Conversation Acts in Task-Oriented Spoken Dialogue *

David R. Traum[†]
Computer Science Department
University of Rochester
Rochester, New York 14627
USA
traum@cs.rochester.edu

Elizabeth A. Hinkelman[‡]
DFKI
Stuhlsatzenhausweg 3
D-W-6600 Saarbruecken 11,
Germany
hinkelman@dfki.uni-sb.de

Technical Report 425

Abstract

A linguistic form's compositional, timeless meaning can be surrounded or even contradicted by various social, aesthetic, or analogistic companion meanings. This paper addresses a series of problems in the structure of spoken language discourse, including *turn-taking* and *grounding*. It views these processes as composed of fine-grained actions, which resemble speech acts both in resulting from a computational mechanism of planning and in having a rich relationship to the specific linguistic features which serve to indicate their presence.

The resulting notion of *Conversation Acts* is more general than speech act theory, encompassing not only the traditional speech acts but turn-taking, grounding, and higher-level *argumentation acts* as well. Furthermore, the traditional speech acts in this scheme become fully joint actions, whose successful performance requires full listener participation.

This paper presents a detailed analysis of spoken language dialogue. It shows the role of each class of conversation acts in discourse structure, and discusses how members of each class can be recognized in conversation. Conversation acts, it will be seen, better account for the success of conversation than speech act theory alone.

---

*To appear, COMPUTATIONAL INTELLIGENCE Special Issue: Computational Approaches to Non-Literal Language, vol 8, no 3, August 1992

[†]supported in part by the NSF under research grant no. IRI-9003841, by ONR under research grant no. N00014-90-J-1811, and by DARPA/ONR under contract N00014-92-J-1512.

[‡]supported by the DFKI

1 Introduction

This paper concerns the underpinnings of conversation, particularly the methods by which participants establish a mutual understanding or common ground of conversational content being discussed. This process of *grounding* [Clark and Schaefer, 1989] introduces opportunities for meaning to differ from classical timeless, compositional meaning as well. Consider

Interviewer: Do you feel that any of these spectacular events of the early '60s, such as the police raid on the Santa Monica concert, that bat, that infamous bat whose head you bit off, the, uh, the those first psychedelic drug parties, ... do you feel they had any impact on your career, like, enhanced your reputation with the adolescent audience?

Rock Star: Yes. No. No, Bill, I believe my rise to fame was inevitable.

When viewed solely at the ordinary level of meaning, the Star's response may appear confused. However, the initial "yes" serves in this context not as an Answer to the question posed, but rather, it grounds the question itself and communicates that the Star has understood it.

1.1 Extending Conversation

Grounding mechanisms are essential to the progress of spoken conversation, as a look at an extended dialogue reveals.

Figure 1 shows a transcript of a short conversation taken from the TRAINS domain[1]. The transcript can be read as follows: the first column lists the utterance (UU) number. The numbering is by turn on the left side of the ".", and utterance within the turn to the right of the ".". Thus UU 15.3 is the 3rd utterance in the 15th turn. Principles for the determination of utterance units are given in [Nakajima and Allen, 1991]. 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, aiding the manager to construct and execute plans in a transportation and manufacturing domain. Once constructed, plans are executed by (simulated) remote agents such as train engineers and factory operators. This dialogue is drawn from a target corpus of spoken conversations with a person playing the role of the system.

In these dialogues, the manager (M) brings to the system (S) a task to accomplish in this domain. In the conversation in Figure 1, this task is revealed in utterance 1.1 - to ship a boxcar of oranges to Bath by 8 am (the current time is midnight). M has a map of the domain, shown in Figure 2. S has a similar map and access to additional information, such as the time required for transit and loading and manufacturing tasks. M and S are charged with constructing a plan to satisfy this task (which the actual computer system would then convey to the agents in the (simulated) TRAINS world for execution).

One of the most striking features of task-oriented dialogues is the pervasiveness of acknowledgement and agreement signals. If S had more initiative, some such signals might

[1]TRAINS Conversation 91-6.1 in [Gross *et al.*, 1992]

UU#	Speaker: Utterance
1.1	M: okay, the problem is we better ship a boxcar of oranges to Bath by 8 AM.
2.1	S: okay.
3.1	M: now ... umm ... so we need to get a boxcar to Corning, where there are oranges.
3.2	: there are oranges at Corning
3.3	: right?
4.1	S: right.
5.1	M: so we need an engine to move the boxcar
5.2	: right?
6.1	S: right.
7.1	M: so there's an engine at Avon
7.2	: right?
8.1	S: right.
9.1	M: so we should move the engine at Avon,
9.2	: engine E,
9.3	: to .. (inc)
10.1	S: engine E1
11.1	M: E1.
12.1	S: okay
13.1	M: engine E1, to Bath, to (inc)
13.2	: or, we could actually move it to Dansville, to pick up the boxcar there
14.1	S: okay
15.1	M: um and hook up the boxcar to the engine,
15.2	: move it from Dansville to Corning,
15.3	: load up some oranges into the boxcar,
15.4	: and then move it on to Bath.
16.1	S: okay.
17.1	M: how does THAT sound?
18.1	S: that gets us to Bath at 7 AM,
18.2	: and (inc)
18.3	: so that's no problem.
19.1	M: good.
20.1	S: ok.

Figure 1: Sample Conversation from TRAINS Domain

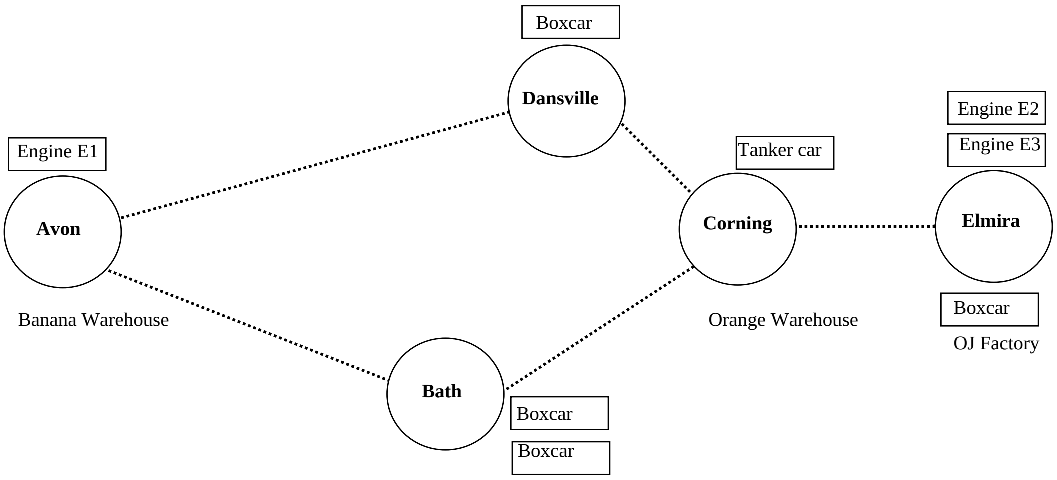

Figure 2: Trains World Set-up for Conversation in Figure 1

be implicit in S's more extended contributions, but here they are explicit and readily identifiable. One complication that arises is similar to that of the Rock Star in the previous example. Utterance 16.1 differs from its neighbors 12.1 and 14.1 in that it signals acknowledgement only, without agreement. This paper will focus on dialogue features like acknowledgement, which serve in coordination and maintenance of the dialogue itself, rather than as a direct part of the domain information communicated.

1.2 From Speech Acts to Conversation Acts

Our approach to dialogue is a generalization of speech act theory, a theory of *Conversation Acts*. Conversation acts extend prior computational speech act work to certain types of coordinated activity that take place between agents in a conversation. In order to accomplish this, we have eliminated some assumptions common in prior speech act work (e.g. [Allen and Perrault, 1980; Bunt, 1989; Cohen and Perrault, 1979; Litman and Allen, 1990; Perrault, 1990]), namely:

1. Utterances are heard and understood correctly by the listener as they are uttered, moreover, it is mutually expected by both participants that this will be the case.
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.

Each of these assumptions is too strong for 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 the listener has understood. Some acknowledgements are made with explicit utterances (e.g. 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 communication failure, 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.

Conversation Acts model discourse as a collection of joint speaker-hearer actions, whose performance results in meaning specific to a particular conversation and grounded to the satisfaction of both participants. Conversation Acts provide a more comprehensive approach to communication than speech act theory alone; one which better accounts for the degree to which meaning is conveyed at all.

In Section Two we will introduce four classes of conversation acts, of which grounding acts and traditional speech acts are two. Section Three is a detailed examination of the conversation acts present in the conversation in Figure 1. Section Four suggests ideas on how the classes of acts may actually be recognized and integrated. Section Five briefly distinguishes this proposal from similar taxonomies of conversational action.

2 Conversation Acts

We distinguish four levels of action necessary for maintaining the coherence and content of conversation. Action attempts at any of these levels can be signalled directly by surface features of the discourse, although usually a combination of surface features and context will be necessary to disambiguate acts. Reading Table 1 from top to bottom, progressive levels are typically realized by larger and larger chunks of conversation, from turn-taking acts, usually realized sub-lexically to argumentation acts which can span whole conversations. It is important to note, however, that according to the terminology of [Halliday, 1961] these classes are *levels* of language description, and not *ranks*. That is, the distinction between these classes is more like between that of phonology and syntax rather than than between a word and a phrase; e.g. there is no grammar which will build up a grounding act as an ordered collection of turn-taking acts. This notion will be elaborated on below in section 5.

2.1 The Core Speech Acts: DU Acts

In adapting speech act models to spoken discourse, we maintain traditional speech acts such as **Inform**, **Request** and **Promise**, referring to them as *Core Speech Acts*. To model the multi-utterance exchanges necessary for mutual understanding of Core Speech Acts, we posit a level of structure called a **Discourse Unit** (DU). A DU consists of an initial presentation, and as many subsequent utterances by each party as are needed to make the act mutually understood, or *Grounded*. The initial presentation is best considered a Core Speech Act *attempt*, which is not fully realized until its DU is grounded. A minimal DU contains an initial presentation and an acknowledgement (which may be implicit in the next presentation by another speaker). However, it may also include any repairs or continuations that are needed to realize the act. A discourse unit corresponds more or less to a top level *Contribution*, in the terminology of [Clark and Schaefer, 1989].

Discourse Level	Act Type	Sample Acts
Sub UU	**Turn-taking**	take-turn keep-turn release-turn assign-turn
UU	**Grounding**	Initiate Continue Ack Repair ReqRepair ReqAck Cancel
DU	**Core Speech Acts**	Inform WHQ YNQ Accept Request Reject Suggest Eval ReqPerm Offer Promise
Multiple DUs	**Argumentation**	Elaborate Summarize Clarify Q&A Convince Find-Plan

Table 1: Conversation Act Types

2.2 Argumentation Acts

We may build higher level discourse acts out of combinations of core speech acts. We may, for instance, use an **inform** act in order to summarize, clarify, or elaborate prior conversation. A very common argumentation act 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], e.g the whole conversation in Figure 1. The kinds of actions generally referred to as *Rhetorical Relations* [Mann and Thompson, 1987] take place at this level, as do many of the actions signalled by cue phrases, and so called *Adjacency Pairs* [Schegloff and Sacks, 1973].

2.3 Grounding Acts: UU Acts

An *Utterance Unit* (UU) is defined as more or less continuous speech by the same speaker, punctuated by prosodic boundaries. 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 (see below). Grounding Acts include:

Initiate An initial utterance component of a Discourse unit - traditionally this utterance alone has been considered sufficient to accomplish the core speech act. An `initiate` usually corresponds to the (first utterance in the) presentation phase of a top level Contribution in [Clark and Schaefer, 1989].

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. For example, the compound plan elaboration UU 15.2-15.4 in Figure 1 are all (at the Grounding Act Level) continuations of the DU begun by UU 15.1.

Acknowledge Shows understanding of a previous utterance. It may be either a repetition or paraphrase of all or part of the utterance (e.g. UU 11.1), a *backchannel response* (e.g. e.g. UU 2.1, 16.1), or implicit signalling of understanding, such as by proceeding with the initiation of a new DU which would naturally follow the current one in the lowest level argumentation act. Typical cases of implicit acknowledgement are answers to questions, (e.g. UU 18.1). Acknowledgements are also referred to by some as *confirmations* (e.g. [Cohen and Levesque, 1991]) or *acceptances* (e.g. [Clark and Schaefer, 1989]). We prefer the term *acknowledgement* as unambiguously signalling understanding, reserving the term *acceptance* for a Core Speech Act signalling agreement with a proposed domain plan.

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 change the interpretation of the speaker's intention. A `repair` can change either the content or Core Speech Act type of acts in the current DU (e.g. a tag question can change an `Inform` to a `YNQ`). `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 are argumentation acts.

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 `acknowledge` only by intonation. A `ReqRepair` invokes a discourse obligation on the listener to respond with either the requested repair, or an explicit refusal or postponement (e.g a followup request).

ReqAck Attempt to get the other agent to acknowledge the previous utterance. This invokes a discourse obligation on the listener to respond with either the requested acknowledgement, or an explicit refusal or postponement (e.g a followup repair or repair request).

Cancel Closes off the current DU as ungrounded. Rather than repairing the current DU, a `cancel` abandons it; the underlying intention, if it is still held, must be expressed in a new DU. An example of a cancel is UU 13.2, which retracts the suggestion (started in UU 13.1) to go to Bath for the needed Boxcar before S has a chance to respond.

2.4 Turn-taking Acts: Sub UU Acts

We posit a series of low level acts to model the turn-taking process [Sacks *et al.*, 1974; Orestrom, 1983]. The basic acts are **keep-turn**, **release-turn** (with a subvariant, **assign-turn**) and **take-turn**.

There may be several turn-taking acts in a single utterance. The start of an utterance might be a take-turn action (if another party initially had the turn), the main part of the utterance might be keeping the turn, and the end might release it. Conversants can attempt these acts by any of several common speech patterns, ranging from propositional (e.g. "let me say something") to lexical (e.g. "umm" in UU 3.1) to sublexical. Many turn-taking acts are signalled with different intonation patterns and pauses. Although a conversant can attempt a turn-taking action at any time, it will be a matter of negotiation as to whether the attempt succeeds. Conversational participants may engage in a "floor battle" where one tries to keep the turn while another tries to take it. 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, if cooperative, may then facilitate it (e.g. refraining from taking a turn when signalled that another wants to keep it, or releasing when another wants to take the turn).

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. Pauses are opportunities for anyone to take the turn. "Filling" pauses with such utterances as "uhh" can signal desire to keep the turn through what might otherwise be seen as a `release-turn`. Certain pauses are marked by context (e.g a previous topic introduction or request) as to who has the turn. Even here, an excessive pause can open up the possibility of a `take-turn` action by another conversant.

Release turn actions are usually 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. Another is naming the next speaker.

Another act, which would be necessary in a face-to-face, or multi-channel communication situation would be **pass-up-turn**. Back-channel items such as "okay" and other signals of attention such as gestures are often (e.g. [Yngve, 1970; Duncan and Niederehe, 1974]) analyzed together as not taking a turn, leaving the previous speaker in control. Because (in our domain setting) all of these items must proceed through the same (the audio) channel, and the other speaker does stop and wait for the response before proceeding on (e.g in utterance 13.2 - 15.2) we analyze these short utterances as a **take-turn** followed quickly by a **release-turn**. The only thing we might classify as **pass-up-turn** would be silence.

3 Conversations Acts in the Sample Conversation

We have distinguished four very different communicative functions, and introduced the principal actions that serve each. Now we will show how these communicative functions are realized in our conversation data.

Figures 3 and 4 show the TRAINS conversation from figure 1 again, with relevant utterance-final features annotated using the Pierrehumbert pitch description system [Pierrehumbert and Hirschberg, 1990]. This system uses two underlying pitch primitives: H (high), and L (low). An intonational phrase is composed from lexical pitch accents (with stressed syllables marked with "*"), a phrase accent at the end of each intermediate phrase, and a final boundary tone designated with "%". There is also a scheme for realization of these underlying pitch features in the utterance. These judgements were made without reference to processed digital signals, so the reader must use them only as a rough guide. We have marked only clearly identifiable features of the utterances; note that the division of the text into lines often corresponds to a phrase accent but not an utterance boundary tone. We have also annotated final lowering, a less local phenomenon, with dots.

3.1 Turn-Taking Acts

Turn Taking acts are difficult to illustrate with a transcript, since they depend heavily on timing and prosodic features. Nevertheless, we will point out a few in the example conversation. Keep-turn actions are realized in several different ways in this dialogue. In utterance 3.1, this is the main purpose of the items "now" and "umm". Here, the "now" could also signal a topic shift (from specifying the goal to working on the plan, see Section 3.4 below). But this would still be consistent with the fact that M has clearly not thought out what to say next, and wants time to work it out rather than letting S take over, or permitting awkward silence. In the same vein, the flat end of 3.1 might momentarily be seen as a release-turn, but the quick start of UU 3.2 is a clear keep-turn. The stretched endings in 15.1-15.3 all signal keep-turns as well, allowing continuation past both clause boundary and semantic completion (see [Ford and Thompson]).

Release-turn actions are, in the ReqAck forms 3.3, 5.2, and 7.2, signalled by HH% contours. The turn is also released after the wh-question form in 17.1, and after declarative sentences 1.1, 15.4 and 18.3. Note that in a two-party conversation there is relatively little difference between a release-turn and an assign-turn, since there is only one potential respondent.

Within each turn, a take-turn action occurs at the beginning of the first utterance (beginning of utterance x.1 for any x). Of special note is the pause in 9.3, which S takes as an opportunity to insert repair 10.1 although there is no prior release-turn from M.

This conversation is unusual in having no overlapped speech or floor contention. Other dialogues from this study [Gross *et al.*, 1992] contain many examples of take-turn and keep-turn acts which fail when the other party does not yield the turn. They also contain failed release-turns and assign-turns, where the floor is not taken up by the other party.

```
1.1 okay, the problem is we better ship  a boxcar of oranges to Bath  by 8 AM.

2.1 okay.

     *       *      L*LL%                    *    H      *    H .........*..   H
3.1 now ... umm ... so     we need to get a boxcar  to Corning, where there are oranges.

                          H*   L
3.2 there are oranges at Corning

       H*HH%
3.3 right?

       L*LL%
4.1 right.

                    *                  *  *
5.1 so we need an engine to move the boxcar

       H*HH%
5.2 right?

       L*LL%
6.1 right.

                        ....L*  H
7.1 so there's an engine at Avon

       H*HH%
7.2 right?

       L*LL%
8.1 right.

                            . . . .      *   H
9.1 so we should     move the engine at Avon ,

           *H
9.2 engine E,

9.3 to .. (inc)

            **HH%
10.1 engine E1

     **
11.1 E1.

     *  *H
12.1 okay
```

Figure 3: TRAINS Domain Conversation with Intonational Features: First Part

```
         **H    H*
13.1 engine E1, to Bath, to...

     H*                                H*                 .........*..*.. H*..LL%
13.2 or, we could actually move it to Dansville, to pick up the boxcar there

     * * H
14.1 okay

                         H*     ........*    H
15.1 um  and hook up the boxcar to the engine,

                H*     ......H*.....H
15.2 move it from Dansville to Corning,

                H               .... H*.   H
15.3 load up some oranges     into the boxcar,

               ..............*  LL%
15.4 and then move it on to Bath.

     * * H
16.1 okay.

      *   *    H* ...*  LL%
17.1 how does THAT sound?

      *    *          *   ...H*...LL%
18.1 that gets us to Bath at 7  AM,

18.2 and (inc)

                .........  LL%
18.3 so that's no problem.

       L*LL%
19.1 good.

20.1 ok.
```

Figure 4: TRAINS Domain Conversation with Intonational Features - Second Part

3.2 Grounding Acts

Figure 5 shows the conversation from Figure 1 labelled with the grounding acts which correspond to each utterance. Each act is subscripted with the number of the DU of which it is a part. Repairs, Continues, and demonstration style acknowledgements have in parentheses the UU which they are most directly connected to.

UU Act	UU#	Speaker: Utterance
$init_1$	1.1	M: okay, the problem is we better ship a boxcar of oranges to Bath by 8 AM.
ack_1	2.1	S: okay.
$init_2$	3.1	M: now ... umm ... so we need to get a boxcar to Corning, where there are oranges.
$init_3$	3.2	: there are oranges at Corning
$reqack_3$	3.3	: right?
ack_3 $init_4$	4.1	S: right.
ack_4 $init_5$	5.1	M: so we need an engine to move the boxcar
$reqack_5$	5.2	: right?
ack_5 $init_6$	6.1	S: right.
ack_6 $init_7$	7.1	M: so there's an engine at Avon
$reqack_7$	7.2	: right?
ack_7 $init_8$	8.1	S: right.
ack_8 $init_9$	9.1	M: so we should move the engine at Avon,
$repair_9$(9.1)	9.2	: engine E,
$cont_9$(9.1)	9.3	: to .. (inc)
$repair_9$(9.2)	10.1	S: engine E1
ack_9(10.1)	11.1	M: E1.
ack_9	12.1	S: okay
$init_{10}$	13.1	M: engine E1, to Bath, to (inc)
$cancel_{10}init_{11}$	13.2	: or, we could actually move it to Dansville, to pick up the boxcar there
ack_{11}	14.1	S: okay
$init_{12}$	15.1	M: um and hook up the boxcar to the engine,
$cont_{12}$(15.1)	15.2	: move it from Dansville to Corning,
$cont_{12}$(15.2)	15.3	: load up some oranges into the boxcar,
$cont_{12}$(15.3)	15.4	: and then move it on to Bath.
ack_{12}	16.1	S: okay.
$init_{13}$	17.1	M: how does THAT sound?
ack_{13} $init_{14}$	18.1	S: that gets us to Bath at 7 AM,
$cont_{14}$(18.1)	18.2	: and (inc)
$cont_{14}$(18.1)	18.3	: so that's no problem.
ack_{14} $init_{15}$	19.1	M: good.
ack_{15}	20.1	S: ok.

Figure 5: Conversation with Grounding Acts

We can see all three types of acknowledgements in this short dialogue. UU 11.1 is a demonstration style acknowledgement: repeating of "E1" by M demonstrates explicit receipt of the repair in 10.1. UUs 2.1, 12.1, 14.1, 16.1, and 20.1 are backchannel acknowledgements, claiming receipt but not demonstrating what they are acknowledging. UUs 5.1, 7.1, 9.1, and 19.1 are implicit acknowledgements, recognizable as initiations of DUs whose contents are next steps after the current DU in argumentation level acts. Thus 5.1 and 7.1 cover further steps in the domain plan (see Figure 7, below), and 19.1 gives a relevant evaluation. UUs 4.1, 6.1, and 8.1 are a middle ground between the paraphrase type (in virtue of the repeated lexeme, though with different intonation) and the implicit type - affirming the checks made in the previous turn.

UUs 3.3, 5.2, and 7.2 are all requests for acknowledgement, making more explicit and intense the discourse obligation to acknowledge the current DU which is normally an implicature of a release-turn after an initiate action. An alternative interpretation for the grounding acts performed by these utterances would be to see them as tag-style repairs to the Core Speech Act types in their respective DUs changing the types from informs to questions, but this analysis is deemed unlikely for reasons given in Section 3.3, below.

`Initiate` can be distinguished from `continue` mainly by context. If there is an ungrounded open DU for which the current utterance forms a syntactic continuation, the current utterance is seen as a `continue`. The very same utterance occurring after an interjected acknowledgement is an `initiate` of a new DU. Thus UU 15.2 is a `continue`, while UU 15.1 is an `initiate`, though on the syntactic level and the domain plan level both just continue the plan from UU 13.2. Note that if the acknowledgement in UU 14.1 were absent, 15.1 would be marked as a `continue` as well. Notice further that 3.2 is labelled an `initiate`, in spite of DU #2 still being open and ungrounded. This is because of the abrupt change in sentence and speech act between 3.1 and 3.2. The importance of the distinction is this: when we get an acknowledgement, how much stuff is being acknowledged? A backchannel acknowledgement such as UU 16.1 grounds the entire DU - the `initiate` in 15.1 and the subsequent continuations in 15.2, 15.3, and 15.4. It does not ground 13.2, because that is already grounded. Similarly, 4.1 explicitly grounds only 3.2 and 3.3, leaving in question whether 3.1 is grounded (see discussion below).

The status of UU 9.2 is also somewhat controversial - relying mainly on one's theory of sentence syntax and whether it is repairing a (potentially) inadequate referring expression ("the engine at Avon") in UU 9.1, or the two together form a complex referring expression, giving both name and location. It seems that certain complex syntactic phenomena such as asides, vocatives, tags, and left and right dislocation might potentially be seen as separate acts which are really connected only by conversational structure, but a detailed proposal is beyond the scope of the present work (but see also [McCawley, 1988] pp. 763-766, [McCawley, 1989] for a similar proposal).

This conversation has no repair-requests (assuming the ReqAck analysis for UU 3.3, 5.2, and 7.2), but a simple example might be if UU 10.1 had had a rising intonation, or was replaced with "engine what?".

UU 13.2 can be seen as a cancel of DU #10, as well as a new `initiate`, in virtue of the "or" and "actually" phrasing. The suggestion in 13.1 to move Engine E1 to Bath is abandoned and left ungroundable. 14.1 grounds the suggestion to move to Dansville, not the disjunction of Bath or Dansville. The explicit cancel distinguishes DU #10, which is certainly ungrounded from DU # 2, which has a more questionable status. Even though UU 3.2 initiates a new DU, it is still possible to ground DU #2 after this point.

3.3 DUs and Core Speech Acts

Table 2 shows more information about the DUs in the conversation, listing for each DU which agent was the initiator, what were the types of its constituent Core Speech Acts (superscripted with the performing agent), and the UUs which comprise it. Ungrounded DUs (e.g. DU #10) have their UU list concluded with a "*". DU #2's status is questionable (see discussion below) hence the "?*". UUs which were intended to be part of the DU, but which were abandoned (e.g. UU 9.3, 18.2) are listed in parens. That they were abandoned can also be seen in Figure 5 by the UU arguments of their successor acts, which refer to prior UUs.

DU#	Initiator	Core Speech Act types	Included UUs
1	M	inform^M suggest(goal)^{M}accept^S	1.1 2.1
2	M	inform^M suggest^M	3.1 ?*
3	M	check^M ?suggest^M	3.2 3.3 4.1
4	S	inform-if^S ?accept^S	4.1 5.1
5	M	check^M	5.1 5.2 6.1
6	S	inform-if^S	6.1 7.1
7	M	check^M ?suggest^M	7.1 7.2 8.1
8	S	inform-if^S ?accept^S	8.1 9.1
9	M	suggest^M accept^S	9.1 9.2 (9.3) 10.1 11.1 12.1
10	M	suggest^M	13.1 *
11	M	suggest^M accept^S	13.2 14.1
12	M	suggest^M	15.1 15.2 15.3 15.4 16.1
13	M	request(eval)^M	17.1 18.1
14	S	inform^S accept^S	18.1 (18.2) 18.3 19.1
15	M	eval^M	19.1 20.1

Table 2: DU Acts from Conversation

In DU #1, M both informs S of the designated problem and suggests that this be the goal of the plan they construct. There are several paths by which this utterance can be recognized as a suggestion: one is inference from S's expectation that M will propose such a goal (see Section 3.4). A second is that in a cooperative environment an inform of a need can be sufficient to convey the suggestion of addressing it. S's acknowledgement in UU 2.1 grounds the DU and also accepts the suggestion to work on achieving the suggested goal.

UU 3.1 in DU #2 also provides two core Speech acts - a literal inform of the obligation and an indirect suggestion of a necessary subaction which can be recognized as such in the context of problem-solving in the TRAINS domain. It is uncertain whether or not this DU is ever grounded. The content certainly ends up in the final plan, but it could have gotten there just as well through DUs #3, #11, and #12. UU 4.1 might be an (implicit) acknowledgement of DU #2 as well as of DU #3, but this is not certain. Another possibility is that UU 6.1 acknowledges DU #2. The reasoning for this is as follows: UU 6.1 grounds DU #5, claiming to have understood UU 5.1. But UU 5.1 uses a definite reference "the boxcar" which is not licensed by non-linguistic context (The map in Figure 2 shows more than one boxcar, any of which would be fine in a plan to move oranges to Bath). The only thing that licenses this use is the mention of "a boxcar" in UU 3.1 - the boxcar which has to get to Corning to get the oranges. So by claiming understanding of UU 5.1, S is indirectly claiming understanding of (at least part of) UU 3.1, and thus grounding at least that part (if it is not already grounded).

The surface actions in DUs #3,#5, and #7 are all clearly `check` actions in spite of their surface declarative form, in virtue of the respective knowledge preconditions [Bunt, 1989; Beun, 1989]. A check differs from an ordinary Yes-No question in that for a Yes-No question the initiator does not know the answer, whereas for a check the initiator does know and is making sure both conversants are in synch. A check is also of the *question* type rather than the *inform* type, as it requires positive confirmation, not just neutral acknowledgement (replacing UU 4.1 with "okay" does not satisfy the discourse obligation, and replacing it with "oh" - which signals a change in information state in the "oh" producer (see [Heritage, 1984]) violates the presupposition behind the utterance that this information was known by the responder).

We can see by looking at Figure 2 (which both conversants had independent access to) that the contents of DUs #3, and #7 were already privately known. The knowledge in DU #5 is background knowledge of the physics of the TRAINS world which could conceivably be unknown to a novice M, but it would surely be known to S, and the confident tone of voice does not indicate a YNQ. Given that the background knowledge and prosodics already indicate a check for each of UUs 3.2, 5.1, and 7.1, it is unnecessary to regard UUs 3.3, 5.2, and 7.2 as repairs, as we might do if we had initially considered these DUs to contain informs.

In addition, DUs #3 and #7 might have indirect suggestion readings. It is consistent to regard UU 3.2 as suggesting which oranges to use in the plan (although this could have been done previously with UU 3.1 or subsequently with 15.2, 15.3). It is also consistent to regard UU 7.1 as suggesting the Engine to use, though this is made more explicit in DU #9. If we did not have intonation to mark these acts as checks, and since the inform possibilities are ruled out by the domain knowledge, these suggestions would be our only likely alternative (see [Traum, 1991a] for examples in another dialogue where this is indeed the case). As it stands, we can not be sure exactly what was meant or understood, though it doesn't matter for the success of the conversational goals.

In a like manner, DUs #4 and #8 have surface forms of Inform-if (a specialization of inform telling whether or not a proposition is deemed true), though they may also be accepting the suggestions described above.

UU 13.1 in DU #10 clearly begins a suggestion, a followup to that in DU #9, but it is cancelled and does not appear in the final plan.

DU #12 ends up as one big compound suggestion which is acknowledged but NOT accepted with UU 16.1. That this is not an accept (as UUs 2.1, 12.1, and 14.1 are) is indicated by the lengthening of the second syllable which suggests a lack of commitment. That M also interpretated it this way is suggested by his followup request in UU 17.1, explicitly asking for evaluation of this proposed plan.

3.4 Argumentation Acts

Unlike the other classes of acts described here, argumentation acts build up hierarchically within the same class. At the high level, acts are mainly derivative of the domain task structure - what the conversation is being used by the participating agents to do. At the lower level, cross-domain rhetorical practices are more common - it doesn't matter so much what the task is, there are standard conventional means of achieving it.

The top level argumentation acts in the TRAINS Conversations are shown in Figure 6. The Manager must first specify the goal, then the participants must construct the plan and the System must verify that the plan meets the agreed upon goals. While, due to different initial knowledge, the Manager is primarily responsible for specifying the goals and the System for verifying the plan, all parts of this process must be grounded to the satisfaction of both parties for the conversation to proceed and conclude successfully. The initiative for constructing the plan is left unspecified by the domain, although in the conversations we have collected in [Gross *et al.*, 1992] the System plays a mainly passive role (as in this dialog), offering his own suggestions only when asked.

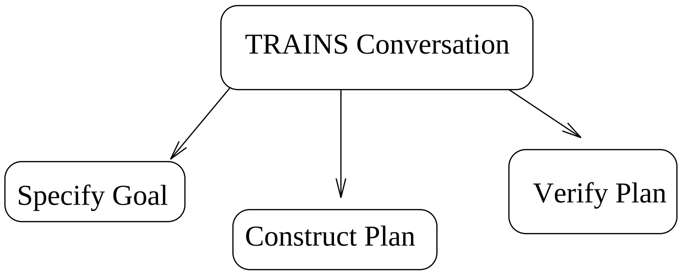

Figure 6: Top Level Trains Conversation Plan

We can see that our sample conversation breaks very neatly along this top-level division: DU #1 is the goal specification, DUs 2-12 are concerned with constructing the plan, and DUs 13-15 are concerned with verifying the plan. That things turn out so neatly here is mainly an artifact of the simplicity of the goal in this problem. For more complex problems the breakdown is more often by a vertical decomposition, where a subgoal is specified and then the plan to achieve that subgoal is constructed and verified at which point the process repeats with another subgoal. Also the steps of constructing and verifying the plan even within a specified subgoal are often intermixed. [Poesio, 1991] presents a slightly different top level decomposition based on this tighter coupling.

The construction of the domain plan is based on standard plan reasoning techniques, applied to the TRAINS domain [Ferguson, 1991]. Figure 7 provides a typical plan decomposition for this problem (shipping a boxcar of oranges to Bath) in a situation such as this one in which boxcar and oranges and engine are all initially in separate locations. The actions in the plan are numbered arbitrarily for ease of reference, and the arrows represent enabling dependencies in the performance of the actions. For example Action (5), Coupling the engine to the boxcar, cannot be performed until the engine has been moved to the location of the boxcar, which in turn cannot be performed until the location of the boxcar and the engine have been identified. Of course, we may talk about these actions in any of a number of orders, not just the order of eventual execution.

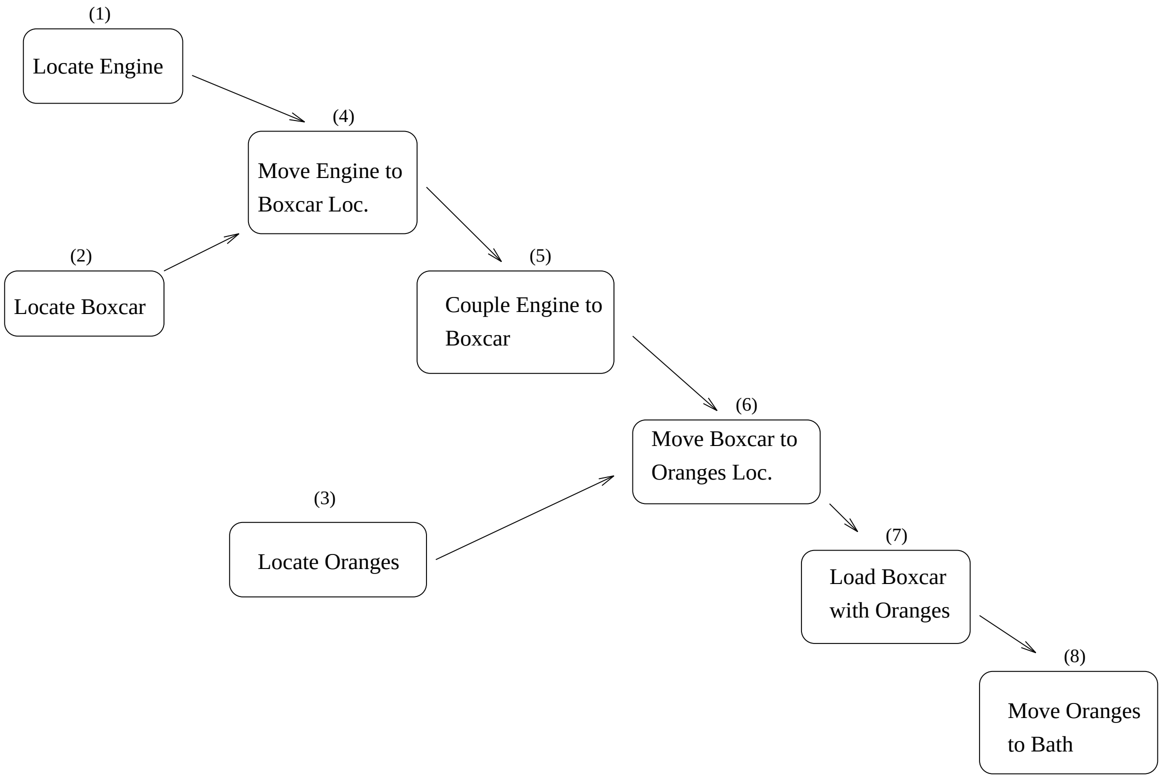

Figure 7: Domain Plan for Moving Oranges to Bath

We can track the connections between the various suggestions in the conversation rather transparently using this abstract plan recipe. The conversational references to the abstract domain plan recipe are shown in Figure 8. The numbers beneath the actions represent the DUs in which these steps are described. Thus, for example, Action (4), moving the engine to the boxcar location is mentioned in DUs 9, 10 (which is left ungrounded) and 11. Action (5) is referred to in DUs 11 and 12, specifically in UU 15.1 of DU 12. The initial TRAINS World set-up in Figure 2 is repeated here for convenience.

DU #2 mentions Action (6), specifying as well, the location of the oranges (3) but not the particular boxcar (2). As said above, whether this act is grounded or not is not clear, because M immediately goes on to ground (3) explicitly in the Q&A argumentation act consisting of DUs #3 and #4 together. Then DUs #5 & #6 ground the enablement link from Action (1) to Action (6) (shown only indirectly in Figure 7 through Actions (4) and (5)). DUs #7 & #8 ground the location of the Engine (part of (1)), and DU #9 elaborates this identification, naming the engine as well. The domain planning behind these utterances has finally gotten to the deepest points in the dependencies, and thus the first steps in actual execution. DU #9 also begins to mention Action (4), but before the destination can be established (UU 9.3 begins to suggest a destination), the repair of the Engine name, UU 10.1, occurs and this needs to be grounded. DU #10 and it's replacement #11 elaborate on DU #9, filling in (2), and completing (4). DU #11 then goes on to suggest (5). DU #12 continues the elaboration, repeating the suggestion of (5) in UU 15.1, and then going on in the next three utterances to suggest (6) (which may have already been suggested in

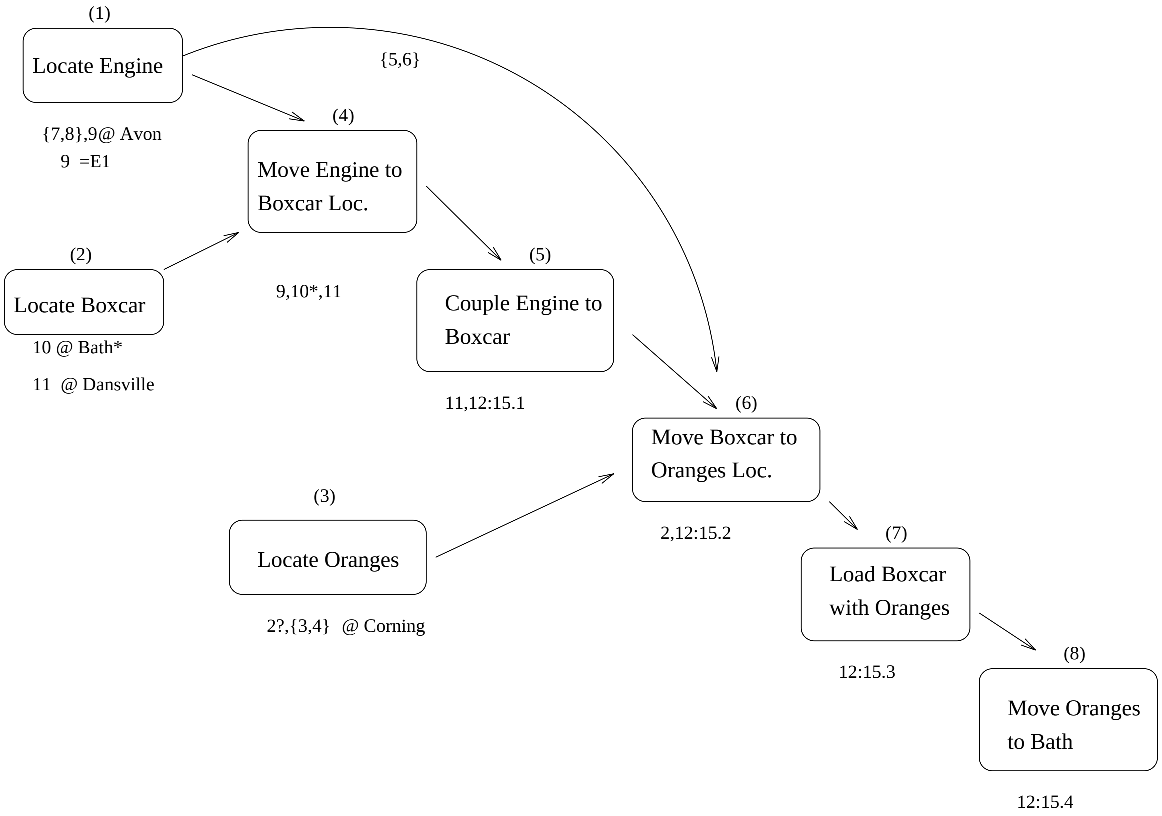

Figure 8: Domain Plan for Moving Oranges to Bath

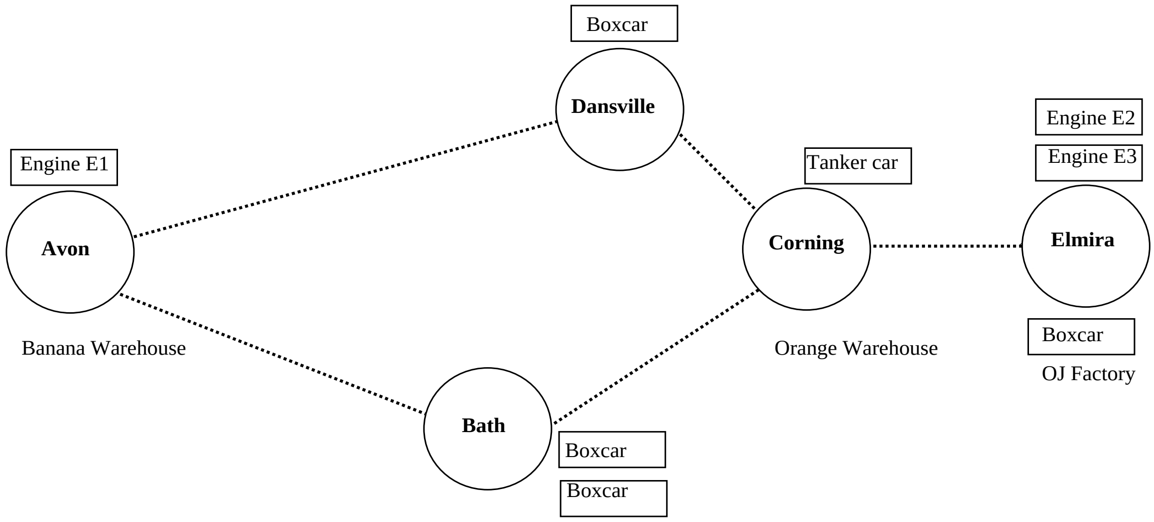

Figure 2 (reprise): Trains World Set-up for Example Conversation

(2)), (7), and (8), completing the plan construction phase.

The plan evaluation phase is begun by DU #13, which is a request to evaluate the whole plan in DUs #3 through #12, referred to by the strongly intonationally marked "THAT" in UU 17.1. This plan is to be evaluated in terms of satisfying the goal set forth in DU #1, as can be determined with access to the top level decomposition in Figure 6. The system's first reply, UU 18.1 is a calculation of the time this whole plan will take. The timings were never brought out in the conversation, but S has the information that it takes 3 hours to go from Avon to Dansville, almost no time to couple to the boxcar, 1 hour to move to Corning, 1 hour to load the oranges, and 2 hours to travel to Bath. Thus the whole plan takes 7 hours, and with a starting time of midnight that adds up to 7am as the time of plan completion. This `inform` supports the requested evaluation, and then S accepts the plan and provides the answer to the question with UU 18.3. M expresses his own approval with DU #15, and the participants now have a grounded plan which they mutually accept as satisfying the grounded goal, and the conversation is completed.

At a lower level of argumentation, we have discourse coherence cues for argumentation relations. Q&As for grounding information content are seen in DU pairs #3&#4, #5&#6, #7&#8, and #13&#14. DU #11 elaborates on the plan suggested in DU #9, and this is further elaborated in DU #12. We have many suggest & accept pairs as well, often within the scope of a single discourse unit (e.g. DUs #1, and #11).

4 Recognizing Conversation Acts

Agents participating in discourse are primarily concerned with problem solving, informational, or social goals. The process of serving these goals through discourse includes recognition of the various conversation acts. Although the recognition process varies according to the class of conversation acts, common elements emerge. For example, each recognition engine can be thought of an evidence combination device, which monitors speech input for certain sets of features. The features serve as evidence for a range of interpretations within the class of acts, but may be overridden by evidence from other features. Also, recognizers may have additional sources of evidence, such as predictions based on previous conversational state. An example of this is the state accumulated in tracking a discourse unit through a sequence of grounding acts (see Section 4.2).

4.1 Turn-Taking Acts

Recognition of turn-taking acts is highly dependent upon the social setting and discourse context. Some social settings have preallocated turns or highly formalized turn selection processes. Others have one participant who serves as an arbiter "granting" the turn to a requesting party. In casual conversation, the turn-taking process is generally determined on-line, through use of the same channel that the turn-taking system is regulating.

Two contextual notions are useful for tracking turn-taking in casual task-oriented conversations such as those in the TRAINS domain, these are the *turn* and *local initiative*. Each of these may be said to be held by one (or none) of the participants at any given time in the conversation. The turn is crucially important in recognizing turn-taking attempts, since the turn-taking acts are formulated in these terms. A take-turn attempt may only be performed by an agent without the turn, whereas a release-turn, and keep-turn may only be performed by the current turn holder. A take-turn attempt can be recognized as any speech by a non-turn-holder. The other notions are more difficult to distinguish.

Local initiative[2] can be glossed as providing the answer to the question of who has most recent discourse obligation - who is expected to speak next according to the default plans for simplest satisfaction of conversational goals. E.g. a question or request will produce a discourse obligation on the other party to respond to the request (either by satisfying it, accepting it as a responsibility to be performed later, or denying it). If there are no local obligations, the local control may be derived from the higher level goals and expectations, such as the top-level plan in Figure 6.

Certain indications signal either a keep-turn (e.g. filled pauses, lengthened words, pauses at non-constituent boundaries, and "continuing intonation") or release-turn. Local initiative is important in determining whether a neutral utterance ending is seen as a keep-turn or release-turn. Thus a question will impose a discourse obligation on the listener, and although the current speaker may retain the turn and follow the question with, for example, a clarification, the next neutral ending will be seen as a release-turn. In an analogous way, if the current speaker makes a neutral ending when she is still expected to speak, e.g. after an introduction or in the middle of a list, then this is still keeping the turn. In conversations such as those in the TRAINS corpus, where verbal acknowledgements are important for grounding the content, there is always a mild obligation on the listener to respond, and thus it takes an explicit continuation signal (either intonation or content) to prevent a neutral ending from releasing the turn.

4.2 Grounding Acts and DUs

Recognition of Grounding Acts is highly dependent on the local linguistic context. Only certain acts will be possible for an agent to perform in a given state of the conversation, and the same utterance will be interpreted differently based upon its surroundings. Section 4.2 describes a finite automaton for tracking the state of a current DU, outlining the preferred and possible acts from each state. Section 4.2 then relates some principles for recognizing particular acts in utterances, given these states.

Construction of Discourse Units

We name the agents taking part in constructing a DU as follows: the *Initiator* is the one who performs the `initiate` act to start off the DU. The other participant is called the *Responder*. Agents may take different roles in different DUs in a mixed initiative conversation. 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

[2] Roughly the same notion as *Control* in [Walker and Whittaker, 1990], although we use a more fine-grained notion of utterance types.

has been completed, or just what it would take to complete one, only certain patterns of actions are allowed. For instance, a speaker cannot acknowledge his own immediately prior utterance. He may utter something (e.g. "ok") 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. Similarly, a speaker cannot `continue` an utterance begun by another speaker. Depending on context, this will be interpreted as either an acknowledgement (e.g. if one is just completing the other's thought), a `repair` (if one is correcting to what *should* have been said), or an `initiate` of a new DU (if this is new information).

We can identify at least seven different possible states for a DU to be in. These can be distinguished by their relevant context: what acts have been performed and what is preferred to follow, as shown in Table 3.

State	Entering Act	Preferred Exiting Act
S	———	Initiate^{I}
1	Initiate^{I}	Ack^{R}
2	ReqRepair^{R}	Repair^{I}
3	Repair^{R}	Ack^{I}
4	ReqRepair^{I}	Repair^{R}
F	Ack$^{\{I,R\}}$	Initiate$^{\{I,R\}}$ (next DU)
D	Cancel$^{\{I,R\}}$	Initiate$^{\{I,R\}}$ (next DU)

Table 3: Meanings of Discourse Unit States

Acts in this table are superscripted with the initial of the agent who performs them, "I" for the *Initiator*, and "R" for the *Responder*. State S represents a DU that has not been initiated yet. State F represents one that has been grounded, though we can always add on more, as in a further acknowledgement or some sort of repair. State D represents an abandoned DU, ungrounded and ungroundable. The other states represent DUs which still need one or more utterance acts to be 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 utterance, and may request that the Responder repair, this would be state 4.

Although these states have acts which are in some sense *preferred*, any of a number of acts can follow at any given state. Table 4 shows a finite state machine which gives the possible transitions from state to state and tracks the progress of Discourse Units. 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 `initiate` (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 in one of the non-final states, then its contents should not be seen as grounded.

Next Act	In State						
	S	1	2	3	4	F	D
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	
Cancel^{I}		D	D	D	D	D	

*repair request is ignored

Table 4: DU Transition Diagram

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. 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 [Traum, 1991a]. 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[3], 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 also be analyzed in terms of the preconditions and effects of actions, as sketched in [Traum, 1991b], but having an explicit

[3]except in the obvious case after a ReqRepair (states 2 and 4)

model of the nature given here may serve to repair interactions, and make processing more efficient.

Recognizing Grounding Acts

In a situation in which there are no accessible DUs, the only possible grounding act would be an `initiate` - thus any utterance that attempted to change the mutual beliefs of the agents would be an `initiate`. `Initiate` can be recognized in other locations as an utterance which conveys new content to be mutually believed which is not a syntactic or semantic continuation or correction of an extant unacknowledged unit. A syntactic/semantic continuation of a unit which has not been acknowledged will be seen as a `continue`. The difference is this: items which are grouped together as part of the same discourse unit will be acknowledged together and those which are not (i.e. the second utterance is marked as a new `initiate` rather than a `continue`) have the option of having one but not the other acknowledged.

Acknowledgements come in three types. Backchannel responses (e.g. "okay", "uh huh") in the proper context (state 1 for the Responder or 3 for the Initiator) directly signal acknowledgement of the current DU. A paraphrase or completion of the other's sentence may be either an acknowledgement, a repair, or a repair request. Questioning intonation will signal a repair request, and acknowledgements are distinguished from repairs by having the same or expected content as opposed to a replacement or new content. Implicit acknowledgements can also be recognized by an initiation of a new DU which forms a next step in a current argumentation act, e.g. an answer acknowledges the previously initiated question.

`Repair` is any utterance which replaces any of the content of the current DU. This change may be either to the explicit content of a previous utterance or to the presuppositions. Repairs by the Initiator are often signalled by cue phrases such as "I mean", "that is", or "I'm sorry". Cancels indicate a dropping of the intention o complete the current DU. Signals include, "forget it", "never mind", or dropping an utterance part way through and starting up with something else.

4.3 Core Speech Acts

The general model that we assume for core speech act recognition is that of [Hinkelman and Allen, 1989; Hinkelman, 1990]. This work emphasized the role of surface signals in recognition of speech act type, using patterns of lexical, syntactic, and semantic features to generate a set of hypotheses about speech act type. The set of hypotheses was then further filtered, by testing for each the plausibility of inferences which it would license about relevant propositional information in context. If these inferences were in contradiction to the current knowledge state, this would be grounds for elimination. If the remaining speech act interpretations provided an inadequate basis for action, Allen-style reasoning [Allen, 1983] could be invoked. One hope was that this model would have good real time properties, and the fact that this model is sensitive to surface characteristics of an utterance makes it a good candidate for recognition in spoken dialogue, where forms may be incomplete or interrupted.

Consider for a moment the properties of multiclause speech acts. It transpires that some multiclause utterances require (unsurprisingly) individual analysis of each clause, for instance:

```
It is with great regret that I announce my resignation.
                                          (nominal perf.)
                               (Announce, perform. verb.)
(Inform, as indicated by declarative syntax and cleft.)
```

which is an actual resignation as well as an announcement and an inform act [Sadock, 1974]. (Note that "regret" is not performative.) Under some conditions the constituent clauses are blocked from speech act status, but we discuss this elsewhere [Hinkelman and Allen, 1992]. There are also data, perhaps ill-formed but nevertheless understandable, in which the speaker's intent shifts:

```
I was hoping you could please give me at least a ride.
                     (Request: imperative + please  )
                  ^^ the point of crossover.
(Inform: declarative; likely indirect request semantics. Incomplete.)
```

Here the speaker essentially slips from a very indirect request to a direct one, via a lexical item conventionally associated with aux-inversion requests. The speaker may also fail to track clause embedding well:

```
Your point has been made and we are please asking you to leave.
We want them to please, think about this child...
```

Clause-by-clause analysis enables integration of core speech act recognition into the framework of turn-taking and grounding. The surface-oriented component of speech act recognition examines each input phrase, as well as any larger constituents that syntactic analysis identifies, and attempts to build speech act interpretations for them. If core speech acts are found, (and there are defaults for most syntactic phrases), they are subjected to verification as plausible plans, as above. In the simpler clause-by-clause model it is desireable to buffer interpretations until the utterance boundary is found, to be sure that the topmost node does not block clause interpretations. The integrated model must also rely on utterance boundaries for this purpose. The final set of interpretations is passed on to DU tracking (see Section 4.2) for preservation across utterance boundaries.

Clause level tracking is useful immediately at the beginning of our example dialogue. Consider the system's view of utterance 1.1. Ignoring the initial "okay", clause level analysis yields an inform, and a declaration of a problem can be a directive act as well. The idiomatic "<person> better" is a forceful suggestion, and unifies with the directive interpretation to yield an overall suggestion act for the utterance. Testing against propositional information in context supports this conclusion.

Consider our example dialogue from the system's point of view, beginning at 3.1. Core speech act recognition gets no interpretation for the first two isolated words, though other modules may. Being declarative, the next clause is an Inform, again an Inform of a need and therefore potentially carrying the force of a Suggestion. Perhaps relative clauses, such as "where there are oranges" default to Inform, or perhaps they don't warrant any default at all. Intonationally, this utterance is a series of phrases with H phrase accents; the last is possibly a bit more elongated than the others, and fails of any conclusive fall. It is thus bounded by the next syntactic unit's beginning, and at this point the interpretation for the entire sentence is that of the first clause.

UU 3.2 and 3.3 follow without strong intonational boundaries. UU 3.2 has declarative syntax but ends in a continuation rise, and 3.3 resolves the interpretation to a confirmation question, or "Check". This interpretation is easily verified as a plausible plan, since both speaker and hearer have written copies of the information contained in the utterance.

4.4 Argumentation Acts

Argumentation Act recognition starts with reference to *Discourse Scripts* [Poesio, 1991]. These represent conventional knowledge of such things as adjacency pairs (e.g Q-A, greeting-greeting), discourse obligations, and high level conversational tasks such as those represented in Figure 6. These scripts represent background assumptions of the expected coherence relations, which can be applied to the current situation to allow recognition of probable speaker intention. First parts of discourse scripts will make their next parts conditionally relevant, for instance, the first utterance after a question will be seen as an answer unless it gives explicit signals to the contrary (e.g. a repair or follow-up question).

The main surface indications of argumentation relations are certain types of cue words. "So" is a very good signal for a summary or deduction function. These generally cue items which, while they haven't appeared explicitly in the previous conversation, are inferable from what has gone on before combined with background knowledge. Thus, the inform in DU #2 is in a summary relation to DU#1, in virtue of the decompositional plan knowledge that there is only one orange source and a boxcar is needed to carry oranges. Similarly for the check in DU #9. The "so" in DU #5 seems like it could be either a summary or just a topic progression, another common use of "so". "And" is a good signal of elaboration or continuation, e.g. UUs 15.1, 15.4.

Domain Plan knowledge will also be extremely important at this level. Often it is not necessary to know the precise relationship between two segments of conversation. As long as the conversants are attending to building up the domain plan and seeing how the contents fit togther, that is enough to get by without explicit identification of the rhetorical relationships. Knowing these relationships, however, can provide important clues to the domain plan recogniser as to how to fit a new item into the plan.

5 Related Classification Schemes

There have been quite a few previous attempts to categorize acts in discourse into different groups, however, we believe none of the previous classifications have the range of coverage that the current scheme has. Most schemes either treat only one or two of the levels we have here, or try to combine everything into a system of rankings, where one group is composed of items at a lower rank, the way grammatical phrases are composed of words. An example is the classification scheme proposed in [Sinclair and Coulthard, 1975] and later modified in [Coulthard *et al.*, 1981] and [Stenstrom, 1984]. This taxonomy is one of ranks within the same level, a level called "Discourse", with the following ranks form smallest to largest: act, move, exchange, sequence, transaction. This system corresponds most closely to the argumentation level, although the exchange rank is very similar to a DU in our terms, consisting of an initiation possibly followed by a response and feedback. Grounding and Turn-taking are not explicitly covered, although there are acts such as *acknowledge* and *reply* in [Coulthard *et al.*, 1981] and *repeat*, *backchannel*, *request acknowledgement* in [Stenstrom, 1984].

Although the levels of action we have discussed in this paper are all manifested through the same channel of spoken language, the levels represent coordination of different types of activity. Turn-taking coordinates who is in immediate control of the speaking channel and should have the attention of the participants. Grounding coordinates the state of mutual understanding on what is being contributed. Argumentation coordinates the higher discourse purposes that the agents have for engaging in the conversation. Core Speech Acts coordinate the local flow of changes in belief, intentions, and obligations.

6 Conclusion

This model of conversation takes a significant bite into the problems raised by the extremely fine-grained interactions in spoken discourse. We hope that the lower three tiers will be valid in a variety of domains, though the cues for realization of particular acts will be different. We also hope that this taxonomy will provide a fruitful basis for re-examination of the issues of high-level discourse structure.

Current work includes continued testing of the scheme against the TRAINS corpus, and implementation of the recognition ideas in Section 4. When the recognition algorithms for the four levels have been thoroughly tested, we will have good evidence that conventional literal meaning is thoroughly interlarded with intentions at many levels.

Acknowledgements

The classification scheme for conversation acts in Section 2 was devised by the first author in collaboration with James Allen. We would also like to thank Derek Gross and Shin'ya Nakajima for collecting the TRAINS dialogues and providing the initial transcriptions which have been used as data for this analysis. All three have been active participants in the continued development of the theory of conversation acts. The Rock Star example in Section 1 was suggested by Stephen P. Spackman.

References

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

[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.

[Beun, 1989] Robbert-Jan Beun, *The Recognition of Declarative Questions in Information Dialogues*, PhD thesis, Katholieke Universiteit Brabant, 1989.

[Bunt, 1989] H. C. Bunt, "Information Dialogues as Communicative Action in Relation to Partner Modelling and Information Processing," In M.M Taylor, F. Neel, and D. G. Bouwhuis, editors, *The Structure of Multimodal Dialogue*. Elsevier Science Publishers B.V., 1989.

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

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

[Coulthard *et al.*, 1981] M. Coulthard, M. Montgomery, and D. Brazil, "Developing a Description of Spoken Discourse," In M. Coulthard and M. Montgomery, editors, *Studies in Discourse Analysis*, pages 1–50. Routledge & Kegan Paul, 1981.

[Duncan and Niederehe, 1974] Starkey Duncan, Jr. and George Niederehe, "On Signalling That It's Your Turn to Speak," *Journal of Experimental Social Psychology*, 10:234–47, 1974.

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

[Ford and Thompson] Cecelia Ford and Sandra Thompson, "On Projectability in conversation: Grammar, Intonation, and Semantics," presented at the Second International Cognitive Linguistics Association Conference, August, 1991.

[Gross *et al.*, 1992] Derek Gross, James Allen, and David Traum, "The Trains 91 Dialogues," TRAINS Technical Note 92-1, Computer Science Dept. University of Rochester, to appear 1992.

[Halliday, 1961] M. A. K. Halliday, "Categories of the Theory of Grammar," *Word*, 17:241–92, 1961.

[Heritage, 1984] John Heritage, "A change-of-state token and aspects of its sequential placement," In J. M. Atkinson and J. Heritage, editors, *Sturctures of Social Action*. Cambridge University Press, 1984.

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

[Hinkelman and Allen, 1989] Elizabeth A. Hinkelman and James F. Allen, "Two Constraints on Speech Act Ambiguity," In *Proceedings ACL-89*, pages 212–219, 1989.

[Hinkelman and Allen, 1992] Elizabeth A. Hinkelman and James F. Allen, "Speech Act Interpretation without Literal Meaning," *Computational Linguistics*, 1992, forthcoming.

[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.

[Mann and Thompson, 1987] William C. Mann and Sandra A. Thompson, "Rhetorical Structure Theory: A Theory of Text Organization," Technical Report ISI/RS-87-190, USC, Information Sciences Institute, June 1987.

[McCawley, 1988] James McCawley, *The Syntactic Phenomena of English*, University of Chicago Press, 1988.

[McCawley, 1989] James D. McCawley, "Individuation in and of Syntactic Structures," In Mark Baltin and Anthony Kroch, editors, *Alternative Conceptions of Phrase Structure*, chapter 7, pages 117–138. Univeristy of Chicago Press, 1989.

[Nakajima and Allen, 1991] Shin'ya Nakajima and James F. Allen, "A Study on the Roles of Prosody in the Cooperative Dialogue," In *Research on Modelling of Speech Dialogue and its Computational Processing: Final Reports*, pages 144–152. National Science Foundation, December 1991.

[Orestrom, 1983] Bengt Orestrom, *Turn-Taking in English Conversation*, Lund Studies in English: Number 66. CWK Gleerup, 1983.

[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.

[Pierrehumbert and Hirschberg, 1990] Janet Pierrehumbert and Julia Hirschberg, "The Meaning of Intonational Contours in the Interpretation of Discourse," In P. R. Cohen, J. Morgan, and M. E. Pollack, editors, *Intentions in Communication*. MIT Press, 1990.

[Poesio, 1991] Massimo Poesio, "Expectation-based Recognition of Intentional Structure," In *Working Notes AAAI Fall Symposium on Discourse Structure in Natural Langauge Understanding and Generation*, November 1991.

[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.

[Sadock, 1974] Jerrold M Sadock, *Meaning*, Toward a Linguistic Theory of Speech Acts, 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 and Sacks, 1973] E. A. Schegloff and H. Sacks, "Opening Up Closings," *Semiotica*, 7:289–327, 1973.

[Sinclair and Coulthard, 1975] J. M. Sinclair and R. M. Coulthard, *Towards an analysis of Discourse: The English used by teachers and pupils.*, Oxford University Press, 1975.

[Stenstrom, 1984] Anna-Brita Stenstrom, *Questions and Responses*, Lund Studies in English: Number 68. Lund : CWK Gleerup, 1984.

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

[Traum, 1991b] David R. Traum, "Towards A Computational Theory of Grounding in Natural Language Conversation," Technical Report 401, Computer Science Dept. University of Rochester, October 1991.

[Walker and Whittaker, 1990] Marilyn Walker and Steve Whittaker, "Mixed Initiative in Dialogue: An Investigation into Discourse Segmentation," In *Proceedings ACL-90*, pages 70–78, 1990.

[Yngve, 1970] Victor H. Yngve, "On Getting A Word In Edgewise," In *Papers from the Sixth Regional Meeting*, pages 567–78. Chicago Linguistic Society, 1970.