," In Aravind K. Joshi, Bonnie L. Webber, and Ivan A. Sag, editors, *Elements of Discourse Understanding.* Cambridge University Press, 1981.

[Clark and Schaefer, 1989] Herbert H. Clark and Edward F. Schaefer, "Contributing to Discourse," *Cognitive Science*, 13:259 – 94, 1989.

[Clark and Wilkes-Gibbs, 1986] Herbert H. Clark and Deanna Wilkes-Gibbs, "Referring as a Collaberative Process," *Cognition*, 22, 1986.

[Cohen, 1978] Phillip R. Cohen, *On Knowing What to Say: Planning Speech Acts*, PhD thesis, University of Toronto, 1978, Reproduced as TR 118 Department of Computer Science, University of Toronto.

[Cohen and Levesque, 1990a] Phillip R. Cohen and Hector J. Levesque, "Performatives in a Rationally Based Speech Act Theory," In *Proceedings ACL-90*, pages 79–88, 1990.

[Cohen and Levesque, 1990b] Phillip R. Cohen and Hector J. Levesque, "Persistence, Intention, and Commitment," In P. R. Cohen, J. Morgan, and M. E. Pollack, editors, *Intentions in Communication.* MIT Press, 1990.

[Cohen and Levesque, 1990c] Phillip R. Cohen and Hector J. Levesque, "Rational Interaction as the Basis for Communication," In P. R. Cohen, J. Morgan, and M. E. Pollack, editors, *Intentions in Communication.* MIT Press, 1990.

[Cohen and Levesque, 1991a] Phillip R. Cohen and Hector J. Levesque, "Confirmations and Joint Action," In *Proceedings IJCAI-91*, pages 951–957, 1991.

[Cohen and Levesque, 1991b] Phillip R. Cohen and Hector J. Levesque, "Teamwork," *Nous*, 35, 1991.

[Cohen and Perrault, 1979] Phillip R. Cohen and C. R. Perrault, "Elements of a Plan-Based Theory of Speech Acts," *Cognitive Science*, 3(3):177-212, 1979.

[Demazeau and Muller, 1990] Y Demazeau and J. P. Muller, editors, *Decentralized A.I.*, Elsevier Science Publishers B. V., 1990.

[Demazeau and Muller, 1991] Y Demazeau and J. P. Muller, editors, *Decentralized A.I. 2*, Elsevier Science Publishers B. V., to appear 1991.

[Dipert, 1989] Randall R. Dipert, "The Structure of Agency: Issues in the Representation of Agency and Action," In *Proceedings of the First Annual SNePS Workshop*, pages 55–66, 1989, Available as TR 89-14 Dept. of Computer Science, State University of New York at Buffalo.

[Ferguson, 1991] George Ferguson, "Domain Plan Reasoning in TRAINS-90," TRAINS Technical Note 91-2, Computer Science Dept. University of Rochester, 1991.

[Frohlich and Luff, 1990] David Frohlich and Paul Luff, "Applying the Technology of Conversation to the Technology for Conversation," In Paul Luff, Nigel Gilbert, and David Frohlich, editors, *Computers and Conversation*. Academic Press, 1990.

[Galliers, 1989] Julia Rose Galliers, *A Theoretical framework For Computer Models Of Cooperative Dialogue, Acknowledgeing Multi-Agent Conflict*, PhD thesis, Open University, 1989, Also available as TR #172 from Cambridge University.

[Galliers, 1990] Julia Rose Galliers, "Belief Revision and A Theory of Communication," Technical Report 193, Cambridge University Computer Laboratory, 1990.

[Grosz and Sidner, 1986] Barbara Grosz and Candice Sidner, "Attention, Intention, and the Structure of Discourse," *CL*, 12(3):175–204, 1986.

[Grosz *et al.*, 1986] Barbara J. Grosz, Karen Spark Jones, and Bonnie Lynn Webber, editors, *Readings In Natural Language Processing*, Morgan Kaufmann, 1986.

[Grosz and Sidner, 1990] Barbara J. Grosz and Candace L. Sidner, "Plans for Discourse," In P. R. Cohen, J. Morgan, and M. E. Pollack, editors, *Intentions in Communication*. MIT Press, 1990.

[Halpern and Moses, 1990] J. Y. Halpern and Y. Moses, "Knowledge and Common Knowledge in a Distributed Environment," *Journal of the ACM*, 37(3):549–587, 1990.

[Harman, 1977] Gilbert Harman, "Review of *Linguistic Behaviour* by Jonathan Bennett," *Language*, 53:417–424, 1977.

[Hinkelman, 1990] Elizabeth Hinkelman, *Linguisic and Pragmatic Constraints on Utterance Interpretation*, PhD thesis, University of Rochester, 1990.

[Hintikka, 1962] Jaakko Hintikka, *Knowledge and belief; an introduction to the logic of the two notions*, Cornell University Press, 1962.

[Hobbs and Evans, 1979] Jerry Hobbs and David Evans, "Conversation as Planned Behavior," Technical Report 203, SRI International, December 1979.

[Johnson-Laird, 1982] P. N. Johnson-Laird, "Mutual ignorence: Comments on Clark and Carlson's Paper," In N. V. Smith, editor, *Mutual Knowledge*. Academic Press, 1982.

[Konolige, 1988] Kurt Konolige, "Hierarchic autoepistemic theories for nonmonotonic reasoning: preliminary report," In *Proceedings of Second International Workshop on Nonmonotonic Reasoning*, pages 42–59, 1988.

[Levesque *et al.*, 1990] Hector J. Levesque, Phillip R. Cohen, and Jose H. T. Nunes, "On Acting Together," In *Proceedings AAAI-90*, pages 94–99, 1990.

[Levinson, 1983] Stephen C. Levinson, *Pragmatics*, Cambridge University Press, 1983.

[Lewis, 1969] David K. Lewis, *Convention: A Philosophical Study*, Harvard University Press, 1969.

[Litman, 1985] Diane J. Litman, *Plan Recognition and Discourse Analysis: An Integrated Approach for Understanding Dialogues*, PhD thesis, University of Rochester, 1985, Reproduced as TR 170 Department of Computer Science, University of Rochester.

[Litman and Allen, 1990] Diane J. Litman and James F. Allen, "Discourse Processing and Common Sense Plans," In P. R. Cohen, J. Morgan, and M. E. Pollack, editors, *Intentions in Communication*. MIT Press, 1990.

[Lochbaum *et al.*, 1990] Karen E Lochbaum, Barbara J. Grosz, and Candace L. Sidner, "Models of Plans to Support Communication: An Initial Report," In *Proceedings AAAI-90*, pages 485–490, August 1990.

[Luff *et al.*, 1990] Paul Luff, Nigel Gilbert, and David Frohlich, editors, *Computers and Conversation*, Academic Press, 1990.

[Maida, 1984] Anthony S. Maida, "Belief Spaces: Foundations of a computational Theory of Belief," Technical Report Cs-84-22, Pennsylvania State University Department of Computer Science, December 1984.

[Moore, 1989] Johanna D. Moore, *A Reactive Approach to Explanation in Expert and Advice Giving Systems*, PhD thesis, University of California, Los Angeles, 1989.

[Nakajima and Allen, 1991] Shin'ya Nakajima and James F. Allen, "A Study of Pragmatic roles of prosody in the TRAINS Dialogs," TRAINS Technical Note, Computer Science Dept. University of Rochester, forthcoming 1991.

[Perner and Garnham, 1988] Josef Perner and Alan Garnham, "Conditions for Mutuality," *Journal of Semantics*, 6:369–385, 1988.

[Perrault, 1984] C. R. Perrault, "On the Mathematical Properties of Linguistic Theories," *Computational Linguistics*, 10:165–176, 1984, Also appears in [Grosz *et al.*, 1986].

[Perrault, 1990] C. R. Perrault, "An Application of Default Logic to Speech Act Theory," In P. R. Cohen, J. Morgan, and M. E. Pollack, editors, *Intentions in Communication.* MIT Press, 1990.

[Perrault and Allen, 1980] C. Raymond Perrault and James F. Allen, "A Plan-Based Analysis of Indirect Speech Acts," *American Journal of Computational Linguistics*, 6(3-4):167–82, July-December 1980.

[Pollack, 1990] Martha Pollack, "Plans as Complex Mental Attitudes," In P. R. Cohen, J. Morgan, and M. E. Pollack, editors, *Intentions in Communication.* MIT Press, 1990.

[Raudaskoski, 1990] Pirkko Raudaskoski, "Repair Work in Human-Computer Interaction," In Paul Luff, Nigel Gilbert, and David Frohlich, editors, *Computers and Conversation.* Academic Press, 1990.

[Reiter, 1990] Ehud Reiter, "The Computational Complexity of Avoiding Conversational Implicatures," In *Proceedings ACL-90*, pages 97–104, 1990.

[Reiter, 1980] R. Reiter, "A Logic for Default Reasoning," *AI*, 13(1,2):81–132, April 1980.

[Sacks *et al.*, 1974] H. Sacks, E. A. Schegloff, and G. Jefferson, "A Simplest Systematics For the organization of Turn-Taking for Conversation," *Language*, 50:696–735, 1974.

[Schegloff *et al.*, 1977] E. A. Schegloff, G. Jefferson, and H. Sacks, "The Preference for Self Correction in the Organization of Repair in Conversation," *Language*, 53:361–382, 1977.

[Schegloff, 1987] Emmanuel A. Schegloff, "Recycled Turn-beginnings: A precise repair mechanism in conversation's turn-taking organization," In Graham Button and John R. E. Lee, editors, *Talk and Social Organization.* Multilingual Matters, 1987.

[Schiffer, 1972] Stephen R. Schiffer, *Meaning*, Oxford University Press, 1972.

[Searle, 1990] John R. Searle, "Collective Intentions and Actions," In P. R. Cohen, J. Morgan, and M. E. Pollack, editors, *Intentions in Communication.* MIT Press, 1990.

[Searle and Vanderveken, 1985] John R. Searle and Daniel Vanderveken, *Foundations of Illocutionary Logic*, Cambridge University Press, 1985.

[Sperber and Wilson, 1986] Dan Sperber and Deirdre Wilson, *Relevence: Communication and Cognition*, Harvard University Press, 1986.

[Suchman, 1987] Lucy A. Suchman, *Plans and Situated Actions*, Cambridge University Press, 1987.

[Traum, 1991] David R. Traum, "The Discourse Reasoner in TRAINS-90," TRAINS Technical Note 91-5, Computer Science Dept. University of Rochester, 1991.

[Traum and Allen, 1991] David R. Traum and James F. Allen, "Causative Forces in Multi-agent Planning," In Y Demazeau and J. P. Muller, editors, *Decentralized A.I. 2*. Elsevier Science Publishers B. V., to appear 1991.

[Turner, 1989] Elise Hill Turner, *Integrating Intention and Convention To Organize Problem Solving Dialogues*, PhD thesis, Georgia Institute of Technology, 1989.