# Towards a Uniform Model of Conversational Context

Massimo Poesio
University of Edinburgh
Centre for Cognitive Science

David R. Traum
TECFA, FPSE
Universite de Geneve

February 7, 1996

**Abstract**

Natural language dialogue systems require contextual information for a variety of processing functions, including reference resolution, speech act recognition, and dialogue management. While much has been written about individual contextual problems, many of the proposed representations are mutually incompatible, not usable by an agent involved in a conversation, or both.

We present a model of context that begins to reconcile in a general and systematic fashion the differences between the discourse context models used for reference resolution, conversation act recognition, and dialogue management which can be used by a system dealing with conversations. Starting from the technical solutions adopted in DRT, we show first how to obtain a discourse model that, while preserving DRT's basic ideas about referential accessibility, includes information about the occurrence of speech acts and their relations. Next, we show how the information about speech acts can be used to formalize the basic ideas of Grosz and Sidner's model of discourse structure. Finally, we extend this model to incorporate an account of the grounding process.

[1] Address: 2 Buccleuch Place, Edinburgh EH8 9LW, Scotland, UK. Phone: +44 31 650 698, Fax: +44 31 650 4587, Email: poesio@cogsci.ed.ac.uk

[2] Address: , 9 Route de Drize, Bat D, CH-1227 Carouge, Switzerland. Telephone Number: +41 22 705 9696, Fax: +41 22 342 8924, Email: David.Traum@tecfa.unige.ch

# 1 Motivations

A vast amount of context is needed to participate in and understand even the simplest conversations. This context includes a large store of (assumed) common knowledge between the participants: about language itself, about conventions of social interaction (e.g., that a question should be responded to with an answer), about the domain of discourse (e.g., that trains move along rail lines, not through the air), as well as more local information shared between just the participants themselves. The individual mental states of the agents (e.g., beliefs, and intentions) are an important element of the context of conversation, as well. And previous utterances in the conversation itself will be important for recognizing and performing appropriate future utterances.

This paper is about the representation of discourse context for systems that engage in naturally occurring conversations. The motivation is provided by work on the TRAINS project at the University of Rochester, one of whose aims is the development of a planning assistant able to engage with its user in spoken conversations in the domain of transportation by train [Allen *et al.*, 1995].

The TRAINS prototype must perform several kinds of linguistic activities that depend on a context, including reference resolution, conversation action recognition and dialogue management. In developing the prototypes, we were confronted with the problem of the poor fit between, on the one hand, the linguistically motivated theories of context developed to account for the semantics of anaphora (e.g., [Kamp and Reyle, 1993]); on the other hand, the models of context proposed for action recognition and dialogue management, whose emphasis is on capturing the effects of speech acts on the beliefs, intentions, and obligations of the participating agents [Allen, 1983; Carberry, 1990; Cohen and Levesque, 1990; Perrault, 1990; Traum and Allen, 1994]. Both traditions resulted in very detailed proposals about context and context update;[1] but the resulting models of context differed significantly. And it is not possible to simply adopt one or the other model; while the linguistically motivated theories of context integrate well with current theories of semantic interpretation, their relation with current work on planning and plan recognition is less clear; the opposite is true of theories of context based on actions and their effects. Unfortunately, a complete language understanding system needs to accomplish all of these tasks.

For example, ever since Grosz's dissertation [Grosz, 1977] it has been known that subordination relations between discourse goals affect the structure of global

[1] See, e.g., [Kamp and Reyle, 1993] for the details of a semantic treatment of anaphoric reference, and [Cohen and Levesque, 1990] about the effect of speech acts on the mental state of agents.

focus, hence the `pragmatic' accessibility of referents. A model of context that contains information about the intentions of the conversational participants is therefore advantageous for reference resolution purposes; but the models of context developed for intention recognition and dialogue control purposes do not provide enough information about which referents are available at each point in the conversation. On the other hand, the discourse models proposed for reference resolution purposes typically `abstract away' from aspects of the input utterance that are not relevant for reference resolution: thus utterances of acknowledgments, such as *okay*, and other discourse particles are typically ignored. But these aspects of the utterance are crucial for speech act recognition and for tracking the state of the conversation for dialogue management purposes. Thus, in a complete NL conversation system, if the model of context used for reference resolution does not provide this information, the dialogue manager either has to do without it, or it has to recover this information on its own, typically reduplicating some effort.[2]

Providing such a unified representation is not simply a matter of adopting a common representation formalism. The difference in emphasis in the research traditions working on the sub-problems have led to some real differences in interpretation of some of the overlapping areas. A case in point is the use of *common ground*. Most work on reference resolution makes the assumption that felicitous reference requires mutual knowledge [Clark and Marshall, 1981], and that, further, everything that is uttered in a shared situation immediately becomes a part of the common ground and, hence, is available for reference. This assumption is important so that, *e.g.,* a speaker can refer to an object mentioned in the immediately preceding utterance.

As shown in [Clark and Wilkes-Gibbs, 1986; Clark and Schaefer, 1989; Traum and Hinkelman, 1992], however, this assumption of immediate addition to common ground is an idealization that is not generally realistic in actual conversation: utterances by one party must be recognized and acknowledged by the other before becoming part of the (assumed) common ground. It is important for a system engaging in extended dialogues (rather than just iterative question answering exchanges) to have an awareness of the grounding process, so that acknowledgments and repairs can be both recognized and produced when appropriate.

It is not simply a case of one tradition "being right" in this matter. Consider the following utterance in a conversation betweeen agents **A** and **B**:

(1) There is a tanker at Avon

[2] Part of the problem is that the linguistically-oriented models of context were developed to deal with *text* rather than dialog.

In a conversation, this could easily be followed by any of the following (or numerous other possibilities):

| | | |
|---|---|---|
| 2a | A | `Let's pick it up.` |
| 2b | B | `Let's pick it up.` |
| 2c | B | `Uh huh.` |
| 2d | B | `There's a what?` |
| 2e | A | `I mean a boxcar.` |
| 2f | A | `Did you hear me?` |

The first example shows that the object mentioned in (1) (the tanker at Avon) must be accessible to *anchor* the referring expression *it* in 2a. The second example shows the other speaker can also make use of the mentioned entity without first grounding it. Thus, if common ground is a precondition for felicitous definite reference[3], the entity must be part of the common ground. On the other hand, the third example shows a potential response to (1) which would have the effect of grounding the content of that utterance. If it were already part of the common ground, then there woudl be no reason for B to make such an utterance. More compellingly, a response such as example 2d would show that the utterance (in particular the mentioned object) is *not* understood, and thus not part of the common ground. The possibility of such a response means that one cannot safely assume that the content of (1) is immediately part of the common ground. 2e shows a repair by A, showing that even if B might have assumed commonality, such commonality would have been wrong, because A mis-spoke. Finally, 2f shows that speakers do not always assume that their utterances have been heard and understood without feedback. Since all of these sorts of examples occur frequently in task-oriented conversation, a comprehensive system must be able to represent the effects of each on the current context, as well as decide if and when it is appropriate to perform one or another.

We outline a model of context that tries to reconcile in a general and systematic fashion the differences between the discourse models used for reference resolution, conversation act recognition, and dialogue management in a system dealing with conversations. Our model is meant to be one that an agent can use as an internal, on-line representation of context while engaging in conversation. This work is based on merging some of the ideas from two recent works within these different traditions, namely [Poesio, 1994] and [Traum, 1994]; yet, we believe that the result is in a sense more than the sum of its parts. The resulting model is novel in several respects. In the rest of this paper, we gradually develop this model, starting from

[3] Clark and Marshall [1981] show with a series of examples that mere availability is not enough for felicitous definite reference, when a grounded alternative is also possible.

traditional basics of the reference resolution tradition and progressively adding in other features of context from other traditions. As the ideas from both of these models of context were implemented as modules within the same conversation system (using conversion languages for inter-module communication), we have reasons for optimism, both for the soundness of the contextual model, and for the utility of the result.

We concentrated our effort on providing a clean conceptual interface between the models that resulted from our joint work on TRAINS. In doing so, we emphasized clarity over efficiency. Other factors that led the development of this model include (i) the desire to make this work compatible with the results of the linguistic work on the semantics of discourse, and (ii) our view of discourse interpretation as a form of defeasible inference, hence, our attempt to avoid introducing in our theory extra-logical mechanisms. Both aspects will be discussed in the paper.

## 2 A Minimal Representation of Context: Discourse Representation Theory

From the point of view of reference resolution, the most fundamental property of utterances is that they UPDATE the existing context, i.e., they augment it with new discourse referents and / or new propositional information [Karttunen, 1976; Webber, 1979]. For example, an utterance of *There is an engine at Avon* has the effect of making a new DISCOURSE REFERENT available for subsequent pronominalization—the utterance can be followed by the utterance *The boxcar is also at Avon and is hooked to it*, where *it* refers back to *an engine*. The minimal requirement for a formalization of context, in a reference resolution perspective, is that it account for this `update potential' of utterances.

First order logic does not satisfy this requirement: the `natural' formalization of *There is an engine at Avon* is $(\exists\, x\, \mathbf{engine}(x) \wedge \mathbf{at\text{-}avon}(x))$; the variable $x$ is boound within the scope of the existential quantifier and is not available for subsequent reference (i.e., conjoining this expression with something like $P(x)$ would not have the desired result of making $x$ bound by the existential quantifier). In Grosz and Sidner's theory, context update is modeled via the focus spaces stack mechanism, i.e., outside the logic used to represent the meaning of statements; in this way, however, the effect of anaphoric links on the truth conditions of a sentence cannot be accounted for. The formalisms proposed in Artificial Intelligence to model intention recognition are simply not concerned with this issue. But modeling update is the *raison d'etre* of Discourse Representation Theory (DRT) [Kamp, 1981; Kamp and Reyle, 1993] and related `dynamic' theories such as Dynamic Predicate Logic [Groenendijk and Stokhof, 1991]. It seems appropriate therefore to use such theories as the starting point for our formalization of context. In particular, we will start from DRT, which among them is the one with the more conventional semantics.

DRT can be summarized as the claim that the model of a discourse—for reference purposes, at least—is a DISCOURSE REPRESENTATION STRUCTURE (DRS): a pair consisting of a set of DISCOURSE REFERENTS and a set of CONDITIONS (propositions) that is typically represented in `box' fashion. For example, the sentence in (2) is represented as in (3).

(2) Engine E2 is at Avon.

(3)
$$\boxed{\begin{array}{l} x\ w \\ \hline \textbf{engine E2}(x) \\ \textbf{Avon}(w) \\ \textbf{at}(x,w) \end{array}}$$

A DRS is best thought of as a database including discourse referents in addition to facts, and that can include additional, subordinate databases. At the same time, DRSs can be thought of as logical expressions: (3) is true wrt a model M and a situation (or world) $s$ if there is a way of assigning objects in $s$ to the discourse referents $x$ and $w$ such that all of the conditions in the box are true of these objects in $s$.

We use the term ROOT DRS to indicate the DRS that represents the whole common ground, i.e., that corresponds to what is usually called the `model of discourse'. The root DRS is updated after every sentence by a DRT CONSTRUCTION ALGORITHM whose first step is to add to the previous discourse model an `uninterpreted condition' that provides information about the syntactic structure of the sentence and the interpretation of its lexical items, followed by the application of rewrite rules called CONSTRUCTION RULES which may add new conditions or new referents, or `flesh out' the interpretation of the uninterpreted condition. Updating the model of discourse in (3) with the content of sentence *The boxcar is also at Avon and is hooked to it* yields the `extended' DRS in (5) for the discourse in (4). Note that two new discourse referents have been introduced, $y$ and $u$, for the definite *the boxcar* and the pronoun *it* respectively. The `box' obtained by the construction algorithm also contains the discourse referents introduced by the first sentence, which are thus `accessible' for reference purposes.[4]

[4]The DRT construction algorithm will be discussed in more detail below. For a detailed discussion, see [Kamp and Reyle, 1993].

(4) Engine E2 is at Avon. The boxcar is also at Avon and is hooked to it.

(5)

| *x w y u* |
|---|
| **engine E2**($x$) |
| **Avon**($w$) |
| **at**($x$,$w$) |
| |
| **boxcar**($y$) |
| **at**($y$,$w$) |
| **hooked-to**($y$,$u$) |
| $u = x$ |

Many recent theories of pronoun interpretation, definite description interpretation, and VP ellipsis resolution have found it convenient to adopt a formal discourse model such as DRT, because of its explicitness. The basic version of DRT, however, concentrates on representing semantic information, and abstracts away from all information of a pragmatic nature, including information that is needed for reference resolution purposes such as the fact that a certain speech act occurred at time $t$, or that agent $a$ has certain beliefs and intentions.[5]

# 3 Context and the Discourse Situation

The one aspect in common among existing theories of reference resolution, intention recognition, and dialogue management is the assumption that context includes information about SPEECH ACTS (or intentions) and the way they are arranged by subordination relations. In a system like TRAINS-93, for example, this information is needed by all modules of that make use of context: the reference resolution module (to build up segments of discourse and implement pragmatic constraints on accessibility), the speech act recognition module of course, and the dialogue manager (to decide what to do next).

We intend to show that the context-description tools used in DRT can be adapted to the purpose of representing a context containing this information.[6] Assuming that the context includes information about speech acts means taking a different position from DRT with respect to what the common ground is, however. In DRT, the discourse model captures the truth conditions of a text; but the fact that speaker

[5] This latter concern is addressed in [Asher, 1986; Kamp, 1990].

[6] The model of context developed by Asher, Lascarides, and Oberlander more or less in parallel with our work on TRAINS (see, e.g., [**?**]), while similar in many respects to the model discussed here, crucially differs from it on this point. Asher *et al.* are concerned with texts rather than conversations; hence, they adopt a theory of discourse structure which is not based on speech acts.

A told B that P is not part of the truth conditions of a sentence. Including speech acts in the discourse model amounts to a shift from a model of sentence *meaning* to a model of sentence *use*.

We will use as the conceptual basis for the unification between DRT and speech act-based models developed for user modeling and dialogue management an idea proposed in Situation Semantics [Barwise and Perry, 1983; Devlin, 1991]. In Situation Semantics, the common ground is assumed to consist of shared information about the DISCOURSE SITUATION, which is the situation that the participants of a conversation find themselves in. The discourse situation includes the actions the agents have been performing, as well as aspects of their mental states, including their beliefs and intentions. The discourse situation also includes information about one or more DESCRIBED SITUATIONs, the situations that are the topic of the conversation. The discourse situation and the described situation(s) need not be the same. In the case of the TRAINS conversations, for example, the described situation is typically a TRAINS domain plan that the two participants are elaborating, whereas the discourse situation consists of information about what has been happening during the conversation itself.

We would like to preserve as much of DRT as possible, and simply reinterpret the DRT representation we are constructing. In other words, we would like to use the root DRS as a representation of the discourse situation, whose conditions assert facts about the occurrence of speech acts; the content of speech acts would then characterize the assertions made about one or more described situations. This is essentially the way we are going to proceed; however, issues of both a conceptual and a technical nature have to be addressed when doing so. At the very least, it is necessary to make sure that our model of context still makes discourse referents accessible for anaphoric reference; in order to ensure this property, however, some care is needed. For example, under the standard semantics for DRT, in a representation as in (7) for the illustrative mini-dialog in (6) between speakers A and B, the discourse referent *x* `evoked' by the NP *engine E2* would not be accessible to the pronoun *it* in the next sentence, represented in (7) as the discourse referent *u*.

| | | |
|---|---|---|
| | A: | Engine E2 is at Avon. |
| (6) | B: | The boxcar is also at Avon and is hooked to *it*. |

(7)
$$\begin{array}{|l|}\hline \textbf{tell}(A,B, \begin{array}{|l|}\hline x\ w \\ \hline \textbf{engine}(x) \\ \textbf{Avon}(w) \\ \textbf{at}(x,w) \\ \hline \end{array}) \\ \\ \textbf{tell}(B,A, \begin{array}{|l|}\hline y\ u \\ \hline \textbf{boxcar}(y) \\ \textbf{at}(y,w) \\ \textbf{hooked-to}(y,u) \\ u = x \\ \hline \end{array}) \\ \hline \end{array}$$

A more general problem with (7) is that it doesn't say that both assertions are `about' the same situation. `Vanilla' DRT allows us to specify the semantics of predicates like **tell**, one of whose arguments is a DRS in either of two ways. The first possibility is to treat **tell** as an extensional predicate: we could, that is, let the embedded DRS be evaluated with respect to the same situation in which the condition asserting the occurrence of the telling event is evaluated. But in this way we would `mix' facts about what's going on in the situation in which A and B are speaking, with facts about the situation they are talking about. Under this interpretation, in fact, (7) would assert of a single situation that in that situation, *x* is an engine and is at Avon, *y* is a boxcar, is at Avon and is hooked to *x*, and also that in that same situation, A tells B that *x* is at Avon, and B tells A that *y* is at Avon and hooked to *x*. This interpretation is often unproblematic, but in conversations the speakers may be talking about situations distinct from the one that includes the discourse situation. In the TRAINS conversations,[7] for example, assertions such as those in (6) might not be describing the state of the world which A and B are currently in, but, rather, a possible state of the world in the future, resulting from a few steps of the plan that A and B are trying to agree upon.

Alternatively, **tell** could be treated as an opaque predicate like **believe**, i.e., we could require the contents of the DRSs serving as third argument of a **tell** relation to be evaluated at a situation determined by the modality and its first two arguments. This solution is not problem-free either, though, since conversations can be about more than one situation or possible world. For example, the participants in the TRAINS conversations may (and typically, do) discuss alternative ways of

[7] Transcribed in [Gross *et al.*, 1993].

achieving the goal, each of which corresponds to a different possible situation.

A solution to this problem was proposed in [Poesio, 1994], where a version of DRT that addresses these issues and is based on the hypotheses about the common ground made in Situation Semantics was proposed, called CONVERSATION REPRESENTATION THEORY (CRT). CRT makes use of the same basic representational tools as DRT, but the Root DRS is taken to consist of information about the occurrence of speech acts (or, as we will call them here, CONVERSATIONAL EVENTS), and each utterance introduces two situations: a conversational event[8] and a described situation. A CRT representation of the dialog in (6) is shown in (8).

$$
(8)\quad \begin{array}{|l|}\hline ce1\ t1\ ce2\ t2\ s \\ \hline ce1 : \mathbf{tell}(A,B,s : \begin{array}{|l|}\hline x\ w \\ \hline \mathbf{engine}(x) \\ \mathbf{Avon}(w) \\ \mathbf{at}(x,w) \\ \hline \end{array}) \\ ce1\ @\ t1 \\ ce2 : \mathbf{tell}(B,A,s : \begin{array}{|l|}\hline y\ u \\ \hline \mathbf{boxcar}(y) \\ \mathbf{at}(y,w) \\ \mathbf{hooked\text{-}to}(y,u) \\ u = x \\ \hline \end{array}) \\ ce2\ @\ t2 \\ \hline \end{array}
$$

The DRS in (8) represents a discourse situation in which two conversational events occurred, *ce1* and *ce2*, both of which are about the described situation *s*. The first two conditions assert that *ce1* is an event of A telling B that the described situation *s* includes an engine located at Avon, and that *ce1* took place at time *t1*. The third and fourth condition assert that *ce2* is an event, taking place at time *t2*, of B telling A that the described situation *s* also contains a boxcar *y*, also in Avon, and hooked to *u*.

The semantics of CRT is based on that for Episodic Logic, a version of situation theory developed by Hwang and Schubert [1993]. The important characteristic of Episodic Logic, for our purposes, is that it assumes a universe containing objects called SITUATIONS, which can be thought of as `small possible worlds'. The lan-

[8] In Situation Semantics, events are considered a special type of situation.

guage can make assertions about these objects—most importantly, it can assert that something is true at a certain situation; and propositions are taken to be functions from situations to truth values.

The language of Episodic Logic is augmented in CRT with two new kinds of expressions: DRSs, and expressions of the form $s : \phi$. Both of these expressions denote propositional objects. A DRS is interpreted as a simultaneous existential quantification (i.e., $\boxed{\begin{array}{l} x\ y \\ \hline \Phi \\ \Psi \end{array}}$ is treated as a function which is true of a situation $s$ if there exist a $x$ and a $y$ in $s$ such that $\Phi$ and $\Psi$ are true of $s$. In first approximation, then, the expression $s$ : K, then, asserts that a situation $s$ is of the type specified by the DRS K.

In fact, expressions of the form $s$ : K also have the effect of `switching' the evaluation of the variables inside K to a new assignment associated with the situation $s$. Expressions in CRT are assigned a value with respect to a situation $s$, a variable assignment $g$, and a set of CASES $C$, one for each situation. An expression of the form $s$ : K asserts that $s$ is of the type specified by K, and furthermore, it shifts the parameters of evaluation so that the value of the discourse markers occurring in K is provided by the case associated with the value of $s$, i.e., C($[\![s]\!]$) (where $[\![s]\!]$ is the semantic value of $s$ with respect to the original variable assignment and the original cases C).

The solution to the problems with (7) discussed above is based on the fact that all speech acts involve a described situation. The fact that the content of both *ce1* and *ce2* is evaluated with respect to the same situation $s$ ensures that the DRS in the complement of *ce2* is evaluated with respect to the same case that is used to evaluate the expressions in the complement of *ce1*, *e.g.*, that $x$ is accessible from within the complement of *ce2*. Conditions of the form $s$ : K resolve both the problem of specifying the situation of evaluation for the propositional complement of locutionary acts, and the problem of guarantying that discourse referents introduced by one conversational event are accessible from the next.[9]

The process of updating the existing discourse model when a new utterance is made is initiated by CONVERSATIONAL EVENT GENERATION RULES that update the existing model of the discourse situation by adding two new discourse referents (a conversational event and a time) and two conditions, one asserting the

[9]Except for the shift in variable assignment that ensures accessibility, expressions of the form $s$ : K do the same work as the `Austinian Propositions' of Situation Theory [Barwise and Perry, 1983] and of expressions of the form [$\phi$ * s] in Episodic Logic [Hwang and Schubert, 1993].

occurrence of the appropriate speech act, the other recording the time at which it occurred. For the moment being, we will assume that this initial conversational event is a locutionary act[10] (one of **tell**, **ask**, or **instruct**); we will come back on this issue below, when discussing the interpretation process. The process of DRS construction can then proceed, except that the whole process takes place within the DRS that is the complement of the locutionary act. DRSs like (8) are the final result of this process.[11]

# 4 Representing Pragmatic Information

Two kinds of pragmatic information play a fundamental role in all processes we intend to model: information about (certain) mental states of the agents participating in the conversation, and information about the occurrence of speech acts. We begin our discussion of the information of CRT++ by spelling out our assumptions about how this kind of information is captured in our model of the discourse situation.

## 4.1 Mental States

Facts about the mental states of agents play an important role in speech act recognition, reference resolution, and dialog management. We assume that these facts about the (mutually known) intentions, beliefs, and obligations of the participants in a conversation can be characterized as states that are part of the discourse situation. These states can be represented in the language introduced in the previous section by conditions of the form $s$ : K. For example, the fact that A intends boxcar $y$ to be at Bath in some situation $s'$ which extends $s$ (and therefore `inherits' all the individuals that occur in $s$) can be thought of as a state *i1* holding at time *tj*. The occurrence of this state in a discourse situation can be represented by including in

[10] This term was introduced by Austin [Austin, 1962], to mean the utterance of a sentence (in context) with particular sense and reference. We use it here, meaning much the same thing as others (e.g., [Allen, 1983]) mean by SURFACE SPEECH ACT.

[11] Whereas the third argument of the locutionary act **tell** is a proposition of the form $s$ : K, the third argument of the locutionary act **ask** is a QUESTION–the denotation of expressions of the form (s?K)—and the third argument of the locutionary act **instruct** is a situation type, representing an action to be performed. Several ways of specifying what kind of semantic object a question is have been proposed in the literature, among which the best known are the proposals of Karttunen [Karttunen, 1977] and Groenendijk and Stokhof [Groenendijk and Stokhof, 1984]. One possibility is to take the denotation of (s?K) to be a function from situations to a partition of the set of situations. In the case of a yes-no question, for example, the function could map situations into one of two sets: those which would support an answer of `yes' to the question, and those that would support an answer of `no'.

the Root DRS a condition that expresses the presence in the common ground of an intention *i1* of A, as follows:

$$(9)\qquad \begin{array}{|l|} \hline \mathit{ce1}\ \mathit{t1}\ \mathit{ce2}\ \mathit{t2}\ s \ldots s'\,\mathit{i1}\ \mathit{tj} \\ \hline \ldots \\ \ldots \\ s \sqsubseteq s' \\ \mathit{i1} : \mathbf{intend}(A, s' : \boxed{\mathbf{at}(y,\mathit{Bath})}) \\ \mathit{i1}\ @\ \mathit{tj} \\ \\ \hline \end{array}$$

Some properties of mental states follow from the fact that states are just one kind of situation; such properties include, for example, `downward persistence' properties, i.e., the fact that if agent A is in a state MS, and if MS spans the temporal interval I, then A is in that state at all intervals $I'$ such that $I'$ is contained in I. These basic properties should be complemented by axiomatizations of the relevant states; we will not do so here. The reader can assume her/his own favourite formalization of mental attitudes, of which there are many around (e.g., [Cohen and Levesque, 1990] or [Konolige and Pollack, 1993b], for intention). We would just like to emphasize that the facts represented as conditions in the root DRS correspond to *mutually known* facts in other theories; e.g., a condition of the form **believe**($A$,$\Phi$) in the root DRS corresponds to a fact of the form **bmb**($A$,$B$,**believe**($A,\Phi$)) in Cohen and Levesque's [1990] formalism.[12]

## 4.2 Speech Acts

Once the discourse situation has been updated through use of the conversational event generation rules, as described in Section 3, above, there follows a process of illocutionary act recognition: an attempt to determine what was actually done by the speaker in performing the utterance. Again, context plays a crucial role in this process. As described by Austin [1962] and others, utterance of the same sentence in different contexts may be a performance of radically different actions. For instance, in the example developed above, the conversational event *ce1*, from (8), might be any of an **inform** of the engine's location, a **check** to make sure that the agents agree about this fact, a **suggestion** to use the engine in a developing domain plan, etc.

[12] Assuming here that we are representing the discourse situation from A's point of view. If we we modelling from B's point of view, this would be equivalent to **bmb**($B$,$A$,**believe**($A,\Phi$)).

Relationships to previous utterances, and hypotheses about the speaker's mental state (including beliefs, local and global goals, and intentions) will be crucial in forming and evaluating hypotheses about which illocutionary acts have been performed. For example, given the above intention *i1*, from (9), it is probable[13] that the **suggestion** interpretation of *ce1* was meant, since locating an engine is a precondition to the action of moving a boxcar (which will have the intended effect of the boxcar being in the destination).

Illocutionary act recognition algorithms based on those discussed in [Allen and Perrault, 1980; Hinkelman, 1990; Traum and Hinkelman, 1992] have been developed in the TRAINS System [Allen *et al.*, 1995]. Following the implemented system, we adopt here the multi-level `Conversation Acts' theory, presented in [Traum and Hinkelman, 1992]. This theory maintains the classical illocutionary acts (e.g., **suggestion**) as one level (called CORE SPEECH ACTS) which are now, however multi-agent collaborative achievements, taking on their full effect only after they have been *grounded* (see Section 6, below). The theory also includes levels of action for discourse relations (called argumentation acts), grounding, and turn-taking.[14]

As illocutionary act recognition is performed, the recognized acts are also added to the discourse situation. We think of these core speech act(s) as being *generated* by the utterance in the sense of Goldman [1970]. Assuming the **suggestion** interpretation of the locutionary act *ce1* in (8), the information acquired via the speech act recognition process can be represented as in (10).

(10)
$$\begin{array}{|l|}\hline x \ldots pl\ ce1\ sug1 \\ \hline \ldots \\ sug1:\ \mathbf{suggest}(A, \boxed{\mathbf{use}(\{A,B\},x,pl)}) \\ \mathbf{generate}(ce1,sug1) \\ \hline \end{array}$$

This means that (one of) the actions performed in the utterance of *ce1* is a suggestion that the agents use *x* in their ongoing domain plan, *pl*. This will have the further (perlocutionary) effect of B trying to incorporate the use of this object into his idea of the already developing plan, which might necessitate further inference of what might have been implicated by *ce1*. For this example, these might include constraints such that *x* is the engine of a planned move-boxcar action with destination *Bath*. These further IMPLICATURES will also be added into the DRS along

[13] Depending also on other aspects of the mental state and preceding discourse.
[14] A similar theory was also presented by Novick [1988].

with the suggestion, as shown in (11).

(11)
$$\begin{array}{|l|}\hline x\ y \ldots pl\ sug1 \\ \hline \ldots \\ \textbf{implicates}(sug1,pl : \begin{array}{|l|}\hline m \\ \hline \textbf{move-boxcar}(m) \\ \textbf{boxcar}(m,y) \\ \textbf{engine}(m,x) \\ \textbf{dest}(m,Bath) \\ \hline \end{array}\ ) \\ \hline \end{array}$$

### 4.3 Other Forms of Pragmatic Information

In the rest of the paper, we will assume that all properties of the discourse situation can be expressed as properties of some state or event included in the discourse situation. In the next section we will discuss how we propose to capture facts about the structure of discourse. In addition, facts about the interactional state of the dialogue can be represented in a similar manner, such as that a particular speaker has the turn or initiative at a given time.

## 5 Discourse Structure

What makes pragmatic information about speech acts so important for reference resolution purposes is the correlation between the intentional structure and the attentional state or, more in general, the effect of discourse subordination on pragmatic accessibility [Grosz, 1977; Reichman, 1985; Grosz and Sidner, 1986; Fox, 1987]. The first task to which we intend to apply the logic and representational tools introduced in the previous sections is that of reformulating Grosz and Sidner's theory of discourse structure. What we hope to show is that this reformulation largely preserves Grosz and Sidner's hypotheses without introducing extraneous mechanisms such as a focus stack. This is important because allows us to formalize the inferences involved in pushing and popping focus spaces or in moving the focus of attention within the same default reasoning mechanisms that we use to formalize all other forms of discourse interpretation.

As it is well-known, Grosz and Sidner claim that a discourse model involves three distinct, but interrelated components: an INTENTIONAL STRUCTURE consisting of the goals of the interlocutors and the relations among these, an ATTENTIONAL STATE consisting of the topics of conversation which are salient at a given point in the discourse, and a LINGUISTIC STRUCTURE which consists of syntactic information arranged in discourse segments.

As seen in the previous section, we assume that each utterance results in the generation of an illocutionary act. The process of speech act recognition does not simply recognize such illocutionary acts; it also determines relations between a new illocutionary act and some of the previously existing acts in the discourse. In the simpler case, these relations will just be of dominance and precedence; the discourse situation may also include more general rhetorical relations such as **elaboration** or **explanation** [Mann and Thompson, 1987], that we assume to represent a particular form of speech act, called ARGUMENTATION ACTS [Traum and Hinkelman, 1992]. The illocutionary acts and the relations among them correspond to the first component of Grosz and Sidner's model, the intentional structure. For example, the interpretation of the two utterances *Engine E2 is at Avon. It is attached to the boxcar* in which the second utterance constitutes an elaboration of the first would result in the following discourse situation, in which the speech act *ce2* elaborates the speech act *ce1*:

(12)

<table>
<tr><td colspan="2">ce1</td></tr>
<tr><td>ce1 : inform(A,B,s :</td><td><table><tr><td>x</td></tr><tr><td>engine_e2(x)<br>at(x,avon)</td></tr></table>)</td></tr>
<tr><td>ce2 : inform(A,B,s :</td><td><table><tr><td>u y</td></tr><tr><td>u = x boxcar(y)<br>attached-to(u,y)</td></tr></table>)</td></tr>
<tr><td colspan="2">aa1 : elaborate(A,ce1,ce2)</td></tr>
</table>

The second aspect of Grosz and Sidner's model, the linguistic structure, follows from our assumptions about events. We assume that a basic property of the way humans interpret events is that events that are causally or otherwise related are taken to form a `story' or, as we will call it here, a THREAD [Nakhimovsky, 1988; Webber, 1988; Kameyama *et al.*, 1993]. A thread is a special kind of situation, a `plural' situation such that each event in the situation has a `next' and `previous' event, and over which a `current-event' function can be defined; we use the subsituation relation $\sqsubseteq$ to denote inclusion of event in a thread [Barwise and Perry, 1983; Poesio, 1994].

We hypothesize that the grouping of conversational events into DISCOURSE SEGMENTS is just an instance of this more general phenomenon of the grouping of events into threads. Once a connection between an illocutionary act and some discourse purpose is established, the illocutionary act is naturally taken to be part of the thread associated with that purpose. We will use the term CONVERSATIONAL THREADS for threads of conversational events when we want to distinguish between this `technical' notion of discourse segment and the intuitive notion. As we will see below, `discourse segments'—introduced to account for reference facts—are only one type of conversational threads. We will discuss another form of conversational thread in Section 6.

We will represent the relation between the conversational event *sug1* in (10) and its conversational thread (let's call it *ct1*) as follows:

(13)
$$\begin{array}{|l|}\hline \ldots \\ \hline \ldots \\ sug1 \sqsubseteq ct1 \\ \hline \end{array}$$

A conversational event may either become part of one of the existing conversational threads or originate a new one. We currently assume that all core speech acts generated by a conversational event become part of the same minimal discourse segment.

Finally, and most importantly, we propose that the properties of the third component of Grosz and Sidner's model of discourse, the attentional state, that Grosz and Sidner specify by means of the focus space stack, can be made to follow from our proposed formalization of speech acts—according to which each speech act expresses a statement about a described situation—and from the properties of situations. Consider the example in (12). The third argument of the speech act *ce1* specifies new constituents of a situation and new facts about it. (E.g., that it contains the denotation of the discourse referents $x$ and that this object is engine E2.) It should be easy to see that this third argument, besides specifying the propositional content of the speech act, also specifies how that speech act affects a `focus space'—the situation $s$. The next speech act, *ce2*, augments the specification of the focus space / situation $s$ by adding the new discourse referents $u$ and $y$ and by specifing that $y$ is a boxcar, etc. A conversational event may do any of the following:

1. Introduce a new described situation (e.g. situation $s$, in *ce1*), which is analo-

gous to the case in which a new stack is created and a new focus space added on top of it. This is what *ce1* does.

2. Extend the situation described by previous conversational events, which corresponds to the case in which an utterance `adds' material to an existing focus space on the stack (e.g., the re-use of $s$ in *ce2*).

3. Introduce a new described situation which extends an existing one (as in the case in which a subplan is being discussed), which corresponds in Grosz and Sidner's terms to the case in which a new focus space is pushed on top of the focus space stack, allowing access to the previous focus spaces. E.g., after introducing a (possible) situation $s$ by saying something like *We need to get the boxcar to Avon by 5pm*, a speaker might continue by saying *We need an engine to move the boxcar*. In this case, the third argument of the conversational event associated with this new utterance describes a situation $s'$ such that $s \sqsubseteq s'$. Recall that situations are organized in an inclusion hierarchy, and that each constituent of a situation $x$ is also a constituent of every situation $x'$ that extends $x$. Thus, a discourse referent introduced in $s$ is also part of $s'$.

In other words, we propose that the rules of pragmatic accessibility that Grosz and Sidner attribute to the presence of a `stack' in the discourse model, can be derived without need of additional stipulations simply by assuming that the content of each speech act is a statement about a situation, and that situations are organized in a hierarchy. All that we still need to assume is what we believe is the crucial aspect of Grosz and Sidner's proposal, namely, the connection between intentional structure and attentional state. We need to assume, that is, that if *ce1* **dominates** *ce2*, then the described situation of *ce2* is an extension of the described situation of *ce1*; whereas if *ce1* **satisfaction-precedes** *ce2*, then the described situation of *ce2* is the same as the described situation of *ce1*. More formally put,

(14) a. $x : \mathbf{pred}(a,b,s : \Phi) \wedge y : \mathbf{pred}'(c,d,s' : \Phi') \wedge \mathbf{dominates}(x,y) \supset s \sqsubseteq s'$

b. $x : \mathbf{pred}(a,b,s : \Phi) \wedge y : \mathbf{pred}'(c,d,s' : \Phi')$
$\wedge$ $\mathbf{satisfaction\text{-}precedes}(x,y)$
$\supset s = s'$

(where all variables are meant to be universally quantified, and **pred** and **pred′** are supposed to be among the predicates that characterize core speech acts). Similar axioms can also be specified for each argumentation act, depending on whether they specify dominance or satisfaction-precedes relations.

To conclude the discussion, we would like to make a couple of comments. First of all, the model just presented is not strictly equivalent to Grosz and Sidner's. Whereas they had a single stack, we have here one 'stack' for each conversational thread; this makes our model of the attentional state, strictly speaking, closer to the 'graph-structured' stack proposed by [Rosé *et al.*, 1995] than to Grosz and Sidner's. We assume that the discourse situation contains at each point in time information about which situation is the current 'focus space'—which we call DISCOURSE TOPIC; the default will then be to preserve the current discourse topic rather than shift to another one. It should be noted, however, that already the model proposed by Grosz and Sidner was not a stack in a strict sense. They made recourse to several kinds of auxiliary devices, such as 'impenetrable barriers' on the stack, auxiliary stacks, and so forth; all such mechanisms can be replaced much more simply by the kind of structure proposed here.

Secondly, it is important to note that the notion of 'situational inclusion' we are assuming here is not *spatio-temporal* inclusion, but *informational* inclusion; i.e., $s \sqsubseteq s'$ does not assert that the situation $s$ is a subevent of $s'$, but that the latter 'contains more information'—i.e., more objects, and more propositions. (For some discussion of this distinction, see [Hwang and Schubert, 1993].)

Finally, we would like to emphasize an important advantage of this way of looking at each utterance's contribution to the attentional state specifying the focus space; namely, that we can use the existing work in semantics on the dynamics of discourse to specify such contribution. The contribution to the focus space is easy to specify in the case of utterances such as *John likes Mary* or *There is an engine at Avon*; it is more difficult in the case of the sentences one actually finds in task-oriented dialogs, such as *We have to send a boxcar to Avon*. Such utterances can be felicitously continued by utterances such as *We need an engine to move the boxcar*, but not by utterances like *The boxcar is empty*, for the obvious reason that the boxcar introduced in discourse is only a potential boxcar, not an actual one. Much work on specifying the contribution of utterances to discourse has been done in linguistics; by making our theory of discourse compatible with those assumed in this work, we make it much easier for us to take advantage of such work.

## 6 Grounding

Once one starts looking more carefully at the way the common ground is actually established in natural conversations, one realizes that a further departure from the view of a discourse model taken in DRT is required. As described above, DRT (and almost all previous work in the reference resolution, discourse structure, and speech acts traditions) makes use of the assumption that everything that is uttered becomes a part of the common ground immediately, and, hence, is available for reference. As shown, e.g., in [Clark and Wilkes-Gibbs, 1986; Clark and Schaefer, 1989], however, this assumption is simply an idealization. Utterances by one conversant must be recognized and acknowledged by the other before entering common ground. The collaborative process of adding to the common ground (called GROUNDING) thus must include installments by each of the conversants. These installments can introduce new material, or continue, repair or acknowledge previously introduced material. Acknowledgments may be explicit —after the first utterance of (6), B could have acknowledged by uttering something like *okay* or *right*—or tacit —the second utterance in (6) can be interpreted as providing a tacit acknowledgment of the first utterance, while at the same time performing additional conversation acts. The grounding process plays a significant role in shaping the form of actual conversations, as shown by the following example from the TRAINS corpus, in which M explicitly checks each new inference about the plan and S acknowledges them one by one:[15]

(15)
```
3.1  M: now
3.3   : so
3.4   : need to get a boxcar
3.5   : to Corning
3.6   : where there are oranges
3.7   : there are oranges at Corning
3.8   : 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
7.2   : at Avon
7.3   : right
8.1  S: right
```

Clark and Schaefer presented an off-line model of the grounding process in [Clark and Schaefer, 1989], while a computational account was provided in [Traum, 1994] and implemented within the TRAINS system. In [Traum, 1994], participation in the grounding process is viewed as the performance of GROUNDING ACTS, one of the levels of conversation acts from [Traum and Hinkelman, 1992]. According to the theory presented in [Traum, 1994], when an agent utters something, in addition to

[15]Taken from Dialog d91-6.1 in [Gross *et al.*, 1993].

performing core speech acts like those discussed in section 4—such as **suggest**, **inform**, etc.—the agent is also performing one or more grounding acts. Presented material that could be acknowledged together (e.g., with a single *okay*) is grouped together into a DISCOURSE UNIT (DU). Grounding acts include (among others devoted to repairs) **init**, which opens a new DU, **continue**, which adds more material to an already open DU, and **ack**, which makes the contents of the DU enter the common ground. and Each DU has its own state, encoding whether or not the DU has been grounded and which kinds of actions (e.g., acknowledgments or repairs) by each agent are needed to ground the content. Also associated with each DU is a model of what the discourse context (and common ground) would be like if the DU were to be grounded. A bounded stack[16] of accessible DUs is maintained in the model, as context for recognizing and deciding to perform grounding acts. Table 1 summarizes the grounding acts, while table 2 describes the state transitions for DUs which occur after the performance of grounding acts. In this table, superscripts represent the performing agent — I for the intiator of that DU, and R for the responder (the other agent).

| Label | Description |
|---:|---|
| initiate | begins a new DU, content separate from previous uncompleted DUs |
| Continue | continuation of previous material by the same speaker |
| Acknowledge | demonstrates or claims understanding of previous material by other speaker |
| Repair | Correct (potential) misunderstanding of DU content |
| Request Repair | Signal of lack of understanding |
| Request Ack | Signal for other to acknowledge |
| Cancel | top work on DU, leaving it ungrounded and ungroundable |

Table 1: Grounding Acts

The model of context in CRT++, as described in the previous section, already includes two of the features needed to represent the grounding process, namely, the occurrence of conversational events, and the possibility of organizing conversational events into threads. A fairly minor augmentation is the addition of a new relation, **ga**, which holds between a locutionary act and a grounding act if the grounding act has been generated by the performance of the locutionary act. For example, in the exchange in (6), B's utterance can be seen as generating both an (implicit) acknowledgment (**ack**) of the DU that includes A's utterance, as well as

[16]The stack is limited to the $n$ most recently initiated DUs. DUs which have "fallen off" the bottom of the stack are no longer accessible. In the TRAINS-93 implementation, $n$ was set to 3.

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

*repair request is ignored

Table 2: DU State Transition Diagram

initiating (**init**) a new DU for the new content that it contains in its own right. We can represent this by augmenting the Root DRS with the conditions shown in (16). We have used *du1* for the DU acknowledged by B's utterance (containing *sug1* and other information from the interpretation of *ce1*), and *du2* for the DU that *ce2* initiates. Both *du1* and *du2* are represented using conversational threads.[17] The resulting state of the DUs after the utterance is described by conditions like that used to describe the intention i1 in section 4. A predicate such as **state1**(*du2*) has other implications about the mental state of the participating agents, as described in [Traum, 1994]. In particular, if a DU is in state 1, then the initiator intends that the content of the DU be mutually believed, while if it is in state F, the content is believed to be mutually believed. Other states concern also the intentions of the

[17] Although DUs are represented using the same technical device as Discourse Segments (namely, conversational threads), the two concepts should not be confused. Grounding and intentional discourse structure are two different phenomena. Any given conversational event will be part of at least two different conversational threads, one representing its groundedness, and another representing the relation of the purpose of the speaker in making the utterance to that of other utterances. E.g., *ce1* is a part of both *ct1* and *du1*.It is still an open question as to what (if any) the necessary connection is between these two types of structure.

responder and obligations of the two agents.

(16)

| . . . *du1* . . . *ack1 du2 init2* |
|---|
| . . . |
| *ack1* : $\boxed{\textbf{ack}(B,du1)}$ |
| *init2* : $\boxed{\textbf{init}(B,du2)}$ |
| **ga**(*ce2*,*ack1*) |
| *s-du1-ce2* : **stateF**(*du1*) |
| *s-du1-ce2* @ *t2* |
| **ga**(*ce2*,*init2*) |
| *s-du2-ce2* : **state1**(*du2*) |
| *s-du2-ce2* @ *t2* |

Although representing the occurrence of grounding acts and their organization into discourse units is a straightforward matter, that's not all there is to grounding: we also need to be able to distinguish the `grounded' part of a discourse situation from that which is `ungrounded'. Unacknowledged statements result in an ungrounded characterization of that part of the discourse situation; acknowledgments can then be interpreted as moving information from an ungrounded state to a grounded state.

Because of the assumption that everything that gets added to a context becomes part of the common ground, in `vanilla' DRT one can simply assume that the Root DRS represents what the conversational participants mutually believe (or, perhaps, what one participant believes that is mutually believed) without worrying about such mental states any further. But the difference between grounded and ungrounded states is precisely that the conversational participants have agreed on the former, but not on the latter. A simple fix, such as allowing only grounded material into the Root DRS will not work either – material that has been presented but not yet grounded is still available for reference by both conversants. This is exactly the problem mentioned previously in the first section.

What is needed is a broader perspective, dividing the Root DRS into grounded and ungrounded parts, both of which are part of the general discourse context. We propose to model the grounding process as follows. We modify the DRS construction algorithm so that the result of an update due to a conversational event generating an **init**, **continue**, or **repair** grounding act is always an ungrounded DRS. We then define **acknowledgments** as moves of the DRS from an ungrounded to a grounded state.

The discourse situation (root DRS) will now consist of two partitions, both accessible for reference purposes, though in subtly different ways. First, we have the GROUNDED ROOT DRS (GR-DRS), representing the actual common ground. Information within is assumed to be shared by both dialogue participants. In addition, we have an UN-GROUNDED ROOT DRS (UR-DRS), consisting of those conversational threads which represent ungrounded DUs. Each DU thread within the UR-DRS represents a particular agent's view of what should become part of the common ground (e.g., when A initiates a DU, with content $\phi$ about described situation s: **intend**($A$,**MB**$(\{A,B\},s : \phi)$)). Items in the UR-DRS can serve as anchors for referring expressions, but are provisional on the material actually having been understood correctly and later acknowledged. In fact, making reference to an item in an ungrounded DU initiated by the other agent can be a means of acknowledging that thread. In this manner, we can also model the collaborative nature of the referring process itself, in a manner similar to [Clark and Wilkes-Gibbs, 1986; Heeman and Hirst, 1995].[18]

The UR-DRS is a DRS that extends the GR-DRS in the sense that if the GR-DRS is a pair $\langle U,C\rangle$ where U is a set of referents and C a set of conditions, the UR-DRS is a pair $\langle U \cup U', C \cup C'\rangle$, where $U'$and $C'$represent the ungrounded referents and conditions. After processing the acknowledgment performed in in *ce2*, the GR-DRS will look something like (17), while the UR-DRS will also include the information shown in (18)[19], as well.

[18]The formalization of this distinction between GR-DRS and UR-DRS, and of the effect of grounding acts, are based on the hypotheses about mental states and their connection with discourse interpretation developed in work by Asher and Kamp, among others, [Asher, 1993; Kamp, 1990]. Asher and Kamp see discourse interpretation as a process of operations on `mental states', collections of objects of type $\langle M,K\rangle$, where M is a `mental attitude' and K is a DRS, where DRSs also denote objects that correspond fairly closely to Frege's `senses'; the statements about the world that these objects represent are specified by a separate, `referential' interpretation, which is just the semantics seen in the previous sections. schematically stands for seen as

[19]These figures show only the aspects that we have discussed previously in the paper. There are, of course, many additional facts that will be part of these DRS as a result of processing even this short bit of dialogue.

(17)

$$
\boxed{
\begin{array}{l}
s \ldots s' \, i1 \; tj \; pl \ldots ce1 \; t1 \; sug1 \; ct1 \; du1 \ldots ce2 \; t2 \; ack1 \\
\hline
\ldots \\
s \sqsubseteq s' \\
i1 : \textbf{intend}(A, s' : \boxed{\textbf{at}(y,Bath)}) \\
i1 \; @ \; tj \\
\ldots \\
ce1 : \textbf{tell}(A,B, s : \boxed{\begin{array}{l} x \; w \\ \hline \textbf{engine}(x) \\ \textbf{Avon}(w) \\ \textbf{at}(x,w) \end{array}}) \\
ce1 \; @ \; t1 \\
\ldots \\
sug1 : \textbf{suggest}(A, \boxed{\textbf{use}(\{A,B\},x,pl)}) \; \textbf{generate}(ce1,sug1) \\
\textbf{implicates}(sug1,pl : \boxed{\begin{array}{l} m \\ \hline \textbf{move-boxcar}(m) \\ \textbf{boxcar}(m,y) \\ \textbf{engine}(m,x) \\ \textbf{dest}(m,Bath) \end{array}}) \\
sug1 \sqsubseteq ct1 \\
\\
ce2 \; @ \; t2 \\
\textbf{ga}(ce2,ack1) \\
\\
ack1 : \boxed{\textbf{ack}(B,du1)}
\end{array}
}
$$

(18)

$$\begin{array}{|l|}\hline s \ldots du2\ init2 \\ \hline \ldots \\ ce2 : \ \mathbf{tell}(B,A,\ s : \begin{array}{|l|}\hline y\ u \\ \hline \mathbf{boxcar}(y) \\ \mathbf{at}(y,w) \\ \mathbf{hooked\text{-}to}(y,u) \\ u = x \\ \hline \end{array}\ ) \\ init2 : \ \boxed{\mathbf{init}(B,du2)}\ \ldots \\ \\ \hline \end{array}$$

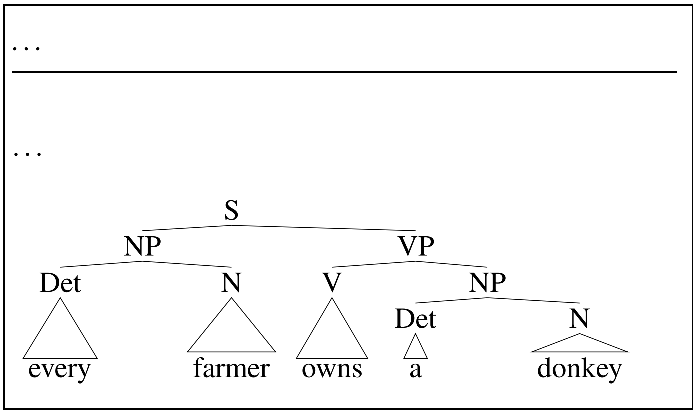
## 7 Building the Discourse Model

In DRT, the discourse model is updated after each utterance by first adding information about the syntactic structure of the utterance (called an UNINTERPRETED CONDITION) to the previous discourse model, and then applying `rewriting rules' called CONSTRUCTION RULES that replace the initial, purely structural representation of the sentence with partially disambiguated interpretations in which some of the ambiguous aspects of the sentence, such as the scope of operators, have been resolved. The construction rules are entirely driven by syntactic information. For example, upon encountering the utterance *Every farmer owns a donkey* at time *t+1*, the discourse model at time $t$ is initially updated by the following uninterpreted condition:

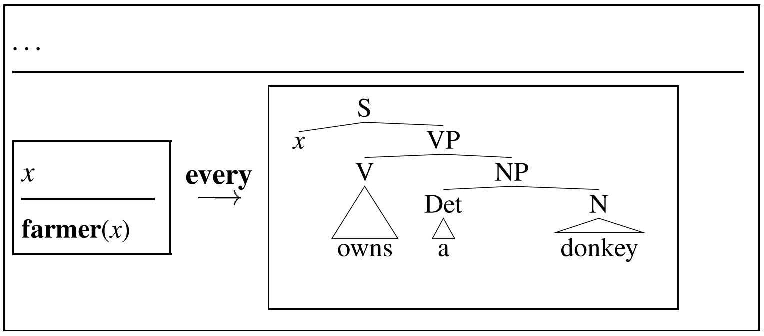
(19)

$$\mathrm{DM}(t) = \begin{array}{|l|}\hline \ldots \\ \hline \ldots \\ \hline \end{array}$$

$\Rightarrow$

DM($t$+$1$) = 


After determining that the quantifier *every farmer* takes wide scope over the rest of the sentence, the initial interpretation is rewritten, resulting in the following partially disambiguated interpretation, still including an uninterpreted condition:

(20) DM($t$+$1$) = 


The process continues until a completely disambiguated interpretation is obtained.

Some characteristics of this view of interpretation are preserved in CRT++; most importantly, the idea that discourse interpretation begins with an undetermined interpretation that gets progressively refined. However, the view of interpretation in CRT++ differs from the one just discussed in two respects. One difference is that we see discourse interpretation as a defeasible reasoning process in which many different sources of information, including syntactic information, referential information, and some aspects of commonsense knowledge, including at least lexical information, interact. Much recent work has tried to develop an `inferential' and `integrated' view of the disambiguation process that does not rely on the assumption that all disambiguated interpretations are produced. The result of these efforts have been several theories of UNDERSPECIFIED REPRESENTATIONS [Alshawi, 1992; Reyle, 1993; Poesio, 1995c]. An underspecified representation, much as an uninterpreted condition in DRT, encodes multiple interpretations; the difference is that the former, unlike the latter, has a proper semantics, and their inferential behavior can be specified by a logic.[20]

A second and more important difference from DRT is that the view of interpretation in CRT++ has been developed to account for the actual nature of spoken language. As anybody who has ever looked at a faithful transcription of spoken conversation knows, spoken language consists for the most part of sentence fragments mixed with repetitions, pauses, repairs, and other disfluencies, all of which however contribute clues to the interpretation process—for example, hesitations and pauses provide important clues for intention recognition. As an example, here is a fragment from a TRAINS transcript:

(21)
```
9.1  M: so we should
9.2   : move the engine
9.3   : at Avon
9.4   : engine E
9.5   : to
10.1 S: engine E1
11.1 M: E1
12.1 S: okay
13.1 M: engine E1
13.2  : to Bath
13.3  : to /
13.4  : or
13.5  : we could actually move it to
        Dansville to pick up the
        boxcar there
14.1 S: okay
```

A model of interpretation that relies on the assumption that input is processed complete sentence by complete sentence could not account for subdialoges like (21), which include embedded repairs before a complete sentence has been performed. What is needed is a theory of interpretation in which the process of building up the syntactic structure of an utterance is mixed with other interpretive processes.

The solution proposed in [Poesio, 1995b; Poesio, 1995a] follows naturally from the view of the discourse model taken in CRT++. Once one accepts that the discourse model consists of information about the discourse situation, it is but a small step to accept that the model may contain information not just about utterances of complete sentences, but also about utterances of incomplete sentences,

[20] On this point, see also [Deemter, 1991].

fragments, or even pauses—that we will call MICRO CONVERSATIONAL EVENTS.

More precisely, we assume that as a consequence of the utterance of the word *the*, the discourse model is updated by recording that an utterance of the string "the" took place, with syntactic category (**cat**) DET and content —i.e., interpretation— $\langle$s,$\lambda$P. P(x) $\wedge$ Q(x)$\rangle$. (This interpretation for the definite determiner derives from the DRT treatment, in which definites and indefinites introduce free variables in context. The expression $\langle$s,$\Phi\rangle$ denotes the value of $\Phi$, with variables evaluated at the assignment associated with $s$.) Using $\rightsquigarrow$ for the function from micro-ces to their contents and **cat** for the function from micro-ces to their syntactic category, this can be represented as follows:

(22)

| $\mu$-$ce_1$,$s$ |
| --- |
| ... |
| $\mu$-$ce_1$ : $\boxed{\textbf{utter}(a,\text{“}the\text{”})}$ |
| $\mu$-$ce_1 \rightsquigarrow \langle s,\lambda P.\ P(x) \wedge Q(x)\rangle$ |
| $\textbf{cat}(\mu\text{-}ce_1)$ = DET |
| ... |

A subsequent utterance of the word *engine* results in the following discourse model:

(23)

| $\mu$-$ce_1$,$s$,$\mu$-$ce_2$ |
| --- |
| ... |
| $\mu$-$ce_1$ : $\boxed{\textbf{utter}(a,\text{“}the\text{”})}$ |
| $\mu$-$ce_1 \rightsquigarrow \langle s,\lambda P.\ P(x) \wedge Q(x)\rangle$ |
| $\textbf{cat}(\mu\text{-}ce_1)$ = DET |
| $\mu$-$ce_2$ : $\boxed{\textbf{utter}(a,\text{“}engine\text{”})}$ |
| $\textbf{cat}(\mu\text{-}ce_2)$ = N |
| ... |

Note that the discourse model at this stage is radically underspecified: not only it does not include information about a complete sentence, it does not even include information about the lexical interpretation of the word string "engine," or indeed abot the syntactic relation between the determiner and the noun, if any. This initial information is the basis for all inferential processes that deal with discourse interpretation: from lexical disambiguation to speech act determination. The task of the lexical disambiguation processes, for example, is to specify the content of lexical items, e.g., to find a predicate that can serve as interpretation of the word-string "engine" in the current context; the result of this disambiguation process is to augment the discourse model with information of the form $\mu\text{-}ce_2 \rightsquigarrow \langle s, \mathbf{engine}_1 \rangle$, where $\mathbf{engine}_1$ is one of the senses of *engine*. The task of syntactic disambiguation is to assemble micro-conversational events in larger micro-conversational events, e.g., to infer from (23) the occurrence of a micro-conversational event of syntactic category NP and constituents $\mu$-ce$_1$ and $\mu$-ce$_2$, as follows (we use $\prec$ to represent the function from micro-ces to the unique micro-ce they are a syntactic constituent of):[21]

(24)

$$
\begin{array}{|l|}
\hline
\mu\text{-}ce_1, \mu\text{-}ce_2, \mu\text{-}ce_3 \\
\hline
\ldots \\
\mu\text{-}ce_1 : \boxed{\mathbf{utter}(a, \text{“}the\text{”})} \\
\mu\text{-}ce_1 \rightsquigarrow \langle s, \lambda P.\, P(x) \wedge Q(x) \rangle \\
\mathbf{cat}(\mu\text{-}ce_1) = \text{DET} \\
\mu\text{-}ce_2 : \boxed{\mathbf{utter}(a, \text{“}engine\text{”})} \\
\mathbf{cat}(\mu\text{-}ce_2) = \text{N} \\
\mathbf{cat}(\mu\text{-}ce_3) = \text{NP} \\
\mu\text{-ce}_2 \prec \mu\text{-ce}_3 \\
\mu\text{-ce}_1 \prec \mu\text{-ce}_3 \\
\ldots \\
\hline
\end{array}
$$

The task of reference resolution is to determine the content of referential noun phrases, i.e., to add to the discourse model facts of the form $\mu\text{-}ce_3 \rightsquigarrow \langle s, \lambda P.\, P(a) \rangle$, where *a* is an accessible antecedent in context. (See [Poesio, 1994; Poesio, 1993] for details about definite description interpretation.) Finally, the task of intention recognition is to infer the conversation acts generated by micro-conversational events (conversation act recognition is described in [Traum and Hinkelman, 1992; Traum, 1994]).

A few comments are in order. First of all, we assume here that such disambiguation processes can be formalized as defeasible inferences; for a formalization of lexical disambiguation, syntactic disambiguation, and scope disambiguation using Reiter's Default Logic, see [Poesio, 1995a]. One might claim that (default) inference is not the most efficient way to do parsing. This representational framework allows one to represent coherent (underspecified) forms for all stages in the parsing process, however using it in a system does not require that all processing be performed as (default) logical inference. While this kind of inference is necessary in some cases, such as for interpreting disfluencies and partial utterances, one can

[21] The `inferential' model of syntactic disambiguation assumed here bears some connections with the view of parsing assumed in Categorial Grammar, see [Poesio, 1995a].

still use traditional (or statistical) techniques to recognize a larger conversationa event, from which the occurance of the intermediate micro conversational events can be deduced (if needed).

Secondly, we would like to point out that the view of interpretation just presented leads naturally to an *incremental* view of utterance processing. This is consistent with the most recent findings of the psychological literature on human language processing, and anyway the transcripts of spoken conversations abound with illustrations of conversants responding to partial utterances. An example of incremental interpretation is in (21): it is clear from the way `S` reacts in 10.1 to `M`'s utterances in 9 that at that point `S` has already determined the referent of the definite description *the engine at Avon* and noticed that it is not `E`, but `E1`. The model of interpretation in CRT++, just sketched can represent fairly naturally what took place in that subdialogue, whereas this would be impossible for a more traditional model which requires full sentence representation, such as that used in the TRAINS-93 system [Traum *et al.*, to appear 1996].

Thirdly, we would like to note that in a model such as the one we are presenting there is no need to stipulate the existence of `surface speech acts' (e.g., **tell** or **ask**) or `referential acts' used in accounts of speech act inference such as [Allen and Perrault, 1980]. The only reason for introducing such acts in one's formalization was to serve as a form of underspecified representation that could `trigger' the intention recognition process; this purpose is served by micro-conversational events in CRT++.

Finally, the idea that different discourse interpretation processes communicate by updating a single model is clearly related to work on blackboard architectures, such as [Hayes-Roth, 1985]. One way to see the micro-conversational events model is as a way to spell out the details of a conversation processing model consistent with such an architecture, and to relate it to work on semantics.

The grounding acts described in the previous section are actually realized at the level of the numbered utterances shown in (21), rather than at the complete sentence level conversational events. **Continue** grounding acts bind these smaller utterances together into compound contributions within the same DU. In addition, **repair** grounding acts modify the contents of DUs as they are being built up, in order to handle repairs such as that in (21).

# 8 Contextual-based Processing

There are a number of processes that a conversational agent must perform involving context. The representation of context developed in the previous sections provides both a framework and general mechanisms for forming specific theories and algorithms to handle the various sub-problems in a context of a unified whole. Thus, e.g., as reference resolution and discourse coherence can play important roles in determining grounding, in the current framework one can directly include rules about these phenomena as part of the grounding process.

While we are still in the preliminary stages regarding the development of these algorithms, many ideas can carry over from previous work. While the details of these are beyond the scope of the current paper, in this section we present brief descriptions of some of these processes, with pointers back to the papers describing the details in similar logical frameworks.

## 8.1 Conversation Act Recognition

As stated in section 7 above, recognizing conversation acts amounts to a process of inferring what was meant by the micro-conversational acts and their combinations. The method of recognizing acts at each of the levels that was implemented in the TRAINS system can be easily transferred to the CRT++ model. For each level, a two-stage process was used, first positing potential acts from syntactic or lexical features (e.g., an **inform** from a declarative sentence, a **release-turn** from a rising intonation, **temporal-succession** from the words “and then”, or an **acknowledgement** from the word “ok”) or expectations from the context (e.g., an **acknowledgement** by a responder with the turn when a DU in state 1 is in focus, a **suggest** during the plan-formation phase in a TRAINS dialogue, or an **answer** following a question), followed by a process of *pruning* the hypotheses by considering what was actually said and how it relates to the contextual structure, including sometimes inference and plan recognition. For example, although a declarative utterance describing information that is part of the common ground cannot (according to the standard felicity conditions) be an **inform**. Likewise, in the dialogue fragment in (15), utterance 5.1 will be posited to be in a followup relation, based on the clue-word “so”. However, the plan reasoner will not be able to connect it up to the previous content act in 3.7, and decided it must be related to 3.4-6. If it were unable to make a sufficient connection, this hypothesis would have been dropped as well. Liekwise, for grounding acts, the relevence of new material to old will be important in establishing which act is being performed. If new material is related to old material by a different speaker (in the appropriate states), this can be an acknowledgement, even without an explicit cue word like “ok”. **Repairs** can be distinguished from **continues** by *deleting* or *replacing* explicit or implicit content in a current DU, rather than just adding on more material.

## 8.2 Dialogue Management

The theory of context representation outlined above can also serve as a basis for dialogue management in a conversation system. The TRAINS dialogue manager, described in [Traum and Allen, 1994], decided what the system should do or say, based on the relationships between aspects of the discourse context and mental state. These aspects included: obligations to say something (e.g., to answer a question or repair a nonunderstanding), prior intentions to say something, conversational goals (e.g., to form a domain plan), and ungrounded DUs (for performing acknowledgments and repairs). As all of these aspects can be represented in the current framework, the algorithms developed in [Traum and Allen, 1994] and similar proposals for dialogue management can be easily transported to work with the unified model of context provided above. For example, the TRAINS-93 dialogue manager prioritized the motiviations for action as follows:

| | **Motivation** | **Action** |
|---|---|---|
| 1 | Discourse Obligations | E.g., User YNQ $\rightarrow$ Intend Inform-if |
| 2 | Obligation: Don't interrupt<br>Obligation: other's turn | wait, listen, and comprehend<br>other's utterances |
| 3 | Intended Speech Acts | perform utterances |
| 4 | Obligation: Grounding | E.g., Intend Ack, ReqAck |
| 5 | Discourse Goals:<br>Domain Plan Negotiation | **Intend** Accept, Reject, Propose,<br>Counter |
| 6 | High-level Discourse Goals | E.g., domain planning, execution |

Table 3: Dialogue Manager Action Priorities

All of these involve inspecting some aspects of the context and taking action if particular relationships hold. For instance, at the grounding level, if this is highest priority requiring action, if the agent is a responder to an accessible DU in state 1 or the initator of a DU in state 3, the system will acknowledge, whereas it will request acknowledgement or cancel if it is an initiator of a DU in state 1 or the responder of a DU in state 3. Repairs and requests for repair generally come in at the higher priority level of obligations. Likewise, initates and continues will come in from saying other intended material.

## 8.3 Generation

An explicit model of context such as presented here can also serve as context for NL generation. While the dialogue manager will decide *what* to say, as well as *when* to say it, issues of exactly *how* to realize this content are left up to an NL generator in the TRAINS system. While generation has not been a focus of our work, the same aspects of context (e.g., discourse structure and referent accessibility) that license language interpretation processes should also allow a generator to reason about whether a particular form will convey the desired meaning. For instance, in producing the second utterance in (6), B might take into account the accessibility of the marker for Engine E2 in deciding whether a pronominal reference were possible. Likewise, the existence of this marker only within the UR-DRS would lead B to reason that this utterance would be taken as an acknowledgment (since it would not be sensible if the prior utterance has not been understood. In addition, all of the lexical information from the micro-conversational events is available, so that the participant can produce a continuation, repair, or repsonse using the same or related words. Thus, for example, in a dialogue like that in (15), a generator would be able to produce replies as in 4.1, 6.1, and 8.1 to the preceeding prompts using the same keywords.

### 8.4 Repair and Attitude Revision

Many of the processes described above, especially implicature, but also including speech act interpretation, reference resolution, and decisions on grounding, are provisional assumptions — not guaranteed to be accurate when performed by one conversational participant in an on-line manner. This is not a major problem within our current framework, because the discourse situation is not a single fixed entity that becomes monotonically further specified (as in individual DRS's in DRT). Rather, the discourse situation (as represented by the GR-DRS and UR-DRS) contains all that is known about the conversation at the current time. As time goes on, and more conversational events occur, some conditions in which a DRSwhich characterized a described situation may end up being no longer a part of the discourse situation. A **repair** grounding act will remove at least one condition that is part of the current DU, and replace it with another representing the updated material. Similarly, individual hypotheses about, e.g., what speech act was performed or which object was the anchor of a referring expression can be revised by replacing one condition with another in the updated discourse context. Other important issues include the ability to detect misconceptions or misunderstandings by one agent or another, and deciding when a repair is necessary or has been performed. We are currently working on including a general model of repair phenomena to the unified model. For some preliminary ideas in a related situation theory, see [Traum, 1994].

# 9 Conclusion

The model of context adopted by a system to engage in natural language conversation must include many different sorts of information. Often, the most natural representation of context for performing one type of processing makes it more difficult or impossible to do other types of processing. Based on experiences in confronting and overcoming just this type of issue within the TRAINS system, we have presented the beginning stages of a unified model of conversational context. The model already can support many different kinds of inference, as needed for reference resolution, speech act recognition, and dialogue management, in a way that is accessible for on-line processing algorithms.

CRT++ uses tools from Discourse Representation Theory, but adopts a different view of the common ground, as consisting of information about the discourse situation. This assumption leads to a natural integration of information about speech acts into the model, as well as a simple way of specifying the initial, underspecified interpretation of an utterance.

Our model incorporates many ideas from the theory of discourse structure proposed by Grosz and Sidner, but some of the characteristics of their model are made to follow from more general principles of interpretation; furthermore, we generalize the notion of discourse structure, by allowing utterances to convey different kinds of `discourse purposes', some of which related to `conversational management tasks' such as turn taking or grounding.

Current work includes extending the depth and coverage of this representation and implementation of new algorithms working within the unified framework.

## Acknowledgements

We would like to thank members of the TRAINS project at the university of Rochester, particularly James Allen, George Ferguson, Peter Heeman, Chung-Hee Hwang, and Len Schubert, for collaborations on the TRAINS system which inspired the present work. We would also like to thank the anonymous reviewers for helpful suggestions on this presentation.

## References


[Allen and Perrault, 1980] Allen, J. F. and Perrault, C. 1980. Analyzing intention in utterances. *Artificial Intelligence* 15(3):143–178.

[Allen *et al.*, 1995] Allen, J. F.; Schubert, L. K.; Ferguson, G.; Heeman, P.; Hwang, C. H.; Kato, T.; Light, M.; Martin, N.; Miller, B.; Poesio, M.; and Traum, D. R. 1995. The TRAINS project: a case study in building a conversational planning agent. *Journal of Experimental and Theoretical Artificial Intelligence*, 7:7–48, 1995.

[Allen, 1983] Allen, J. F. 1983. Recognizing intentions from natural language utterances. In Brady, Michael and Berwick, Robert C., editors 1983, *Computational Models of Discourse*. MIT Press.

[Alshawi, 1992] H. Alshawi, editor. *The Core Language Engine*. The MIT Press, 1992.

[Asher, 1986] Asher, N. 1986. Belief in discourse representation theory. *Journal of Philosophical Logic* 15:127–189.

[Asher, 1993] Asher, N. 1993. *Reference to Abstract Objects in English*. D. Reidel, Dordrecht.

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

[Barwise and Perry, 1983] Barwise, J. and Perry, J. 1983. *Situations and Attitudes*. The MIT Press.

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

[Clark and Marshall, 1981] Clark, H. H. and Marshall, C. R. 1981. Definite reference and mutual knowledge. In *Elements of Discourse Understanding*. Cambridge University Press.

[Clark and Schaefer, 1989] Clark, H. H. and Schaefer, E. F. 1989. Contributing to discourse. *Cognitive Science* 13:259 – 94.

[Clark and Wilkes-Gibbs, 1986] Clark, H. H. and Wilkes-Gibbs, D. 1986. Referring as a collaborative process. *Cognition* 22.

[Cohen and Levesque, 1990] Cohen, P. R. and Levesque, H. J. 1990. Rational interaction as the basis for communication. In Cohen, P.R.; Morgan, J.; and Pollack, M., editors 1990, *Intentions in Communication*. Morgan Kaufmann. chapter 12.

[Deemter, 1991] K. van Deemter. *On the Composition of Meaning*. PhD thesis, University of Amsterdam, 1991.

[Devlin, 1991] Devlin, K. 1991. *Logic and Information*. Cambridge University Press, Cambridge, UK.

[Fox, 1987] B. A. Fox. *Discourse Structure and Anaphora*. Cambridge University Press, Cambridge, UK, 1987.

[Goldman, 1970] Goldman, A. 1970. *A Theory of Human Action*. Princeton University Press, Princeton, NJ.

[Groenendijk and Stokhof, 1984] Groenendijk, J. A. G. and Stokhof, M.B. J. 1984. *Studies on the Semantics of Questions and the Pragmatics of Answers*. Ph.D. Dissertation, University of Amsterdam.

[Groenendijk and Stokhof, 1991] Groenendijk, J.A.G. and Stokhof, M.J.B. 1991. Dynamic Predicate Logic. *Linguistics and Philosophy* 14:39–100.

[Gross *et al.*, 1993] Gross, D.; Allen, J. F.; and Traum, D. R. 1993. The TRAINS 91 dialogues. TRAINS Technical Note 92-1, Computer Science Dept. University of Rochester.

[Grosz and Sidner, 1986] Grosz, B. J. and Sidner, C. L. 1986. Attention, intention, and the structure of discourse. *Computational Linguistics* 12(3):175–204.

[Grosz, 1977] Grosz, B. J. 1977. *The Representation and Use of Focus in Dialogue Understanding*. Ph.D. Dissertation, Stanford University.

[Hayes-Roth, 1985] B. Hayes-Roth. A blackboard architecture for control. *Artificial Intelligence*, 26:251–321, 1985.

[Heeman, 1991] Heeman, P. A. 1991. A computational model of collaboration on referring expressions. Master's Thesis, Technical Report CSRI 251, Department of Computer Science, University of Toronto.

[Heeman and Hirst, 1995] P. A. Heeman and G. Hirst. Collaborating on referring expressions. *Computational Linguistics*, 21(3):351–382, 1995.

[Hinkelman, 1990] Hinkelman, E. 1990. *Linguistic and Pragmatic Constraints on Utterance Interpretation*. Ph.D. Dissertation, University of Rochester.

[Hwang and Schubert, 1993] Hwang, C. H. and Schubert, L. K. 1993. Episodic logic: A situational logic for natural language processing. In Aczel, P.; Israel, D.; Katagiri, Y.; and Peters, S., editors 1993, *Situation Theory and its Applications, v.3*. CSLI. 303–338.

[Kameyama *et al.*, 1993] M. Kameyama, R. Passonneau, and M. Poesio. Temporal centering. In *Proc. of the Meeting of the Association for Computational Linguistics*, pages 70–77, Columbus, OH, 1993.

[Kamp and Reyle, 1993] Kamp, H. and Reyle, U. 1993. *From Discourse to Logic*. D. Reidel, Dordrecht.

[Kamp, 1981] Kamp, H. 1981. A theory of truth and semantic representation. In Groenendijk, J.; Janssen, T.; and Stokhof, M., editors 1981, *Formal Methods in the Study of Language*. Mathematical Centre, Amsterdam.

[Kamp, 1990] Kamp, H. 1990. Prolegomena to a structural account of belief and other attitudes. In Anderson, C. A. and Owens, J., editors 1990, *Propositional Attitudes—The Role of Content in Logic, Language, and Mind*. University of Chicago Press and CSLI, Stanford. chapter 2, 27–90.

[Karttunen, 1976] Karttunen, L. 1976. Discourse referents. In McCawley, J., editor 1976, *Syntax and Semantics 7 - Notes from the Linguistic Underground*. Academic Press, New York.

[Karttunen, 1977] Karttunen, L. 1977. The syntax and semantics of questions. *Linguistics and Philosophy* 1:1–44.

[Konolige and Pollack, 1993b] K. Konolige and M. E. Pollack. A representationalist theory of intention. In *Proceedings IJCAI-93*, 1993b.

[Lascarides and Asher, 1991] Lascarides, A. and N. Asher. 1991. Discourse relations and defeasible knowledge. In *Proc. ACL-91*, pages 55–63, University of California at Berkeley.

[Mann and Thompson, 1987] Mann, W. C. and Thompson, S. A. 1987. Rhetorical structure theory: A theory of text organization. Technical Report ISI/RS-87-190, USC, Information Sciences Institute.

[Nakhimovsky, 1988] A. Nakhimovsky. Aspect, aspectual class, and the temporal structure of narratives. *Computational Linguistics*, 14(2):29–43, June 1988.

[Novick, 1988] Novick, D. 1988. *Control of Mixed-Initiative Discourse Through Meta-Locutionary Acts: A Computational Model*. Ph.D. Dissertation, University of Oregon. also available as U. Oregon Computer and Information Science Tech Report CIS-TR-88-18.

[Perrault, 1990] Perrault, C. R. 1990. An application of default logic to speech act theory. In Cohen, P. R.; Morgan, J.; and Pollack, M. E., editors 1990, *Intentions in Communication*. The MIT Press, Cambridge, MA. chapter 9, 161–185.

[Poesio, 1993] M. Poesio. A situation-theoretic formalization of definite description interpretation in plan elaboration dialogues. In P. Aczel, D. Israel, Y. Katagiri, and S. Peters, editors, *Situation Theory and its Applications, vol.3*, chapter 12, pages 339–374. CSLI, Stanford, 1993.

[Poesio, 1994] Poesio, M. 1994. *Discourse Interpretation and the Scope of Operators*. Ph.D. Dissertation, University of Rochester, Department of Computer Science, Rochester, NY.

[Poesio, 1995a] M. Poesio. Disambiguation as defeasible reasoning over underspecified representations. In P. Dekker and J. Groenendijk, editors, *Proc. of the 11th Amsterdam Colloquium*, 1995.

[Poesio, 1995b] M. Poesio. A model of conversation processing based on micro conversational events. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Pittsburgh, 1995.

[Poesio, 1995c] M. Poesio. Semantic ambiguity and perceived ambiguity. In K. van Deemter and S. Peters, editors, *Semantic Ambiguity and Underspecification*. CSLI, Stanford, CA, 1995.

[Reichman, 1985] R. Reichman. *Getting Computers to Talk Like You and Me*. The MIT Press, Cambridge, MA, 1985.

[Reyle, 1993] U. Reyle. Dealing with ambiguities by underspecification: Construction, representation and deduction. *Journal of Semantics*, 10:123–179, 1993.

[Rosé *et al.*, 1995] C. P. Rosé, B. Di Eugenio, L. S. Levin, and C. Van Ess-Dykema. Discourse processing of dialogues with multiple threads. In *Proc. ACL*, Boston, MIT, June 1995.

[Traum and Allen, 1994] Traum, D. R. and Allen, J. F. 1994. Discourse obligations in dialogue processing. In *Proc. of the 32nd Annual Meeting of the Association for Computational Linguistics*, New Mexico. 1–8.

[Traum and Hinkelman, 1992] Traum, D. R. and Hinkelman, E. A. 1992. Conversation acts in task-oriented spoken dialogue. *Computational Intelligence* 8(3). Special Issue on Non-literal Language.

[Traum *et al.*, to appear 1996] D. R. Traum, L. K. Schubert, M. Poesio, N. G. Martin, M. Light, C. H. Hwang, P. Heeman, G. Ferguson, and J. F. Allen. Knowledge representation in the TRAINS-93 conversation system. *International Journal of Expert System*, to appear 1996. Special Issue on Knowledge Representation and Inference for Natural Language Processing.

[Traum, 1994] Traum, D. R. 1994. *A Computational Theory of Grounding in natural language conversation*. Ph.D. Dissertation, University of Rochester, Department of Computer Science, Rochester, NY.

[Webber, 1979] Webber, B. L. 1979. *A Formal Approach to Discourse Anaphora*. Garland, New York.

[Webber, 1988] B. L. Webber. Tense as discourse anaphor. *Computational Linguistics*, 14(2):61–73, June 1988.