Conversational Actions and Discourse Situations

Massimo Poesio
University of Edinburgh
Centre for Cognitive Science and HCRC

David R. Traum
TECFA, FPSE
Université de Genève

Abstract

We use the idea that actions performed in a conversation become part of the common ground as the basis for a model of context that reconciles in a general and systematic fashion the differences between the theories of discourse context used for reference resolution, intention recognition, and dialogue management. We start from the treatment of anaphoric accessibility developed in DRT, and 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 different kinds of 'structure' that play a role in conversation—discourse segmentation, turn-taking, and grounding—can be formulated in terms of information about speech acts, and use this same information as the basis for a model of the interpretation of fragmentary input.

1 Motivations

Although the slogan 'language is (joint) action' is accepted by almost everyone working in semantics or pragmatics, in practice this idea has resulted in theories of the common ground that differ in almost all essential details. We intend to show that this need not be the case, i.e., that the hypothesis that speech act occurrences are recorded in the common ground can serve as the basis for a model of language processing in context that reconciles in a general and systematic fashion the differences between the theories of the common ground adopted in current theories of reference resolution, intention recognition, and dialogue management.

[1]Address: 2 Buccleuch Place, Edinburgh EH8 9LW, Scotland, UK. Phone: +44 131 650 6988, Fax: +44 131 650 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

Our model is meant to be usable by an agent engaging in conversations as an internal, on-line representation of context. Our proposal is motivated 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, intention recognition and dialogue management. The problem is that 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.

One example of poor fit between existing theories of context is the contrast between, on the one hand, linguistically motivated theories of context developed to account for the semantics of anaphora (e.g., [Kamp and Reyle, 1993]); and on the other hand, the models of context proposed for intention 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]. These traditions resulted in very detailed proposals about context and context update;[1] but the resulting models of context differ significantly. 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.

A similar gap exists between the theories of the common ground developed in the speech act tradition and those assumed in Conversational Analysis (for an introduction, see [Levinson, 1983]). Conversational Analysts are concerned with aspects of inter-agent coordination such as turn-taking and the structure of repairs, which are of great importance for dialogue management but are typically ignored by traditional theories of speech acts. Unfortunately, a complete language understanding system needs to accomplish all of these tasks.

The contents of the paper are as follows. We concentrate first on introducing our position about what the common ground ought to contain; later we get to the issue of how the common ground is established. Section 2 is a brief introduction to Discourse Representation Theory (DRT), a theory of context developed in formal semantics that embodies the traditional basics of the reference resolution tra-

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

dition; we take a version of this theory as starting point for our formalization. In Section 3 we review the arguments for assuming that the common ground includes pragmatic as well as semantic information, and we propose a theory of the information about the discourse situation shared by the participants in a conversation centered around information about the occurrence of speech acts. We also show that this theory can be formalized using technical tools very similar to those proposed in DRT. In Section 4 we develop the theory to include an account of discourse structure. In Section 5 we discuss our theory of interpretation, and in Section 6 we use various ideas introduced in the previous sections to give an account of the grounding process. Although we do review very quickly in Section 5 how we assume this information about the discourse situation is used for interpretation, there is no space for an extensive discussion, and to talk about other modules of our system that rely on this information, such as the Dialogue Manager; these topics are discussed at length in [Poesio, 1994; Traum, 1994] and more briefly in [Allen *et al.*, 1995; Traum *et al.*, to appear 1996].

Most of our discussion below is based on transcripts of spoken conversations collected as part of the TRAINS project [Gross *et al.*, 1993]. These are conversations between two humans, a MANAGER and a SYSTEM, whose task is to develop a plan to transport goods around a simplified TRAINS WORLD consisting of cities connected by railway. The System and the Manager can't see each other, and communicate via microphones and earphones. They each have copies of a map of the Trains World.

2 A Minimal Representation of Context: Discourse Representation Theory

We will use the word 'context' to refer to the information that a conversant brings to bear when interpreting utterances in a conversation. Not everybody agrees on what this 'information' is; but virtually all researchers in the field agree that it includes at least the COMMON GROUND among the participants [Stalnaker, 1979; Clark and Marshall, 1981], i.e., the information that they share. Evidence for this role of the common ground is presented, e.g., by Clark and Marshall, who show that the felicitous use of referring expressions crucially depends on the correctness of the speaker's assumptions about the common ground.

From the point of view of reference resolution, the crucial information provided by context is which referents are available, and the fundamental property of utterances is that they add new discourse referents (as well as propositional information) to the common ground [Karttunen, 1976; Webber, 1979]. For example, an utterance of *There is an engine at Avon* has the effect of making two new DISCOURSE REFERENTS available for subsequent pronominalization, so that the utterance can be followed by the utterance *It is hooked to a boxcar*, where *It* refers back to *an engine*. The minimal requirement for a theory of context to be used for reference resolution is that it accounts 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 $(\exists\, x\ \exists\, w\ \mathbf{engine}(x) \wedge \mathbf{Avon}(w) \wedge \mathbf{at}(x,w))$, in which the variable x is bound within the scope of the existential quantifier and is not available for subsequent reference (i.e., conjoining this entire expression with something like P(x) would not have the desired result of making x bound by the existential quantifier).[2] The formalisms proposed by Artificial Intelligence researchers to model intention recognition are equally inappropriate, since reference is not one of the concerns. In Grosz and Sidner's theory [1986], context update is modeled via the focus space 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.

Modeling update is the *raison d'etre* of Discourse Representation Theory (DRT) [Kamp, 1981; Heim, 1982; Kamp and Reyle, 1993] and related 'dynamic' theories [Groenendijk and Stokhof, 1991]. It seems therefore appropriate to use such theories as the starting point for our formalization of context. We start from DRT, which has the most conventional semantics of all these theories.

2.1 Basic DRT

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 (facts about these discourse referents) that is typically represented in 'box' fashion. For example, the sentence in (1) is represented as in (2).

(1) There is an engine at Avon.

(2)
$$\begin{array}{|l|}\hline x\ w \\ \hline \mathbf{engine}(x) \\ \mathbf{Avon}(w) \\ \mathbf{at}(x,w) \\ \hline \end{array}$$

DRSs are logical expressions that capture the intuitions about discourse models expressed in [Karttunen, 1976; Webber, 1979], but can be given a precise semantics:

[2] These problems with first order logic are extensively discussed in chapter 1 of [Heim, 1982].

(2) 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. This semantics makes (2) logically equivalent to the existentially quantified statement ($\exists\, x\ \exists\, w$ **engine**(x) $\wedge$ **Avon**(w) $\wedge$ **at**(x,w)).

In general, a DRT characterization of a context consists of several DRSs 'embedded' within a distinguished DRS that represents the whole common ground. We use the term ROOT DRS to indicate this DRS, which is built incrementally and updated by each sentence in a discourse. The effects of a sentence on the common ground are specified by a DRS CONSTRUCTION ALGORITHM that adds new discourse referents and new conditions to the root DRS. For example, the effect of sentence *It is hooked to a boxcar* on the root DRS in (2) is the DRS in (4), specifying an interpretation for the discourse in (3). Note that (4) has two more discourse referents than (2), y and u, interpreting the definite *a boxcar* and the pronoun *it* respectively. The 'box' in (4) also contains the discourse referents introduced by the first sentence, which are thus accessible for reference purposes —in the sense that conditions like u **is** x, asserting that the denotation of the discourse referent u is identical with the denotation of the discourse referent x introduced in the first sentence, may be part of the interpretation of the text even though they refer to variables introduced as part of the translation of the first sentence.[3]

(3) There is an engine$_i$ at Avon. It$_i$ is hooked to a boxcar.

(4)

x w y u
engine(x)
Avon(w)
at(x,w)
boxcar(y)
hooked-to(u,y)
u **is** x

Most current work on the semantics of pronominal anaphora, definite descriptions, and ellipsis is cast in terms of formal discourse models such as DRT because of their explicitness. Hence, by adopting a model of context with a clear connection to DRT we can tie our pragmatic theories of reference resolution or intention recognition to work on semantics. However, the 'vanilla' version of DRT (i.e., the version presented in [Kamp, 1981] and revised in [Kamp and Reyle, 1993]) has three problematic characteristics from our perspective.

The first problem is that the emphasis in DRT is on representing the semantic

[3] For a detailed discussion of the DRT construction algorithm, see [Kamp and Reyle, 1993].

aspects of context; DRT abstracts away from all information of a pragmatic nature, including information that is needed for reference resolution purposes.[4] Secondly, a much simplified view of the process by which the common ground is updated is assumed in DRT, whereas, as we will see below, much of what is going on in a conversation are contributions concerned with ensuring that the participants's views of the common ground are synchronized. Finally, the construction algorithm as formulated by Kamp and Reyle does not assign an independent interpretation to each sentence, let alone to contributions to a text smaller than a sentence (this problem is known as 'lack of compositionality'). The algorithm is formulated as a set of rewrite rules that transform syntactic representations of complete sentences into conditions of a DRS. But as we will see below, the contributions to a dialogue are most often fragments (rather than complete sentences) and semantic interpretation processes begin well before a sentence is completed.

We will address the first two problems in the following sections. The third problem, DRT's lack of compositionality, has motivated much research in the formal semantics community; this work resulted in many alternative formulations of the construction algorithm that are compositional in the sense that the interpretation of a sentence is derived from the interpretations of the lexical items and 'local' composition operations [Groenendijk and Stokhof, 1991; Muskens, 1994].

The theory of interpretation we present below builds on Muskens' proposal, which we summarize in section 2.2. This section is somewhat technical, and may be skipped by those readers who are willing to accept our claim that sentences of English can be mapped into DRT expressions by means of techniques for semantic composition analogous to those used in theories of semantic interpretation such as Montague Grammar [Montague, 1973]. This is done in Muskens' theory by interpreting DRT expressions as expressions of a typed logic; the mapping also gives us a proof theory for our language. The readers skipping 2.2 may still want to give a look at the grammar at the end to get an idea of how this might be done.

2.2 Muskens' Compositional Reformulation of DRT

The basic idea of Muskens' approach (as well as of most 'dynamic' theories) is to capture the update properties of sentences by treating them as transitions among STATES: intuitively speaking, a sentence like (1) is thought of as specifying a transition from initial states in which x and w are not available for reference to ones in which they are and in which, furthermore, x and w satisfy the conditions imposed by the sentence. The sentence that follows (1) will, in turn, specify a transition from

[4] This limitation has been a concern for other researchers as well, with the result that other reformulations of DRT exist. We will briefly discuss some of these proposals below.

whatever state is the result of the transition specified by (1) to a new state in which additional conditions are imposed on *x* and *w*.

In the approach we are considering, this idea is formalized by thinking of DRSs as relations among states [Poesio, 1991b; Muskens, 1994]. We translate sentences as DRSs, and introduce a concatenation relation among DRSs ';' which allows us to compose transitions as follows: if K and K′ are DRSs, K;K′ specifies a transition from a set of initial states to a new set of states that satisfy both the constraints imposed by K and those imposed by K′. In symbols, and using Muskens' linear notation for DRSs according to which [$u_1, \ldots, u_n \mid \varphi_1, \ldots, \varphi_m$] is the same DRS as

$u_1, \ldots, u_n$
$\varphi_1, \ldots, \varphi_m$

,

this 'relational' interpretation of (2) can be specified in semi-formal terms as in (5), whereas the compositional interpretation of (3) is as in (6), in which the ';' operator is used. In (5), the denotation of a DRS is specified as a relation between states, i.e., a set of pairs $\langle$i,j$\rangle$ of states. Observe that in (6) each sentence gets an independent interpretation (a DRS) and these interpretations are then composed together.

(5) $[\![$[*x*,*w*|**engine**(*x*),**Avon**(*w*),**at**(*x*,*w*)] $]\!]$ = {$\langle$i,j$\rangle$ | *j* differs from *i* at most over *x* and *w*, and the values assigned by *j* to *x* and *w* satisfy $[\![$**engine**(*x*)$]\!]$, …, $[\![$**at**(*x*,*w*)$]\!]$}

(6) [*x*,*w*|**engine**(*x*), **Avon**(*w*),**at**(*x*,*w*)]; [*y*,*u*|**boxcar**(*y*),**hooked-to**(*u*,*y*), *u* **is** *x*]

Very briefly, Muskens' proposal is as follows. He arrives at a compositional formulation of the DRS construction algorithm by interpreting DRSs and conditions as expressions of a special form of type theory that includes, in addition to the two primitive types of Montague's Intensional Logic [Montague, 1973] *e* ('entities') and *t* ('truth values'), two new primitive types: the type *s* of states, and the type π of discourse referents. Muskens proposes to use constants of type π as the interpretation of noun phrases. He assumes a constant **V** of type $\langle\pi,\langle s,e\rangle\rangle$, i.e., denoting a function from discourse referents and states to entities; this function specifies the object associated with discourse referent *d* at state *i*. Muskens also introduces a relation $i[u_1, \ldots, u_n]j$ which holds between states *i* and *j* if *j* differs from *i* at most over the values assigned to discourse markers $u_1, \ldots, u_n$.

- $i[u_1, \ldots, u_n]j$ is short for $\forall v\ (u_1 \neq v \wedge \ldots \wedge u_n \neq v) \rightarrow (\mathbf{V}(v, i) = \mathbf{V}(v, j))$
- $i[]j$ is short for $\forall$v $\mathbf{V}(v,i) = \mathbf{V}(v,j)$.

We will see below that it is this relation that specifies the crucial ‘update’ aspect of the interpretation of DRSs.

The type-theoretic interpretation of the constructs of DRT is specified as follows: let K and K$'$ be DRSs, *i* and *j* variables ranging over states, and $\varphi_1,\ldots,\varphi_m$ expressions of type $\langle s,t \rangle$. Then

$\mathbf{R}\{\tau_1,\ldots,\tau_n\}$	is short for	$\lambda\, i.\ \mathbf{R}(\tau_1)\ldots(\tau_n)$
τ_1 **is** τ_2		$\lambda\, i.\ (\tau_1) = (\tau_2)$
not(K)		$\lambda\, i.\ \neg\exists\, j\, K(i)(j)$
K **or** K$'$		$\lambda\, i.\ \exists\, j\ \ K(i)(j) \vee K'(i)(j)$
$K \Rightarrow K'$		$\lambda\, i.\ \forall \mathrm{j}\ \ K(i)(j) \rightarrow \exists k K'(j)(k)$
$[\mathrm{u}_1, \ldots, \mathrm{u}_n \mid \varphi_1, \ldots, \varphi_m]$		$\lambda\, i.\ \ \lambda\, j.\ \ i[\mathrm{u}_1, \ldots, \mathrm{u}_n]j \wedge \varphi_1(j), \ldots, \varphi_m(j)$
$K\,;K'$		$\lambda\, i.\ \ \lambda\, j.\ \exists\, k\ \ K(i)(k) \wedge K'(k)(j)$

Note that a DRS is interpreted as a function from pairs of states onto truth values. A DRS K with discourse referents $\mathrm{u}_1 \ldots \mathrm{u}_n$ and conditions $\varphi_1 \ldots \varphi_m$ is true at state *i* iff there is a state *j* such that $\langle \mathrm{i,j} \rangle$ is in the denotation of K, i.e., such that *j* agrees with *i* over all discourse referents other than $\mathrm{u}_i \ldots \mathrm{u}_n$, and all of φ_l hold at *j*. Muskens' Unselective Binding Lemma asserts that these definitions yield the right semantics for DRSs (e.g., they assign existential force to a DRS like (2)); his Merging Lemma ensures that the DRS for a discourse can be composed piecemeal from the DRSs of single sentences and ‘;’.

Having reinterpreted the constructs of DRT in terms of a logic like the one used by Montague, it is then rather simple for Muskens to specify the translation from a fragment of English into DRT which works much as Montague's own, and in which each word is assigned a lexical semantics that is composed with the semantics of other words to obtain the semantic interpretation of sentences. For example, the semantic interpretation of *There is an engine at Avon* in (2) is obtained by means of the following rules of lexical interpretation, in which it is assumed that *u* is a new discourse referent:

an	$\leadsto$	$\lambda\, P.\ \ \lambda\, Q.\ \ [u|\ ];P(u);Q(u)$
engine	$\leadsto$	$\lambda\, x.\ [\ |\mathbf{engine}(x)]$
Avon	$\leadsto$	$\lambda\, P.\ [u|\mathbf{Avon}(u)];P(u)$
is	$\leadsto$	$\lambda\, P.\ \lambda\, x.\ P(x)$
at	$\leadsto$	$\lambda\, x.\ \lambda\, y.\ [\ |\mathbf{at}(y,x)]$

and the following rules of interpretation for derivation trees, where X$'$ indicates the translation of the constituent of category X:

$$
\begin{array}{lcl}
\mathrm{NP} \rightarrow \mathrm{Det\ N} & \leadsto & \mathrm{Det}'(\mathrm{N}') \\
\mathrm{NP} \rightarrow \mathrm{PN} & \leadsto & \mathrm{PN}' \\
\mathrm{PP} \rightarrow \mathrm{P\ NP} & \leadsto & \lambda\, x.\ \mathrm{NP}'(\lambda\, y.\ \mathrm{P}'(y)(x)) \\
\mathrm{S} \rightarrow \mathrm{NP\ VP} & \leadsto & \mathrm{NP}'(\mathrm{VP}') \\
\mathrm{S} \rightarrow \textit{there}\ \mathrm{V[be]\ NP\ PP} & \leadsto & \mathrm{NP}'(\mathrm{PP}')
\end{array}
$$

A proof theory for the language with DRSs can also be derived from the proof theory of the underlying type theory.[5]

3 Conversational Acts and The Discourse Situation

3.1 Pragmatic Information in the Common Ground

By modeling the common ground as a root DRS we can capture two aspects of the participants' shared knowledge: the antecedents made available during a conversation, and the propositions which have been asserted. Not all contributions to a conversation are assertions, however—one can have questions or instructions, for example—and anyway assertions contribute to the common ground more than their propositional content. In Stalnaker's words,

> The fact that a speaker is speaking, saying the words he is saying in the way he is saying them, is a fact that is usually accessible to everyone present. Such observed facts can be expected to change the presumed common background knowledge of the speaker and his audience in the same way that any obviously observable change in the physical surroundings of the conversation will change the presumed common knowledge. ([Stalnaker, 1979], p. 323)

The view that utterances are observable actions (SPEECH ACTS) whose occurrence is recorded by both participants—as formulated, for example, in Austin and Searle's influential work [Austin, 1962; Searle, 1969]—has been the basis of most work on context in AI [Cohen and Perrault, 1979; Allen and Perrault, 1980; Grosz and Sidner, 1986; Carberry, 1990; Cohen *et al.*, 1990]. In this work, speech acts are seen as actions capable of modifying the mental state of the participants in a conversation; theories of intention recognition such as those proposed by Allen, Carberry, Cohen, Levesque, Perrault, and others in the works mentioned are formulated as theories of what we can infer as we observe a speech act.

[5] For a proof theory working directly off the language of DRT, see [Kamp and Reyle, 1991].

Are these two aspects of the common ground—the characterization of anaphoric information specified in DRT, and the information used to infer intentions—distinct? And if they are, is information about speech acts and intentions used together with information about accessible referents and shared beliefs? There is a sense in which the anaphoric and intentional aspects of the common ground are distinct: while the interpretation of a pronoun affects the truth conditions of a text, the fact that ′speaker A told B that P′ is not part of the truth conditions of sentence P.[6] The goal of DRT is to capture the truth conditions of a text, seen as a sequence of assertions; assuming that information about the occurrence of speech acts is part of the common ground amounts to a shift from modeling sentence *meaning*, as in DRT, to modeling (utterance) *use*. So, is it necessary to consider this pragmatic information if all we are interested in is reference resolution? And conversely, should we worry about anaphoric relations if all we are interested in is the process by which a conversant decides what to say next?

The answer to the second question seems clearly to be yes: we do need to worry about the meaning of what was said to decide how to reply. More interestingly perhaps, the shift to a pragmatic model of the common ground is necessary even if we are only interested in how people understand anaphoric expressions. It has long been known that non truth-conditional information plays a role in referent identification; Grosz [1977], Reichman [1985], Fox [1987] and Grosz and Sidner [1986], among others, showed convincingly that 'pragmatic accessibility' or 'segmentation' effects on reference are inextricably tied with (discourse) intentions and their structure. An example of segmentation can be seen in the fragment of TRAINS conversation d91-6.2 in (7) below, which contains two uses of the definite description *the boxcar*: one at utterance 14.2, the other at utterance 31.2. These two definite descriptions do not refer to the same object, but as the conversants are engaged in different tasks at the two points in time, they do not perceive an ambiguity and have no difficulty in finding the correct antecedent each time.

(7)

	…
13.3	We're gonna hook up engine E2 to the boxcar at Elmira,
13.4	and send that off to Corning
13.5	now while we're loading that boxcar with oranges at Corning,
13.6	we're gonna take the engine E3
13.7	and send it over to Corning,
13.8	hook it up to the tanker car,
13.9	and send it back to Elmira

[6] For a discussion of the problems with the so-called PERFORMATIVE ANALYSIS of sentences, which makes the illocutionary force of an utterance part of its truth conditions, see [Boër and Lycan, 1980].

14.1 S: okay
14.2 We could use one engine to take both the tanker
and the boxcar to Elmira
…
29.3 while this is happening,
29.4 take engine E1 to Dansville,
29.5 pick up the boxcar,
29.6 and come back to Avon
30.1 S: okay
31.1 U: okay
31.2 then load the boxcar with bananas

If we accept Clark and Marshall's claim that whether referring expressions are felicitous depends only on shared information, we are forced to conclude that information about the task structure and how specific utterances are related to it must be part of the common ground. The common ground must therefore include pragmatic information in addition to the truth-conditional information captured by DRT. We will show, however, that the context description tools introduced in DRT can be adapted to the purpose of formalizing a model of language interpretation; our model of the common ground can thus be seen as a generalization of the models of the common ground used in formal semantics.[7]

3.2 Conversation Acts

Many theories of discourse structure have been proposed in the literature. We will adopt a speech act-based account; as we will see in the next sections, a theory of this kind gives us the tools to account not only for the organization of dialogs according to the domain task just discussed, but also for other kinds of structure observable in spoken dialogs, such as the structure of turn-taking and the structure of grounding.

[7] Although we are only interested here in the common ground insofar as interpretation is concerned, there are reasons to doubt that a purely truth-conditional approach is adequate even for the purpose of accounting for the semantics of sentences. A number of expressions can only be interpreted by taking into account that an utterance took place. First of all, there are a number of utterances which do not have any truth-conditional impact, and whose meaning, therefore, can only be explained within a pragmatically based theory of the common ground: for example, the DISCOURSE MARKERS–*right*, *okay*, etc. Indexicals like *I* or *you* are examples of referring expressions whose meaning depends on pragmatic factors; others are expressions like *the former*, *the latter*, or *vice versa*, as well as any expressions in a text that refer to parts of that text—*the following table*, *the list below*, etc. As well, a number of adverbs and adjectives—*frankly*, *in my opinion*, etc.—can only be interpreted with respect to aspects of the discourse situation such as the conversants' opinions. These phenomena all indicate that a generalisation of the notion of common ground like the one proposed here is needed. In fact, the central idea of dynamic semantics—that 'informational update' is the main role of sentences—is already a generalisation in the direction we propose.

In addition, we will also see that information about speech acts is a crucial ingredient of accounts of interpretation that take into account the fragmentary nature of spoken input.[8]

Most classic theories of speech acts concentrate on the actions performed by the conversational participants as a way of 'getting the job done'—e.g., instructions to the other conversant, requests for information necessary to accomplish the task, etc. But these actions are only a part of what happens in conversations; the conversants spend a lot of their time making sure they do not talk over each other and ensuring that 'informational' coordination is achieved. Recent theories of speech acts (e.g., [Novick, 1988; Kowtko *et al.*, 1992; Traum, 1994; Bunt, 1995]) are built on the assumption that a good theory of the actions involved in these aspects of a conversation is as important to a system as a good theory of task-oriented acts.

Following the implemented TRAINS-93 system, we adopt here the multi-level CONVERSATION ACTS theory, presented in [Traum and Hinkelman, 1992]. This theory maintains the classical illocutionary acts of speech act theory (e.g., **inform**, **suggest**), now called CORE SPEECH ACTS. These actions are, however, reinterpreted as multi-agent collaborative achievements, taking on their full effect only after they have been *grounded*, i.e., acknowledged (see Section 6, below). Rather than being actions performed by a speaker to a hearer, the core speech acts are joint actions; the initial speaker and the hearer (called hereafter *initiator* and *responder*, respectively) each contribute actions of a more basic type, the result being the common ground assumed to be the effects of core speech acts.

In addition, Conversation Acts (CA) theory also assumes that three other kinds of speech acts are performed in conversations: acts for TURN-TAKING, GROUNDING, and more complex acts called ARGUMENTATION ACTS that involve more than one core speech act—for example, to perform an elaboration. The four kinds of acts of CA theory are displayed in Table 1. The acts from top to bottom are typically realized by larger and larger chunks of conversation: from turn-taking acts usually realized sub-lexically, to grounding acts which are realized within a single utterance unit (UU),[9] to core speech acts which are only completed at the level of a completed discourse unit (DU),[10] to argumentation acts which can span whole conversations.

[8]The DIT model being developed by Bunt [1995] is also based on speech acts. This dependence on speech acts is the main difference between our model and the STDRT model of context developed by Asher, Lascarides, Oberlander and others more or less in parallel with our work on TRAINS (see, e.g., [Lascarides and Asher, 1991; Asher, 1993]). One reason for the difference is that Asher *et al.* are concerned with texts rather than conversations.

[9]Utterance Units roughly correspond to intonation phrases, although long pauses are also taken as unit boundaries. See [Traum and Heeman, 1996] for an empirical investigation of the appropriate utterance unit boundaries for grounding.

[10]Discourse Units are discussed further in Section 6.

The table also shows some representative acts for each class.

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**, **ynq**, **check**, **eval** **suggest**, **request**, **accept**, **reject**
Multiple DUs	Argumentation	**elaborate**, **summarize**, **clarify** **q&a**, **convince**, **find-plan**

Table 1: Conversation Act Types

We will discuss grounding acts and argumentation acts in the following sections. Some of the core speech acts used in TRAINS-93 are characterized as follows:

inform Initiator presents responder with new information in an attempt to add a new mutual belief.

ynq Initiator asks responder to provide information that initiator is missing but suspects that responder may know; imposes a discourse obligation [Traum and Allen, 1994] on responder to evaluate and respond to the request.

check Like a **ynq**, but initiator already suspects the answer; initiator wants to move the proposition in question from individual to mutual belief, or bring the belief into working memory.

eval An evaluation by the initiator of the "value" of some (physical or intentional) object.

suggest Initiator proposes a new item as part of a plan.

request Like a **suggest**, but also imposes a discourse obligation to respond.

accept Initiator agrees to a proposal by responder.

reject Initiator rejects a proposal by responder.

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**, **assign-turn** and **take-turn**. These are recognized when one conversant desires to keep speaking, stop speaking, get another conversant to start, or to start talking. Usually single words or tunes in the speech stream are enough to signal one of these acts.

In addition to these four sets of illocutionary and perlocutionary acts,[11] there is also a class of *locutionary acts*, consisting of the single act **utter** of "uttering a sound". (We also refer to acts of this type as "surface speech acts" or simply "utterances".) Locutionary acts will be discussed in Section 5. Austin distinguished several levels of these acts, including the *phonetic act* (making certain noises), the *phatic act* (uttering words and constructions that are part of a specific lexicon / grammar), and the *rhetic act*, (using that construction with a definite sense and reference). We will see in Section 5 how our treatment of locutionary acts relates to this classification.

We follow Goldman [1970] in positing a *generation* relationship between these various acts. The more conventional and intentional level acts are conditionally generated by the performance of appropriate acts at lower levels, given the proper context. Locutionary acts generate the four levels of conversation acts. While argumentation acts are in turn generated by (sequences of) core speech acts, there is no such relationship between, e.g., grounding acts and core speech acts. Both are generated by the corresponding locutionary acts.

3.3 Discourse Situation and Described Situation

Our unification between DRT and speech act-based models for user modeling and dialogue management is rooted in ideas about the common ground developed in Situation Semantics [Barwise and Perry, 1983; Devlin, 1991]. Situation Semantics is based on a theory of information according to which what we know is organized in SITUATIONS—'chunks' of facts and objects. In particular, it is assumed that the common ground of a conversation includes shared information about the DISCOURSE SITUATION, which is the situation that the participants in a conversation find themselves in. As the title says, our theory is a theory of the effect of

[11] Austin [1962] distinguished between illocutionary acts which were conventional expressions of the initiator's intentions, and perlocutionary acts which are not completely in the initiator's control to perform, such as "convince", in which the hearer's mental state must change for the act to occur. Turn-taking and argumentation acts have more of the flavour of perlocutionary acts, while grounding and core speech acts have more of the flavour of illocutionary acts. However, since we now recognize that even core speech acts require participation of the responder (at least to the extent of registering, understanding, and acknowledging them), we have muddied rather than clarified the illocutionary/perlocutionary distinction.

conversational acts on discourse situations.[12]

The discourse situation includes the (speech) actions the agents have performed, as well as information about their mental states, such as information about their beliefs, intentions, and their VISUAL SITUATION—i.e., what they can see around them.[13] The discourse situation also includes information about one or more DESCRIBED SITUATIONS: these are the situations that the conversants talk about. Although in simple cases the discourse situation and the described situation are the same, this is not true in general. The participants in a TRAINS conversation, for example, may discuss the TRAINS world, a simplified abstraction of reality consisting of information about towns, railways, and available engines; or they may talk about the actions included in the domain plan they are developing, and about the state of the world resulting from these actions; or indeed they may talk about the discourse situation. It is important to keep this information distinct, as what is true in one situation may not be true in others. For example, in the TRAINS world there is an orange juice factory at Elmira, while there is none in the real city of that name in western New York; it takes 5 hours to go from Avon to Bath in the TRAINS world, while in reality it only takes 2 hours; and so forth. This suggests that we want to keep the 'real world' and the 'discourse situation' separate from the TRAINS world represented on the map used by the conversants that we study.

In fact, more than one described situation can be discussed in a dialogue. The participants in a TRAINS conversation often talk about the state of the world resulting from the execution of some steps in the plan. In the following fragment of dialogue 91-6.2, for example, the Manager is talking about the situation resulting from a previously planned action of moving engine E2 from Elmira to Dansville after hooking it to a boxcar:

(8)	135.2	M:	take the boxcar
	135.3	:	that's hooked up to engine E2
	135.4	:	which came from Elmira
	135.5		and is at / now at Dansville

Engine E2 is not at Dansville in the TRAINS world at the time the Manager is speaking: the system has to realize this, or else it would establish a goal of correcting the

[12] It is important to keep in mind that information about the discourse situation is only a part of the common ground between the participants in a conversation, which consists of all the information that they assume to share, including general information about, say, the town they live in, the organization of the society in which they live, etc. We will not consider these other aspects of the common ground; see, e.g., [Clark, 1996] for a preliminary discussion.

[13] See [Poesio, 1993] for an account of how visual information is used to resolve certain cases of definite descriptions.

Manager (and generate *there is no engine at Dansville right now*).[14]

3.4 Speech Acts and Dynamics

The way in which we obtain a theory of the common ground that accounts both for the effect of utterances on anaphoric accessibility and for the role of pragmatic information about speech acts is by reinterpreting a DRT-style representation of the common ground. Whereas in DRT the root DRS specifies conditions on the described situation, we use it as a representation of the discourse situation, in the sense that the conditions in the root DRS specify what speech acts took place and what effects they have on the mental state of the participants. The content of speech acts, in turn, specifies the assertions made about one or more described situations.

In doing so we had to address issues of both a conceptual and a technical nature. At the very least, it is necessary to make sure that our model of context still makes discourse referents accessible for anaphoric reference. Under the standard semantics for DRT, the interpretation of the illustrative mini-dialog in (9) between conversants A and B represented by the DRS in (10) would not make the discourse referent x 'evoked' by the NP *an engine* accessible to the pronoun *it* in the next sentence, represented in (10) as the discourse referent u.

(9)
A: There is an engine at Avon.
B: *It* is hooked to a boxcar.

(10)

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

[14] The fact that the participants in the TRAINS conversations typically discuss alternative way to achieve goals may be considered additional (if controversial) evidence for the claim that more than one described situation may be discussed in a conversation, as each alternative subplan might be considered a separate, 'possible' situation. The ontological status of subplans is rather unclear, however (for discussion, see [Poesio, 1994; Traum *et al.*, to appear 1996]). Incidentally, this is an example of the difference between a theory based on linguistic treatments of anaphora and one that simply assumes that all referents are added to a 'focus space'—the former makes distinctions which are much more fine-grained than those available in the latter.

A second problem with an interpretation like (10) is that it doesn't give us the information that both assertions are about the same described situation. In 'vanilla' DRT the semantics of predicates like **inform**, one of whose arguments is a DRS, can be specified in one of two ways. One possibility is to treat **inform** as an extensional predicate: that is, to evaluate the embedded DRS with respect to the same situation in which the condition asserting the occurrence of the telling event is evaluated. But in this way facts about what is going on in the discourse situation would be mixed with facts about the described situation: under this interpretation, (10) 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. As discussed in the previous section, the described situation and the discourse situation do not always coincide in the the TRAINS conversations.

Alternatively, **inform** 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 **inform** relation to be evaluated at a situation determined by the modality and its first two arguments. In other words, we could treat all cases of pronominal reference in a dialogue as instances of MODAL SUBORDINATION [Roberts, 1989]. But, as discussed in the previous section, different speech acts may be about different described situations, which is a problem for this solution.

Our solution to the problem is based on ideas proposed in [Poesio, 1994]. The theory proposed there is based on the treatment of tense, aspect and event anaphora originated by Davidson [1967] and adopted with some variations in DRT and other semantic theories such as Episodic Logic [Hwang, 1992; Hwang and Schubert, 1993]. To explain the observation that it is possible to refer anaphorically not only to ordinary individuals, but also to events[15] and other abstract entities—as in *A: we sent engine E1 to Avon. THAT happened three hours ago.*— Davidson proposed that 'the logical form of action sentences' involves an extra argument denoting the event of performing that action, which is made available for subsequent reference. In DRT, this idea is implemented by assigning to the text in (3) the interpretation in (11). This interpretation, crucially, contains conditions of the form $s : \varphi$, where s is an event or state and φ is a characterization of that state; and the object s is available for subsequent reference [Kamp and Reyle, 1993].

[15] We will use the term 'event' here as synonymous with 'situation' and 'episode'.

(11)
$$\begin{array}{|l|}\hline x\ w\ y\ u\ s\ s' \\ \hline \mathbf{engine}(x) \\ \mathbf{Avon}(w) \\ s : \boxed{\mathbf{at}(x,w)} \\ \\ \mathbf{boxcar}(y) \\ s' : \boxed{\mathbf{hooked\text{-}to}(u,y)} \\ u\ \mathbf{is}\ x \\ \hline \end{array}$$

Now, *conversational* events, such as the occurrence of conversational actions, may serve as antecedents for anaphora as well, as in *A: we need to send an engine to Avon. B: is THAT a suggestion?* This observation led to the first ingredient of the proposal in [Poesio, 1994], namely, the hypothesis that the events that take place in the discourse situation leave a 'trace' in the form of discourse referents just as the events that take place in a described situation do.

The second ingredient of that theory is a theory of update in which it is entire situations that get updated, rather than just assignments, and the dynamic properties of utterances are characterized in terms of (changes to) described situations. In [Poesio, 1994], utterances are transitions from discourse situations to discourse situations, such that the discourse situation resulting from an utterance includes additional conversational events. The propositional content of a conversational event is specified by a DRS which, instead of being interpreted as a relation between assignments as in Muskens' [1994] system, is interpreted as a relation between situations—a transition between the described situation of a previous conversational event and a new described situation. For example, the dialog in (9) results in a discourse situation that includes at least the information in (12).

(12)
$$\begin{array}{|l|}\hline ce1\ \ ce2\ \ s\ s'\ s'' \\ \hline ce1 : \mathbf{inform}(A,B, \begin{array}{|l|}\hline x\ w\ e \\ \hline \mathbf{engine}(x) \\ \mathbf{Avon}(w) \\ e : \boxed{\mathbf{at}(x,w)} \\ \hline \end{array} (s)\,(s')) \\ ce2 : \mathbf{inform}(B,A, \begin{array}{|l|}\hline y\ u\ e' \\ \hline \mathbf{boxcar}(y) \\ e' : \boxed{\mathbf{hooked\text{-}to}(y,u)} \\ u\ \mathbf{is}\ x \\ \hline \end{array} (s')\,(s'')) \\ \hline \end{array}$$

The DRS in (12) represents a discourse situation in which two conversational events occurred, *ce1* and *ce2*. The first two conditions assert that *ce1* is an event of A informing B that the described situation of *ce1*, s', extends a previous situation s, and includes an engine located at Avon. Note that DRSs are still interpreted as relations between state-like objects, i.e., as objects that give you a proposition when applied to two state-like objects, just as in Muskens' system discussed in section 2.2—the difference is that now a DRS denotes a relation between two situations. The next two conditions in (12) assert that *ce2* is an event of B informing A that the described situation s'' also contains a boxcar y, hooked to u, which denotes the same object as x. Note how events in the described situation and in the discourse situation are characterized in the same fashion.[16]

The required dynamics of discourse referents comes about as the result of (i) interpreting DRSs as relations between situations of which the second (the described situation proper) is seen as an informational extension of the first and (ii) making the two situations involved in the characterization of the content of a conversational event like *ce1* globally accessible discourse referents. In (12), for example, the described situation s'' of the conversational event *ce2* coming after *ce1* is an extension of the described situation s' of *ce1*, hence it contains all of the constituents of s'. As we will see below, this way of looking at how discourse referents are made available is closely related to the 'focus space stack' idea of Grosz and Sidner.

These ideas have been implemented by modifying the logic TT_2^4, a version of Situation Theory proposed in [Muskens, 1989], to which we added some of the ideas from [Muskens, 1994] discussed above.[17] Our extension to Muskens' TT_2^4 logic is discussed in some detail in Appendix A. This is a partial typed logic, based on a generalization of Montague's system of types. What is most important here is that, first, Muskens' situations behave like $\langle$world,time$\rangle$ pairs, in the sense that on the one hand they have the property of supporting certain facts, characteristic of worlds; on the other hand, they are temporally ordered by a precedence relation '$<$'. Secondly, we can define a new type of condition, written '$s : \varphi$', and stating that situation s is of type φ; this construct is analogous to constructs used in theo-

[16] A dynamics of situations is needed for more complex cases of anaphora, including so-called 'bridging' references and 'result-state' anaphora [Poesio, 1994]. For example, the action of mixing water with flour originates a situation which includes an additional object that may serve as antecedent for definite descriptions such as *the dough* even though it hasn't been explicitly mentioned. A similar claim that a situational view on dynamics offers a more general theory of anaphoric phenomena is in [Milward, 1995].

[17] The theory in [Poesio, 1994] was based on Episodic Logic, a version of Situation Theory with many points in common with Muskens' proposal, but much richer in many respects [Hwang, 1992; Hwang and Schubert, 1993]. Here we use Muskens' theory as it is simpler and has a clearer connection with the kind of logics typically used in semantics, such as Montague's IL.

ries of events embedded in DRT and Situation Theory. Third, Muskens defines an inclusion relation between situations, '$\subseteq$', such that $s \subseteq s'$ iff the domain of s is a subset of the domain of s', and anything which is definitely true or definitely false at s preserves its truth value at s'. We use the inclusion relation to model information growth. Finally, this new semantics still allows us to build our interpretation of the discourse situation incrementally, by merging together the interpretations of single utterances. Thus, (13) is equivalent to (12). We will discuss how (13) is obtained below, when discussing interpretation.[18]

(13)
$$\begin{array}{|l|}\hline ce1\ \ s\ s' \\ \hline ce1 : \ \textbf{inform}(A,B, \begin{array}{|l|}\hline x\ w\ e \\ \hline \textbf{engine}(x) \\ \textbf{Avon}(w) \\ e : \ \boxed{\textbf{at}(x,w)} \\ \hline \end{array} (s)\,(s')) \\ \hline \end{array} ; \begin{array}{|l|}\hline ce2\ \ s'' \\ \hline ce2 : \ \textbf{inform}(B,A, \begin{array}{|l|}\hline y\ u\ e' \\ \hline \textbf{boxcar}(y) \\ e' : \ \boxed{\textbf{hooked-to}(y,u)} \\ u\ \textbf{is}\ x \\ \hline \end{array} (s')\,(s'')) \\ \hline \end{array}$$

3.5 Mental States

Facts about the mental states of agents play an important role in speech act recognition, reference resolution, and dialog management. Information about the (mutually known) intentions, beliefs, perceptual input, and obligations [Traum and Allen, 1994] of the conversants is also part of the discourse situation. This information can be represented in the language introduced in the previous section by conditions of the form s : K, asserting that STATE s of type K is part of the discourse situation, where a state is a particular type of situation with different properties from events (see, e.g., [Kamp and Reyle, 1993] for a characterization of states and events). 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. The occurrence of this state in a discourse situation can be represented

[18] Whereas the third argument of the illocutionary act **inform** is a proposition of the form $\mathbf{K}(s,s')$, the third argument of the illocutionary act **y-n-q** 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 [1977] and 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'. Ginzburg has been developing a model of the discourse situation motivated by work on the semantics of questions [1995a; 1995b].

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

(14)
$$\begin{array}{|l|}\hline \ldots s' \mathit{i1}\ \mathit{tj} \\ \hline \ldots \\ \mathit{i1} :\ \mathbf{intend}(A,\, s' :\ \boxed{\mathbf{at}(y, Bath)}) \\ \hline \end{array}$$

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, 1993]). We would 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,φ) in the root DRS corresponds to a fact of the form **bmb**(A,B,**believe**(A,φ)) in Cohen and Levesque's [1990] formalism.[19]

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, such as which conversant has the turn or initiative at a given time, can be represented in a similar manner.[20]

4 Discourse Structure and Event Structure

We now turn to the task of developing a theory of discourse structure—or, more precisely, of incorporating into our theory of the discourse situation the main features of the better known and (arguably) most influential account of discourse structure, Grosz and Sidner's theory [1986]. Grosz and Sidner's discourse model involves three distinct, but interrelated components: an INTENTIONAL STRUCTURE

[19] 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,φ)).

[20] For a different view of how information about belief and intentions could be incorporated in DRT, see [Kamp, 1990].

consisting of the goals of the interlocutors and the relations among these; an ATTENTIONAL STATE specifying which entities are most salient at a given point in the discourse; and a LINGUISTIC STRUCTURE which consists of syntactic information arranged in discourse segments. We intend to demonstrate that it is possible to capture Grosz and Sidner's intuitions by using fewer technical tools than they use, and in particular, that the properties of the attentional state they explain by assuming a focus space stack follow from the ideas about speech acts and dynamics discussed in section 3, once we adopt Grosz and Sidner's hypotheses about the hierarchical structure of discourse intentions.

4.1 Intentional Structure and the Hierarchical Structure of Events

Grosz and Sidner (henceforth: G&S) assume that the process of speech act recognition does not simply result in the recognition of the illocutionary acts generated by an utterance; it also relates these acts to the intentions expressed by previous acts. In their theory, utterances (more precisely, discourse segments) are associated with a DISCOURSE SEGMENT PURPOSE (DSP), and these purposes are related in two different ways: by DOMINANCE (when a purpose is part of the achievement of another) and SATISFACTION-PRECEDENCE (when achieving one purpose is a prerequisite for achieving the other).

In Conversation Acts theory, utterances (locutionary acts) contribute to the generation of one or more illocutionary act: these may include turn-taking acts, grounding acts, and core speech acts. I.e., where G&S would talk about the initiator's discourse segment purpose in performing a locutionary act, we will talk about the initiator performing an act generated by the locutionary act. Ultimately we accept G&S's argument that intentions are more 'basic' than acts; we also believe, however, that some discourse purposes occur more frequently than others, and it is therefore likely that they get 'compiled,' i.e., a responder gets to recognize what the initiator is doing without reasoning about the initiator's intentions. A classification of the most common 'acts' in a certain kind of conversations is therefore of great importance when building a system that has to engage in such conversations. We do allow for the possibility of a conversant having intentions than cannot be reduced to (sets of) conversational acts; provided that they get recognized, such intentions could still become part of the discourse situation, as discussed at the end of the previous section.

Much as G&S assume that discourse purposes are related to higher discourse purposes, we assume that conversational acts are related to other conversational acts, as well as to higher-level actions, realized by way of multiple core speech acts and not associated with any utterance in particular. We call these more complex acts

CONVERSATIONAL THREADS; they will be discussed in more detail below.[21] We define relations between conversational actions that reflect the underlying relations between the intentions associated with these actions; thus, we write $\alpha \uparrow \beta$ to indicate that action α is (immediately) dominated by action β, and $\alpha \prec \beta$ to indicate that α (immediately) satisfaction-precedes β. We indicate the transitive closures of these relations by $\uparrow^*$ and $\prec^*$, respectively: e.g., $\alpha \uparrow^* \beta$ iff there is an action δ such that $\alpha \uparrow \delta$ and $\delta \uparrow^* \beta$. In fact, it is these extended relations that correspond more closely to Grosz and Sidner's notion of dominance and satisfaction-precedes.[22]

For example, the interpretation of the two utterances *There is an engine at Avon. It is attached to a boxcar* in which the DSP of the first utterance satisfaction-precedes the DSP of the second would be captured by the following description of the discourse situation, in which $\prec$ holds between the core speech act *ce1* and the core speech act *ce2*:

(15)

$$
\begin{array}{|l|}
\hline
ce1\ ce2\ aa1\ s\ s'\ s'' \\
\hline
ce1 : \mathbf{inform}(A,B, \begin{array}{|l|}\hline x\ w\ e \\ \hline \mathbf{engine}(x) \\ \mathbf{Avon}(w) \\ e : \boxed{\mathbf{at}(x,w)} \\ \hline \end{array} (s)\ (s')) \\
ce2 : \mathbf{inform}(B,A, \begin{array}{|l|}\hline y\ u\ e' \\ \hline \mathbf{boxcar}(y) \\ e' : \boxed{\mathbf{hooked\text{-}to}(y,u)} \\ u\ \mathbf{is}\ x \\ \hline \end{array} (s')\ (s'')) \\
ce1 \prec ce2 \\
\hline
\end{array}
$$

4.2 Attentional State

The second component of Grosz and Sidner's model of discourse is the attentional state. According to G&S, the relative salience of discourse referents is determined by their position in the FOCUS SPACE STACK, a separate component of the discourse model. The discourse referents associated with utterance *u* become part of a FOCUS

[21] We mentioned above how in Conversation Act theory a core speech act is usually the result of a joint effort by the conversants which may involve more than one turn, as discussed below. So in fact even our core speech acts already express 'higher-level discourse purposes'.

[22] Note that '$\uparrow$' is not the same relation as '$\subseteq$'. The former relation has to do with event decomposition, and only applies to events; the latter with informational inclusion, and applies to every situation.

SPACE, a collection of objects and properties which is pushed on the stack on top of the focus spaces associated with discourse segment purposes that dominate or satisfaction-precede *u*. Only the discourse referents currently on the stack are accessible, those closest to the top being more accessible than those below them.[23]

The rules of pragmatic accessibility that G&S attribute to the presence of a stack in the discourse model can be derived in our model without this additional stipulation, simply because the content of each speech act is a statement about a described situation, and situations are hierarchically organized. All that is needed is what we take to be the crucial part of Grosz and Sidner's theory, namely, the hypothesis that the attentional state is parasitic on the intentional structure. We need to assume, that is, that if *ce1* $\uparrow$ *ce2*, then the described situation of *ce1* is included in the described situation of *ce2*; whereas if *ce1* $\prec$ *ce2*, the described situation of *ce2* extends the described situation of *ce1*. More formally put, if **pred** and **pred**$'$ are predicates that characterize core speech acts, then

(16) a. $\forall$e,e$'$,a,b,c,d,s_1,s_2,s_3,s_4,
$e : \mathbf{pred}(a,b,\varphi(s_1)(s_2)) \wedge e' : \ \mathbf{pred}'(c,d,\varphi'(s_3)(s_4)) \wedge e \uparrow e' \rightarrow s_2 \subseteq s_4$

b. $\forall$e,e$'$,a,b,c,d,s_1,s_2,s_3,s_4,
$e : \mathbf{pred}(a,b,\varphi(s_1)(s_2)) \wedge e' : \ \mathbf{pred}'(c,d,\varphi'(s_3)(s_4)) \wedge e \prec e' \rightarrow (s_2 \ \mathbf{is} \ s_3)$

We also assume that the discourse situation contains at each point in time information about which described situation contains at time *t* the information that would be contained in the focus space stack as a whole: we call this situation DISCOURSE TOPIC. Conditions of the form '*discourse-topic(t)* **is** *s*' specify the discourse topic at time *t*. By default, a conversational event is interpreted as extending the current discourse topic, rather than shifting to a new one or returning to an old one.

To see how these definitions do the work of the stack, consider the example in (15). The third argument of the speech act *ce1* specifies that its described situation, s', extends the situation *s* with a new constituent, *x*, and with the information that this object is an engine. The situation s' has the same function as a focus space in G&S's theory; and the described situation of a speech act in the same discourse segment as *ce1* will be an extension of the described situation of *ce1* much in the same way as it would be if we were to associate a focus space with it and put it on a stack whose top is *ce1*'s focus space, as in G&S's theory. The described situation s'' of the next core speech act in (15), *ce2*, extends the focus space / described situation

[23] G&S's model of the attentional state has many points in common with the model developed by Reichman [1985].

of *ce1*, s', by including new discourse referents, *u* and *y*, and by specifing that *y* is a boxcar, etc.

In general, a conversational event may

1. Describe a situation which does not extend the described situation of any previous conversational event. This case corresponds to the case in G&S's theory in which the stack is restarted and a new focus space added on top of it. If we assume that (15) is a complete description of the discourse situation after the second utterance, hence *s* is not the described situation of any other conversational event, this is what *ce1* does.

2. 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, still allowing access to the previous focus spaces. This is what *ce2* does: its described situation (s'') extends the described situation of *ce1*. 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''.

It should be noted that the model of the attentional state we have just discussed is not strictly equivalent to G&S's. Where they have a single stack, we have here one 'stack' for each discourse segment; this makes our model of the attentional state, strictly speaking, closer to the 'graph-structured' stack proposed by [Rosé *et al.*, 1995] than to G&S's. It should be noted, however, that even the model proposed by Grosz and Sidner is not a stack in a strict sense. Their treatment of interruptions, for example, involves auxiliary devices, such as 'impenetrable barriers' on the stack for what they call 'true interruptions' and 'auxiliary stacks' for flashbacks; in fact, this is what they say of 'impenetrable barriers':

> "This boundary is clearly atypical of stacks. It suggests that ultimately the stack model is not quite what is needed. What structure should replace the stack remains unclear to us." ([Grosz and Sidner, 1986], footnote 12).

In our model, a 'true interruption' would be treated rather simply as a conversational event ce_1 with described situation s_1, followed by conversational events $ce_2 \ldots ce_{n-1}$ whose described situation does not extend s_1, followed by a conversational event ce_n whose described situation is an extension of s_1. As for the auxiliary stack, its task is to store those focus spaces that have to be popped from the stack to insert a new focus space 'in between' existing focus spaces.[24] Such insertions 'in between' in our model simply require revising information about situation transitions: thus, if the propositional content of speech act ce_1 was taken to be a transition between situations s_1 ad s_2, and the propositional content of speech act ce_2 was taken to be a transition between situations s_2 and s_3, we can insert a new focus space mapping s_2 into s_4 in between simply by revising what we know about ce_2 and asserting that its content is a transition between s_4 and s_5. In other words, the auxiliary mechanisms proposed by Grosz and Sidner are unnecessary with the theory of described situations proposed here.

4.3 Conversational Threads, Argumentation Acts, and Discourse Scripts

It is a basic fact about the way humans interpret events that they tend to be grouped into larger 'stories' or, as we will call them here, THREADS [Nakhimovsky, 1988; Webber, 1988; Kameyama *et al.*, 1993]. A thread is itself an event, that decomposes hierarchically into its constituent events [Kautz, 1987]. The hierarchical organization of speech acts into larger units or discourse segments (associated with more general discourse purposes) is just an instance of this more general phenomenon of events being grouped into threads, and the relations between DSPs assumed by Grosz and Sidner are those generally assumed to hold between actions (e.g., in Kautz's theory). We will use the term CONVERSATIONAL THREADS for threads of conversational events when we want to distinguish between this 'technical' notion of discourse segment from the intuitive notion.[25]

Our theory of event structure is fairly standard. As discussed in the previous section, we assume that events can be decomposed into smaller events; the relation between events and the threads of which they are a part of, that we indicated with '$\uparrow$', corresponds to the domination relation in theories such as Kautz's. We also assume that each event in a thread has an immediately preceding and immediately following event: this is what is specified by the '$\prec$' relation. Finally, we assume that the perspective from which we view a thread changes over time, i.e., we assume that each thread has a 'current-event' ('now point') at any time *t*.

[24] This is Grosz and Sidner's method for dealing with flashbacks such as *Whoops I forgot about ABC. I need an individual concept for the company ABC.* in dialogues such as *OK. Now how do I say that Bill is ... Whoops I forgot about ABC. I need an individual concept for the company ABC.* According to them, as the DSP of this utterance satisfaction-precedes the DSP of *Now how do I say that Bill is ...* , the focus space associated with the flashback has to be inserted in the stack before the focus space of *Now how do I say that Bill is ...*

[25] The reason for this term is that, 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.

In Conversation Act theory, certain kinds of threads are singled out. We assume that rhetorical 'relations' such as **elaboration** or **explanation** [Mann and Thompson, 1987] are in fact a particular form of conversation act involving multiple core speech acts, called ARGUMENTATION ACTS. These acts implicitly involve domination, satisfaction-precedence, and other relations between the component events, depending on the type of rhetorical relation.[26] For example, the interpretation of *A: There is an engine at Avon. B: It is attached to a 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*:

(17)
$$\begin{array}{|l|}\hline ce1\ ce2\ aa1\ s\ s'\ s'' \\ \hline ce1 :\ \mathbf{inform}(A,B, \begin{array}{|l|}\hline x\ w\ e \\ \hline \mathbf{engine}(x) \\ \mathbf{Avon}(w) \\ e :\ \boxed{\mathbf{at}(x,w)} \\ \hline \end{array} (s)\ (s')) \\ ce2 :\ \mathbf{inform}(B,A, \begin{array}{|l|}\hline y\ u\ e' \\ \hline \mathbf{boxcar}(y) \\ u\ \mathbf{is}\ x \\ e' :\ \boxed{\mathbf{hooked\text{-}to}(y,u)} \\ \hline \end{array} (s')\ (s'')) \\ aa1 :\ \mathbf{elaborate}(B,A,ce1,ce2) \\ \hline \end{array}$$

More in general, we assume that people know a lot about the structure of certain kinds of threads, and use this information to predict what's going to happen next, as well as to 'fill in' holes in the description. The idea that this information about SCRIPTS—threads whose structure and roles are known in advance, such as the sequence of events that takes place in a restaurant—plays a crucial role in natural language processing was explored in well-known work by Schank and associates [Schank and Abelson, 1977]. Similarly, people know a lot about the organization of events in conversations, both at a broad level (e.g., what generally happens in a conversation) and at a more local level (e.g., what to expect after a question); the so-called micro- and macro-structure of conversations has been studied in the field of CONVERSATION ANALYSIS [Schegloff and Sacks, 1973; Sacks *et al.*, 1974]. The connection between work on scripts and work in conversation analysis was explored in AI work on 'dialogue games' [Power, 1979; Kowtko *et al.*, 1992;

[26] A similar position is taken in recent work on rhetorical structure in the generation field [Moore and Paris, 1993].

Airenti *et al.*, 1993] and 'discourse scripts' [Poesio, 1991a; Turner, 1989].

Our reformulation of Grosz and Sidner's notion of discourse structure in terms of threads establishes an explicit connection between work on intention recognition using expectations and work based on the planning paradigm. A discourse script is simply a particular type of thread; by recognizing the thread of which a certain speech act is a part of, and the current position in that thread (as specified by the 'now' point of that thread), we can use expectations to recognize the type of speech act. Our analysis of the macro-structure of the TRAINS conversations is discussed in [Poesio, 1991a], and was used to implement a speech act analyzer using expectations as well as the syntactic information about the utterance to generate hypotheses about speech acts. What's more, our assumption that conversational events are organized into conversational threads is a more general assumption than Grosz and Sidner's idea that core speech acts are organized in discourse segments, since we allow for threads of turn-taking acts and grounding acts as well. The dialog manager of TRAINS-93 relies especially heavily on expectations in this case. We discuss our theory of expectations in grounding in Section 6, below.

5 Micro Conversational Events and Interpretation

As discussed in section 3.2, we assume with Austin that the conversational acts include the locutionary act of uttering a sound; we also share with Stalnaker the assumption that the occurrence of locutionary acts is recorded in the common ground, and that it is this occurrence that triggers the interpretation processes that lead to recognizing instances of the other classes of conversation acts. In this section we look more closely at locutionary acts and at the interpretation process.

5.1 Utterances in Spontaneous Speech

Spontaneous speech consists for the most part of utterance fragments, rather than full sentences; these fragments are mixed with pauses and other hesitations, with repetitions, and with corrections of what has just been said. The fact that turn-taking acts and grounding acts in between utterance fragments which specify parts of a core speech act is evidence that the common ground is updated before a core speech act is completed. Furthermore, these utterance fragments are not simply recorded in the common ground without being interpreted. Utterance fragments such as *the engine at Avon* in (18) trigger (some aspects of) the interpretation process much as complete speech acts would. S's repair in 10.1 indicates that he has already interpreted *the engine at Avon* in 9.2-9.3, and found that M is mistaken.

(18)
```
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
```

According to the theory of locutionary acts we adopt [Poesio, 1995a], a new conversational event is recorded in the common ground each time a conversant utters a sound, no matter whether the sound corresponds to a phoneme, a word, or it's just noise. We call such utterances MICRO CONVERSATIONAL EVENTS (MCE). We use the binary predicate **utter** to characterize locutionary acts at all levels, including MCEs. Using the notation for representing events introduced in the previous sections, the update resulting from an utterance by speaker A of the word *engine* can be characterized as in (19).

(19)
$$\begin{array}{|l|}\hline \mu\text{-}ce_2 \\ \hline \ldots \\ \mu\text{-}ce_2 : \boxed{\textbf{utter}(A, \text{“}engine\text{”})} \\ \hline \end{array}$$

(19) is a 'radically underspecified'[27] characterization of the update to the common ground resulting from an utterance of the word *engine*. This information is available in the common ground before the listener has heard a complete sentence, in fact, even before the syntactic and semantic interpretation of the utterance fragment has been determined. The task of the listener is to complete this initial interpretation by inferring the initiator's intentions, i.e., how this MCE combines with other MCEs to generate one of the four classes of speech acts discussed above.

Part of the information to be inferred is the syntactic category and the meaning the speaker intended for the micro conversational event. Using the symbol **cat** to

[27] This term, derived from work on underspecification in phonology, has been proposed for underspecification in semantics by Pinkal, p.c.

indicate the function from micro-conversational events to their syntactic category, the symbol $\rightsquigarrow$ to denote the function from MCEs to their meaning, and the predicate $\mathbf{engine}_t$ to indicate the sense of *engine* most salient in the TRAINS conversations, the result of lexical access can be characterized as follows:

(20)
$$\begin{array}{|l|}\hline \mu\text{-}ce_2 \\ \hline \ldots \\ \mu\text{-}ce_2 : \boxed{\mathbf{utter}(A, \text{“}engine\text{”})} \\ \mu\text{-}ce_2 \rightsquigarrow \mathbf{engine}_t \\ \mathbf{cat}(\mu\text{-}ce_2) = \text{N} \\ \ldots \\ \hline \end{array}$$

Micro conversational events can also be recognized as being part of larger conversational events. For example, syntax will tell the listener that an utterance of the determiner *an* and an utterance of the noun *engine* immediately following the first may be the decomposition of a larger event of uttering an NP. Such utterance of an NP would dominate the utterances of the determiner and the noun, as in:

(21)
$$\begin{array}{|l|}\hline \mu\text{-}ce_1, \mu\text{-}ce_2, \mu\text{-}ce_3 \\ \hline \ldots \\ \mu\text{-}ce_1 : \boxed{\mathbf{utter}(a, \text{“}an\text{”})} \\ \mathbf{cat}(\mu\text{-}ce_1) = \text{DET} \\ X \rightsquigarrow \lambda P.\ \lambda Q.\ [x|];P(x);Q(x) \\ \mu\text{-}ce_2 : \boxed{\mathbf{utter}(a, \text{“}engine\text{”})} \\ \mathbf{cat}(\mu\text{-}ce_2) = \text{N} \\ X \rightsquigarrow \lambda x.\ [|\mathbf{engine}_t(x)] \\ \mu\text{-ce}_1 \prec \mu\text{-ce}_2 \\ \mathbf{cat}(\mu\text{-}ce_3) = \text{NP} \\ \mu\text{-}ce_3 : \boxed{\mathbf{utter}(a, \text{“}an\ engine\text{”})} \\ \mu\text{-ce}_2 \uparrow \mu\text{-ce}_3 \\ \mu\text{-ce}_1 \uparrow \mu\text{-ce}_3 \\ \ldots \\ \hline \end{array}$$

Other information that can be inferred from the locutionary acts together with the rest of the common ground is the referent of anaphoric expressions, and what illocutionary and perlocutionary acts the initiator intended. 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 locutionary acts in the sense of Goldman [1970]. We will discuss these processes, as well as the details of semantic interpretation, shortly. The discourse situation resulting from an utterance by A of *There is an engine at Avon* interpreted as an **inform** is shown in (22).

(22)
$$
\begin{array}{|l|}
\hline
\textit{la1 s s' ce1} \\
\hline
\textit{la1}: \ \mathbf{utter}(A, \textit{"there is an engine at Avon"}) \\
\textit{la1} \rightsquigarrow \begin{array}{|l|}
\hline
x\ w\ e \\
\hline
\mathbf{engine}(x) \\
\mathbf{Avon}(w) \\
e : \boxed{\mathbf{at}(x,w)} \\
\hline
\end{array} (s)\ (s') \\
\textit{ce1}: \ \mathbf{inform}(A,B, \begin{array}{|l|}
\hline
x\ w\ e \\
\hline
\mathbf{engine}(x) \\
\mathbf{Avon}(w) \\
e : \boxed{\mathbf{at}(x,w)} \\
\hline
\end{array} (s)\ (s')) \\
\mathbf{generate}(\textit{la1},\textit{ce1}) \\
\hline
\end{array}
$$

We should mention here that instead of three separate kinds of locutionary acts, as suggested by Austin, we have a single one, of uttering a sound; the 'phatic' and 'rhetic' effects of utterances are characterized as additional information about micro conversational events, rather than as separate actions. It is also worth emphasizing that the hierarchical organization of locutionary acts is isomorphic to the syntactic structure of utterances, and therefore the structure of locutionary acts can perform the same function played by the separate linguistic component of discourse structure postulated by Grosz and Sidner.

The fact that both locutionary acts and illocutionary acts are present in the common ground originates an interesting question: which acts augment the common ground with new discourse referents? I.e., is it the locutionary or the illocutionary acts that introduce new referents? This problem may not be apparent from (22) since **inform** is a special kind of speech act whose content is a proposition which can be equated with the meaning of the locutionary act, but we will see below cases of acts like **suggest** in which this is not the case.

This question is not answered yet, but there is reason to believe that the dynamics is actually associated with locutionary acts, and this is the solution we adopted. The first reason is that although an utterance can be thought of as a single act at the locutionary level, in general it generates more than one act at the illocutionary level, not all of which have the same 'content' argument; if indeed the dynamics were associated with illocutionary acts, we might expect to see multiple updates at each utterance, which doesn't seem to be the case. Secondly, new discourse referents can be added to the common ground by micro-conversational events, before the illocutionary act(s) can be recognized. Finally, core speech acts like **ynq** may update the common ground while asking to verify a fact (as in *A: is there an engine* *at Avon? B: yes, it is hooked to the boxcar*), but the update itself is not part of what is being queried (e.g., in the example just given the speaker is not asking whether there new discourse referent for *an engine* ought to be added to the common ground).

5.2 Micro Conversational Events and Underspecification

This view of interpretation as a process of 'filling in' gaps in the information initially available to the listener is usually characterized as involving UNDERSPECIFICATION, and has been championed, e.g., by [Schubert and Pelletier, 1982; Hobbs *et al.*, 1993; Alshawi, 1992; Reyle, 1993; Poesio, 1995b; van Deemter and Peters, 1996]. These theories of utterance interpretation all assume that (part of) the reason why people have no trouble in dealing with ambiguous expressions is because they do not generate all interpretations before filtering them, but start with a partially specified interpretation and then generate those few interpretations available in context.

The idea that underspecified representations are partial descriptions of the 'microstructure' of the discourse situation goes further than most current proposals, in that it provides a format in which *syntactic* as well as semantic underspecification can be expressed.[28] A second difference between the form of underspecification just proposed and the alternatives is that by making underspecified representations partial characterizations of the discourse situations we can express them within the logics already introduced for semantic processing (e.g., Muskens' logic TT_2^4), instead of introducing new logics with a special semantics as done, e.g., in [Alshawi and Crouch, 1992; Reyle, 1993; Poesio, 1995b]. The result is a much simpler theory of the interaction between semantics and pragmatics; furthermore, none of the semantic proposals currently available is very satisfactory [Poesio, 1995a].[29]

[28] The theory of micro conversational events is similar to Hobbs' 'flat representations' [Hobbs, 1985; Hobbs *et al.*, 1993] in that it does not presuppose that grammar can provide a sentential interpretation before disambiguation can begin; this is a crucial requirement when developing a theory of interpretation in spoken conversations, as complete sentences are very rare in this kind of data. But where Hobbs' underspecified interpretation combines information about what has been uttered and information about the content of these utterances, the distinction between described situation and discourse situation adopted in our theory allows us to keep these two forms of information distinct, thus avoiding the problems discussed above.

[29] Alternative formulations of 'radically underspecified' theories of interpretation have been developed by Muskens and Pinkal, who do not, however, interpret their underspecified representations as partial characterizations of the discourse situation, but as expressions of a different 'glue language'.

5.3 Interpretation

The theory of locutionary acts just sketched gives us the opportunity to characterize all interpretation processes as a form of defeasible reasoning over underspecified representations [Hobbs *et al.*, 1993; Alshawi, 1992]. As space prevents an extensive discussion of our work on interpretation, we will just give here a few illustrative examples.

We formulate our rules as default inference rules in Reiter's Default Logic [Reiter, 1980]; the disambiguated interpretations of an utterance are obtained by computing the extensions of the default theory $\langle$D,W$\rangle$, where D is the set of interpretation rules and W is the initial, underspecified interpretation. This computation of the extensions takes place incrementally, after every micro-update of the common ground, and may be followed by pruning of some of the hypotheses.[30]

The lexical rules specifying the syntactic category and meaning of lexical items can be specified by 'lexical defaults' as in (23). (We use capital letters to indicate unbound variables.) There is one such lexical default for each lexical rule of grammars such as Muskens' in 2.2. We assume that the lexicon is accessed immediately after a micro conversational update [Tanenhaus *et al.*, 1979], i.e., that the lexical category of a word-string and its meaning are obtained very quickly. Ambiguous word-strings, i.e., word-strings with multiple lexical entries, are associated with multiple default inference rules, all activated in parallel.[31]

(23)

$$\frac{X: \boxed{\mathbf{utter}(A, \text{“}\mathit{engine}\text{”})} : \begin{array}{l} X \rightsquigarrow \lambda x.\ [|\mathbf{engine}_t(x)] \wedge \\ \mathbf{cat}(X) = \mathrm{N} \end{array}}{\begin{array}{l} X \rightsquigarrow \lambda x.\ [|\mathbf{engine}_t(x)] \wedge \\ \mathbf{cat}(X) = \mathrm{N} \end{array}}\ \textbf{LEX-ENGINE}$$

$$\frac{X: \boxed{\mathbf{utter}(A, \text{“}\mathit{a}\text{”})} : \begin{array}{l} X \rightsquigarrow \lambda P.\ \lambda Q.\ [x|];P(x);Q(x) \quad \wedge \\ \mathbf{cat}(X) = \mathrm{DET} \end{array}}{\begin{array}{l} X \rightsquigarrow \lambda P.\ \lambda Q.\ [x|];P(x);Q(x) \quad \wedge \\ \mathbf{cat}(X) = \mathrm{DET} \end{array}}\ \textbf{LEX-A}$$

Syntactic interpretation can also be specified by means of default inference rules like **NP:Det+N** in Figure 1 that specifies how a determiner and a noun combine into a noun phrase. The following abbreviations are used in the figure:

- μ-ce$_1$:[$_{\mathrm{NP}}$:sem k] stands for $\mathbf{cat}(\mu\text{-}ce_1) = \mathrm{NP}$ and $\mu\text{-}ce_1 \rightsquigarrow k$

[30] For a more extensive discussion, see [Poesio, 1996].

[31] This view of the grammar as specifying the syntactic category and meaning of sub-sentential utterance events originated in Situation Semantics [Barwise and Perry, 1983; Evans, 1985] and is the basis for Cooper's Situation Theoretic Grammar [Cooper, 1992].

$$\frac{\begin{array}{l}\mu\text{-ce}_1\text{:[}_{\text{Det}}\text{ :sem }\alpha] \wedge \mu\text{-ce}_2\text{:[}_{\text{N}}\text{ :sem }\beta] \\ \wedge\ \mu\text{-ce}_1 \prec \mu\text{-ce}_2\end{array} \ :\ \mu\text{-ce}_3\text{:[}_{\text{NP}}\text{ :const }\{\mu\text{-ce}_1,\mu\text{-ce}_2\}\text{ :sem }\alpha(\beta)]}{\mu\text{-ce}_3\text{:[}_{\text{NP}}\text{ :const }\{\mu\text{-ce}_1,\mu\text{-ce}_2\}\text{ :sem }\alpha(\beta)]}\ \textbf{NP:Det+N}$$

Figure 1: A defeasible rule for parsing

- $\mu\text{-}ce_3$: $[_{\text{NP}}\ \mu\text{-}ce_1 : \alpha\ \mu\text{-}ce_2 : \beta]$ stands for: $\{\ \textbf{cat}(\mu\text{-}ce_3) = \text{NP},\ \mu\text{-ce}_2 \uparrow \mu\text{-ce}_3,\ \mu\text{-ce}_1 \uparrow \mu\text{-ce}_3\} \cup \{\ \mu\text{-}ce_1 : \alpha,\ \mu\text{-}ce_2 : \beta\}$

It should be easy to see how grammars such as the one presented in section 2.2 can be reformulated in terms of default inference rules along the lines of **NP:Det+N**.[32]

Referential expressions like *the engine* are another example of micro conversational events whose meaning is not specified by the grammar. The initial underspecified interpretation is exemplified by (24):

(24)

$$\begin{array}{|l|}\hline \mu\text{-}ce_1,\mu\text{-}ce_2,\mu\text{-}ce_3 \\ \hline \ldots \\ \mu\text{-}ce_1 :\ \boxed{\textbf{utter}(a,\text{“}the\text{”})} \\ \textbf{cat}(\mu\text{-}ce_1) = \text{DET} \\ \mu\text{-}ce_2 :\ \boxed{\textbf{utter}(a,\text{“}engine\text{”})} \\ \textbf{cat}(\mu\text{-}ce_2) = \text{N} \\ \textbf{cat}(\mu\text{-}ce_3) = \text{NP} \\ \mu\text{-ce}_2 \uparrow \mu\text{-ce}_3 \\ \mu\text{-ce}_1 \uparrow \mu\text{-ce}_3 \\ \ldots \\ \hline \end{array}$$

The task of reference resolution is to determine the meaning of such micro conversational events, i.e., to add to the common ground facts of the form $\mu\text{-}ce_3 \rightsquigarrow \lambda P.\ P(a)$, where a is an accessible antecedent in context and P(a) is a condition.[33]

[32] We want to emphasize that we are not suggesting that parsers should be implented as general-purpose defeasible reasoners. What we are suggesting is that the view just presented is an appropriate characterization of various types of disambiguation processes; more specialized reasoners may be involved in each particular process. Whenever the input is only partially grammatical, however, it is necessary to have a way to codify the interaction between a traditional parser and 'robust' parsing techniques, so that this can be made accessible to explicit repair; this can be done in terms of micro conversational events.

[33] See [Poesio, 1994; Poesio, 1993] for details about definite description interpretation. The current theory can also be used to reformulate work such as [Cohen, 1984; Heeman and Hirst, 1995] that depends on the notion of 'referring acts' introduced in [Searle, 1969]. What we are suggesting is that referring acts can be seen as particular cases of micro conversational events, instead of as instances of a special type of action.

5.4 Interpretation: Inferring Illocutionary Acts

A critical component of the interpretation process is the attempt to determine what was actually done by the initiator in performing the locutionary acts, i.e., to recognize the illocutionary acts he/she performed. Again, context plays a crucial role in this process. As described by Austin [1962] and others, the same sentence can be uttered in different contexts to perform radically different actions. For instance, in the example developed above, the conversational event *ce1*, from (12), 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.

The current point in the current conversational thread, and hypotheses about the initiator'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 (14), it is probable[34] that the **suggest** 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]. Assuming the **suggest** interpretation of the locutionary act *la1* of uttering the sound *there is an engine at Avon*, the information acquired via the speech act recognition process can be represented as in (25).

(25)

$$\boxed{\begin{array}{l} x \ldots pl\ la1\ sug1 \\ \hline \ldots \\ la1:\ \mathbf{utter}(A, \textit{"there is an engine at Avon"}) \\ sug1:\ \mathbf{suggest}(A,B, \boxed{\mathbf{use}(\{A,B\},x,pl)}\ ) \\ \mathbf{generate}(la1,sug1) \end{array}}$$

This means that the actions performed in the utterance of *la1* is a suggestion that the agents use *x* in their ongoing domain plan, *pl*.[35] 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 *la1*. For this example, these might include constraints such that *x* is the engine of a planned move-boxcar action with destination *Bath*. These

[34] Depending also on other aspects of the mental state and preceding discourse.

[35] The representation of plans in TRAINS-93 is discussed in [Traum *et al.*, to appear 1996].

further IMPLICATURES will also be added into the DRS along with the suggestion.

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. This assumption is an idealization, however. As shown, e.g., in [Clark and Wilkes-Gibbs, 1986; Clark and Schaefer, 1989], utterances by one conversant must be recognized and acknowledged by the other before becoming part of the common ground. This collaborative process of adding to the common ground is called GROUNDING. Grounding 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 (9), B could have acknowledged by uttering something like *okay* or *right*—or tacit —the second utterance in (9) 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:[36]

(26)
```
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
```

[36] This fragment is in dialog d91-6.1 in [Gross *et al.*, 1993].

```
7.2    : at Avon
7.3    : right
8.1  S: right
```

6.1 A Computational Model of Grounding

Clark and Schaefer presented an off-line model of the grounding process in [Clark and Schaefer, 1989]; a computational account was presented in [Traum, 1994] and implemented within the TRAINS-93 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 performing core speech acts —such as **suggest**, **inform**, etc.— he/she is also performing one or more grounding acts. Grounding acts include: **init**, which opens a new DU; **continue**, which adds more material to an already open DU; **ack**, which makes the contents of the DU enter the common ground; as well as others, devoted to repairs. Presented material that could be acknowledged together (e.g., with a single *okay*) is grouped together into a DISCOURSE UNIT (DU). 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[37] of accessible DUs is maintained in the model, as context for recognizing and deciding to perform grounding acts. Table 2 summarizes the grounding acts, while table 3 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 initiator of that DU, and R for the responder (the other agent).

6.2 Discourse Units as Conversational Threads

The theory of discourse situations we have been developing up to now already includes two of the features needed by the theory of the grounding process just discussed, namely, that the common ground includes information about the occurrence of different kinds of conversational events, and that conversational events are organized into threads. We are simply going to assume that grounding acts are another kind of conversational event whose occurrence is recorded in the common ground, and that discourse units are threads of grounding acts. We also discussed the idea

[37] 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.

Label	Description
initiate	Begin new DU, content separate from previous uncompleted DUs
continue	Continue previous material by the same speaker
acknowledge	Demonstrate or claim understanding of previous material by other conversant
repair	Correct (potential) misunderstanding of DU content
Request Repair	Signal lack of understanding
Request Ack	Signal for other to acknowledge
cancel	Stop work on DU, leaving it ungrounded and ungroundable

Table 2: Grounding Acts

that listeners exploit their knowledge of 'discourse scripts'—conversational threads that follow a certain routine—to predict what's coming next; our DU State Transition Diagram provides a way of encoding such information about the next moves in the case of the scripts having to do with grounding.

Finally, we assume that the **generate** relation 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 (9), B's utterance can be seen as generating both an (implicit) acknowledgment (**ack**) of the DU that includes A's utterance, as well as initiating (**init**) a new DU for the new content that it contains in its own right. This interpretation of the utterance is captured by assuming that it results in adding to the Root DRS the conditions shown in (27). 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 conversational threads.[38] The resulting state of the DUs after the utterance is described by conditions like that used to describe the intention *i1* in section 3.

The predicates **state1**(*du2*) and **stateF**(*du1*) capture the current state of each discourse unit. Their implications about the mental state of the participating agents are 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 responder and obligations of the two agents.

[38] Although DUs are seen as another instance of the same form of organization which results in 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.

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	
cancel^R			1	1		D	

*repair request is ignored

Table 3: DU State Transition Diagram

(27)

…du1 …ce1 ce2 ack1 du2 init2

…
ack1 : **ack**(*B*,*du1*)
init2 : **init**(*B*,*du2*)

generate(*ce1*,*ack1*)
s-du1-ce2 : **stateF**(*du1*)
generate(*ce2*,*init2*)
s-du2-ce2 : **state1**(*du2*)

Notice that some utterances, (e.g., *okay* following an utterance by another speaker) generate grounding acts (in this case, an **ack**), but do not introduce new discourse referents or generate core speech acts. The model of discourse we are discussing thus provides an interpretation for utterances that perform operations on the common ground that are relevant for reference purposes, although do not directly introduce new referential material.

6.3 Grounded and Ungrounded Aspects of the Discourse Situation

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 conversants mutually believe (or, perhaps, what one participant believes 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. Consider the following utterance in a conversation between **A** and **B**:

(28) A: There is an engine at Avon.

(28) could be followed by any of the following continuations, as well as numerous other possibilities:

(29) a. A: Let's pick it up.
b. B: Let's pick it up.
c. B: Uh huh.
d. B: There's a what?
e. A: I mean a boxcar.
f. A: Did you hear me?

The first example shows that the object mentioned in (28) (the engine at Avon) must be accessible to interpret the pronoun *it* in (29a). The second example shows that the other conversant can also make use of the mentioned entity without first grounding it. Thus, if common ground is a prerequisite for felicitous definite reference,[39] the entity must be part of the common ground. On the other hand, the third example shows a potential response to (28) which would have the effect of grounding the content of that utterance. More compellingly, a response such as (29d) would show that the utterance (in particular the mentioned object) is *not* understood, and thus not part of the common ground. (29e) is a repair by A, showing that even if B might have assumed commonality, such assumption would have been incorrect, because A mis-spoke. Finally, (29f) shows that speakers do not always assume that their

[39] 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.

utterances have been heard and understood without feedback. Since all of these sorts of examples occur frequently in task-oriented conversations, 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.

What is needed is a more general notion of 'discourse model,' which may include both grounded and ungrounded information; grounding acts can then be seen as operations that either add to the ungrounded part of the discourse model, or move material from the ungrounded part to the grounded part.

More precisely, we suggest that the common ground consists of two parts, 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 shared by both conversants. In addition, we have an UN-GROUNDED ROOT DRS (UR-DRS); this is an extension of the grounded root DRS which, in addition, includes all of the ungrounded DUs. Each DU thread within the UR-DRS represents an agent's view of what should become part of the common ground: e.g., when A initiates a DU with content ϕ about described situation *s*, the DU will include the state **intend**(*A*,**MB**($\{A,B\}$,s : ϕ)). Items introduced as part of a DU 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 way, 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].

The grounding process is formalized as follows. 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 DRS GR-DRS;$\langle$U$'$,C$'\rangle$, where U$'$and C$'$represent the ungrounded referents and conditions; as discussed above, this is equivalent to a pair $\langle \mathrm{U} \cup \mathrm{U}', \mathrm{C} \cup \mathrm{C}' \rangle$. Grounding acts are operations on the content of these pairs. The result of an update due to a conversational event generating an **init**, **continue**, or **repair** grounding act is always an ungrounded DRS. **acknowledgments** are moves of part of the current discourse model from an ungrounded to a grounded state. The result of acknowledging discourse unit du_1 consisting of conversational events $ce_i \ldots ce_j$ is to map the GR-DRS K into the new DRS K;[$ce_i,\ldots,ce_j | ce_i : \varphi_i \ldots ce_j : \varphi_j$]. Thus, after processing the acknowledgment performed in *ce2*, the GR-DRS will look something like (30), while the UR-DRS will also include the information shown in (31), as well[40]

[40] These figures show only the aspects of the discourse situation that we have discussed previously in the paper. There are, of course, many additional facts that will be part of these DRSs as a result of processing even this short bit of dialogue. For one thing, we have omitted all of the micro conversational events.

(30)

$$\begin{array}{|l|}\hline
s \ldots s' i1 \;\; pl \ldots la1 \;\; ce1 \; ct1 \; du1 \ldots \;\; ack1 \\ \hline
la1 : \;\; \mathbf{utter}(A, \text{“\textit{there is an engine at Avon}”}) \\
la1 \rightsquigarrow \begin{array}{|l|}\hline x \; w \; e \\ \hline \mathbf{engine}(x) \\ \mathbf{Avon}(w) \\ e : \boxed{\mathbf{at}(x,w)} \\ \hline \end{array} (s)\,(s') \\
ce1 : \;\; \mathbf{inform}(A,B, \begin{array}{|l|}\hline x \; w \; e \\ \hline \mathbf{engine}(x) \\ \mathbf{Avon}(w) \\ e : \boxed{\mathbf{at}(x,w)} \\ \hline \end{array} (s)\,(s')) \\
\mathbf{generate}(la1, ce1) \\
ce1 \uparrow ct1 \\
\mathbf{generate}(la1, ack1) \\
ack1 : \boxed{\mathbf{ack}(B, du1)} \\ \hline
\end{array}$$

(31)

$$\begin{array}{|l|}\hline
s \ldots du2 \; init2 \; la2 \; s' \; s'' \\ \hline
la2 : \;\; \mathbf{utter}(B, \text{“\textit{it is hooked to the boxcar}”}) \\
la2 \rightsquigarrow \begin{array}{|l|}\hline y \; u \; e' \\ \hline \mathbf{boxcar}(y) \\ u \; \mathbf{is} \; x \\ e' : \boxed{\mathbf{hooked\text{-}to}(y,u)} \\ \hline \end{array} (s')\,(s'') \\
init2 : \boxed{\mathbf{init}(B, du2)} \\ \hline
\end{array}$$

7 Conclusions and Open Issues

We have presented a theory of the information about the discourse situation shared by the participants in a conversation that has two main characteristics. First of all, we showed that a number of facts about spoken dialogues—discourse segmentation effects, the fact that speakers need not utter complete sentences, and that they are typically involved in more than one task at once—can be explained starting with only a few, generally accepted assumptions: that all utterances are actions (including ‘micro’ utterances); that utterances generate various sorts of speech acts; and that speech acts, being events, are structured in the same way other events are (i.e., they are organized into larger events whose internal structure is mutually known). Secondly, we showed that a theory of this kind can be formalized by fairly simple extensions of the technical tools developed by formal semanticists, and therefore, the results in one area can be used in the other. This formalization also raises some intriguing issues which we aren't completely resolved, but couldn't even be observed before attempts such as ours—namely, what is the precise impact of anaphoric accessibility of dominance and satisfaction-precedes, and whether discourse update is the result of locutionary or illocutionary acts.

There are of course plenty of open issues; we will mention three. We have only briefly mentioned the problem of the semantics of questions (more precisely, the problem of which kind of objects occur as third argument of a **ynq** speech act) and we haven't mentioned at all the similar problem of the semantics of imperative sentences such as *send the engine to Avon*, which are particularly common in the TRAINS conversations. We think this is perhaps the major formal challenge for a theory of the type we are developing.

A second issue is how to formalize repair processes [Schegloff *et al.*, 1977; Levelt, 1983]. Whereas all speech acts that we have discussed augment the common ground, repairs seem to crucially involve a 'revision' step—i.e., some information seems to disappear. Thus, an account of repair processing seems to involve 'downdate' operations. Such operations are not unconceivable within the formal account we are presented, but the empirical facts are not yet completely clear.

Finally, there is the whole area of the management of the knowledge base encoding information about the discourse situation, rather than the information itself. One issue, for example, is the question of which information about the discourse situation is actually retained, and for how long. (It is well-known from the literature that micro conversational events are stored in short-term memory and disappear pretty quickly—this explains, for example, the short duration of priming effects.) The topic of resource bounds in dialogue is discussed in [Walker, 1993].

Acknowledgements

We would like to thank the members of the TRAINS project at the university of Rochester, particularly James Allen, George Ferguson, Peter Heeman, Chung-Hee Hwang, and Len Schubert, for collaboration on the TRAINS system which inspired the present work and for lots of feedback. Massimo Poesio would like to thank Robin Cooper, Richard Crouch, Jonathan Ginzburg, and Reinhard Muskens for many useful discussions and comments. We would also like to thank the anonymous reviewers for helpful suggestions on this presentation.

References

[Airenti *et al.*, 1993] G. Airenti, B. G. Bara, and M. Colombetti. Conversation and behavior games in the pragmatics of dialogue. *Cognitive Science*, 17:197–256, 1993.

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

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

[Allen, 1983] J. F. Allen. Recognizing intentions from natural language utterances. In M. Brady and R. Berwick, editors, *Computational Models of Discourse*, pages 107–166. MIT Press, Cambridge, MA, 1983.

[Alshawi and Crouch, 1992] H. Alshawi and R. Crouch. Monotonic semantic interpretation. In *Proc. 30th. ACL*, pages 32–39, University of Delaware, 1992.

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

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

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

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

[Boër and Lycan, 1980] S. E. Boër and W. G. Lycan. A performadox in truth-conditional semantics. *Linguistics and Philosophy*, 4:71–100, 1980.

[Bunt, 1995] H. C. Bunt. Dynamic interpretation and dialogue theory. In M. M. Taylor, F. Néel, and D. G. Bouwhuis, editors, *The Structure of Multimodal Dialogue 2*. John Benjamins, Amsterdam, 1995.

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

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

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

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

[Clark, 1996] H. H. Clark. *Using Language*. Cambridge University Press, Cambridge, 1996.

[Cohen and Levesque, 1990] P. R. Cohen and H. J. Levesque. Persistence, intention and commitment. In P.R. Cohen, J. Morgan, and M. Pollack, editors, *Intentions in Communication*, chapter 12. Morgan Kaufmann, 1990.

[Cohen and Perrault, 1979] P. R. Cohen and C. R. Perrault. Elements of a plan based theory of speech acts. *Cognitive Science*, 3(3):177–212, 1979.

[Cohen *et al.*, 1990] P. R. Cohen, J. Morgan, and M. Pollack, editors. *Intentions in Communication*. MIT Press, Cambridge, MA, 1990.

[Cohen, 1984] P. R. Cohen. The pragmatics of referring and the modality of communication. *Computational Linguistics*, 10(2):97–146, April-June 1984.

[Cooper, 1992] R. Cooper. Three lectures on Situation Theoretic Grammar. Research paper, University of Edinburgh, Centre for Cognitive Science, 1992.

[Davidson, 1967] Donald Davidson. The logical form of action sentences. In N. Rescher, editor, *The Logic of Decision and Action*, pages 81–95. University of Pittsburgh Press, Pittsburgh, 1967.

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

[Evans, 1985] D. Evans. *Situations and Speech Acts: toward a formal semantics of discourse*. Garland, New York, 1985.

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

[Ginzburg, 1995a] J. Ginzburg. Resolving questions, i. *Linguistics and Philosophy*, 18(5):567–609, 1995.

[Ginzburg, 1995b] J. Ginzburg. Resolving questions, ii. *Linguistics and Philosophy*, 18(6):567–609, 1995.

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

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

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

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

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

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

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

[Heim, 1982] I. Heim. *The Semantics of Definite and Indefinite Noun Phrases*. PhD thesis, University of Massachusetts at Amherst, 1982.

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

[Hobbs *et al.*, 1993] J. R. Hobbs, M. Stickel, P. Martin, and D. Edwards. Interpretation as abduction. *Artificial Intelligence Journal*, 63:69–142, 1993.

[Hobbs, 1985] J. Hobbs. Ontological promiscuity. In *Proceedings ACL-85*, pages 61–69, 1985.

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

[Hwang, 1992] C. H. Hwang. *A Logical Approach to Narrative Understanding*. PhD thesis, University of Alberta, Department of Computing Science, Edmonton, Alberta, CA, 1992.

[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, 1991] H. Kamp and U. Reyle. A calculus for first order Discourse Representation Structures. Sonderforschungsbereichs 340 Bericht 14, University of Stuttgart, Stuttgart, 1991.

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

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

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

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

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

[Kautz, 1987] H. A. Kautz. *A Formal Theory of Plan Recognition*. PhD thesis, Department of Computer Science, University of Rochester, Rochester, NY, 1987. Also available as TR 215, Department of Computer Science, University of Rochester.

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

[Kowtko *et al.*, 1992] J. C. Kowtko, S. D. Isard, and G. M. Doherty. Conversational games within dialogue. Research Paper HCRC/RP-31, Human Communication Research Centre, June 1992.

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

[Levelt, 1983] W. J. M. Levelt. Monitoring and self-repair in speech. *Cognition*, 14:41–104, 1983.

[Levinson, 1983] S. Levinson. *Pragmatics*. Cambridge University Press, 1983.

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

[Milward, 1995] D. Milward. Dynamics and situations. In *Proceedings of the Tenth Amsterdam Colloquium*, 1995.

[Montague, 1973] R. Montague. The proper treatment of quantification in english. In K.J.J. et al. Hintikka, editor, *Approaches to Natural Language*, pages 221–242. D. Reidel, Dordrecht, 1973.

[Moore and Paris, 1993] J. D. Moore and C. L. Paris. Planning text for advisory dialogues: Capturing intentional and rhetorical information. *Computational Linguistics*, 19(4):651–694, December 1993.

[Muskens, 1989] R. Muskens. *Meaning and Partiality*. PhD thesis, University of Amsterdam, 1989.

[Muskens, 1994] R. Muskens. A compositional discourse representation theory. In P. Dekker and M. Stokhof, editors, *Proceedings of the 9th Amsterdam Colloquium*, pages 467–486, 1994.

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

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

[Orestrom, 1983] B. 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*, chapter 9, pages 161–185. The MIT Press, Cambridge, MA, 1990.

[Poesio, 1991a] M. Poesio. Expectation-based recognition of discourse segmentation. In *Proc. AAAI Fall Symposium on Discourse Structure*, Asilomar, CA, November 1991.

[Poesio, 1991b] M. Poesio. Relational semantics and scope ambiguity. In J. Barwise, J. M. Gawron, G. Plotkin, and S. Tutiya, editors, *Situation Semantics and its Applications, vol.2*, chapter 20, pages 469–497. CSLI, Stanford, CA, 1991.

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

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

[Poesio, 1996] M. Poesio. Underspecification in a theory of utterance processing. In preparation, 1996.

[Power, 1979] R. J. D. Power. The organization of purposeful dialogue. *Linguistics*, 17:107–152, 1979.

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

[Reiter, 1980] R. Reiter. A logic for default reasoning. *Artificial Intelligence*, 13(1–2):81–132, April 1980.

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

[Roberts, 1989] C. Roberts. Modal subordination and pronominal anaphora in discourse. *Linguistics and Philosophy*, 12:683–721, 1989.

[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.

[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.

[Schank and Abelson, 1977] R. C. Schank and R. Abelson. *Scripts, Plans, Goals and Understanding*. Lawrence Erlbaum, 1977.

[Schegloff and Sacks, 1973] E. A. Schegloff and H. Sacks. Opening up closings. *Semiotica*, 7:289–327, 1973.

[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.

[Schubert and Pelletier, 1982] L. K. Schubert and F. J. Pelletier. From English to Logic: Context-free computation of 'conventional' logical translations. *American Journal of Computational Linguistics*, 10:165–176, 1982.

[Searle, 1969] J. R. Searle. *Speech Acts*. Cambridge University Press, New York, 1969.

[Stalnaker, 1979] R. Stalnaker. Assertion. In P. Cole, editor, *Syntax and Semantics*, volume 9, pages 315–332. Academic Press, 1979.

[Tanenhaus *et al.*, 1979] M. K. Tanenhaus, L. A. Stowe, and G. Carlson. Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning and Verbal Behavior*, 18(4):427–440, August 1979.

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

[Traum and Heeman, 1996] D. R. Traum and P. Heeman. Utterance units and grounding in spoken dialogue. to be presented at *4th International Conference on Spoken Language Processing (ICSLP-96)*, October, 1996.

[Traum and Hinkelman, 1992] D. R. Traum and E. A. Hinkelman. Conversation acts in task-oriented spoken dialogue. *Computational Intelligence*, 8(3), 1992. 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] D. R. Traum. *A Computational Theory of Grounding in natural language conversation*. PhD thesis, University of Rochester, Department of Computer Science, Rochester, NY, July 1994.

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

[van Deemter and Peters, 1996] K. van Deemter and S. Peters, editors. *Semantic Ambiguity and Underspecification*. CSLI Publications, Stanford, 1996.

[Walker, 1993] M. A. Walker. *Informational Redundancy and Resource Bounds in Dialogue*. PhD thesis, University of Pennsylvania, 1993.

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

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

A Syntax and Semantics of the Representation Language

Muskens' TT_2^4 logic is a partial typed logic based on a generalization of Montague's system of types. Although Montague's logic is 'intensional'—i.e., it includes expressions of type $\langle s,\alpha \rangle$ that denote functions from ⟨world,time⟩ pairs (also called 'indices') into objects in the domain of type α—there is no basic type *s* in the logic, i.e., there are no expressions denoting world- or situation-like objects. The set of types in TT_2^4 includes such a type; the objects of that type are called SITUATIONS. The set of types of TT_2^4 is the minimal set such that

1. *e*, *t*, and *s* are types, and
2. if α and β are types, then so is $\langle \alpha,\beta \rangle$

For each type an infinite set of variables and constants of that type is assumed. The terms of the language of TT_2^4 are defined in the standard way, as follows:

1. Every constant or variable of any type is a term of that type;
2. If φ and ψ are terms of type *t* (FORMULAE), then $(\neg\varphi)$ and $(\varphi \wedge \psi)$ are formulae;
3. If φ is a formula and *x* is a variable of any type, then $(\forall x\ \varphi)$ is a formula;
4. If A is a term of type $\langle \alpha,\beta \rangle$ and B is a term of type α, then (A B) is a term of type β;
5. If A is a term of type β and *x* is a variable of type α, then $\lambda\, x.$ A is a term of type $\langle \alpha,\beta \rangle$;

6. If A and B are terms of type α, then (A = B) is a formula.[41]

We will also make use in what follows of the defined formula $\mathbf{E}(x,s)$, interpreted as 'x exists in situation s' ([Muskens, 1989], page 71).

A model for the language is defined as a pair $\langle$F,I$\rangle$, where F is a 'TT^4_2 frame' (a set of sets D_α providing interpretations for objects of type α) and I is an interpretation function assigning interpretations in D_α to objects of type α. The expressions just listed have the semantics one would expect; we won't discuss it for brevity.

Two properties of Muskens' system are of interest here. First of all, the language can be used to make properties and predicates explicitly dependent on their 'index of evaluation'—a situation—by assigning them the appropriate type. For example, the natural language verb *run* can be interpreted as a term of type $\langle e,\langle s,t\rangle\rangle$. We can thus explicitly specify in our language what is the case at each situation. Secondly, situations are organized by an inclusion relation $\subseteq$, such that $s \subseteq s'$ iff the domain of s is a subset of the domain of s', and anything which is definitely true or definitely false at s preserves its truth value at s'. We use the inclusion relation to model 'information growth'.

We 'embed' in TT^4_2 some aspects of the system of [Muskens, 1994] by extending the former as follows. (The reader should compare the following quick description with the description of [Muskens, 1994] in section 2.2.) First of all, we define new primitive types π_e and π_s of 'pidgeon holes' (discourse referents) 'containing' objects of e and s, respectively, and we consequently redefine the set of types as the minimal set such that:

1. e, t, π_e, π_s and s are types, and
2. if α and β are types, then so is $\langle\alpha,\beta\rangle$

We use the same notion of frame and model that is used in TT^4_2, except that we require a frame to include nonempty sets D_{π_e} and D_{π_s} used to interpret constants and variables of type π_e and π_s, respectively.

We assume all of the term definitions in TT^4_2, and in addition we allow for infinite constants and variables of types π_e and π_s. We add non-logical constants $\mathbf{V}_e$ and $\mathbf{V}_s$ denoting total functions from 'e-type' and 's-type' pidgeon holes and situations to D_e and D_s, respectively, such that $\mathbf{V}_y(u,s)$ (where y is either e or s) is the object in D_y stored in pidgeon hole u at situation s. These functions play the

[41] The '4' superscript in the name of the logic indicates that TT^4_2 is a four valued logic, whose values are **T**, **F**, **N** (neither) and **B** (both). Negation and conjunction are not sufficient to express all functions over four truth values, hence the language of TT^4_2 also includes the symbols # (denoting **B**) and $\star$ (denoting **N**). We have omitted them from the description as we don't use them in the paper.

same role of Muskens' $\mathbf{V}(u,s)$ function in [Muskens, 1994]. We do not require that the value of $\mathbf{V}_y(u,s)$ be an object in s. (For simplicity, we will omit indices on both pidgeon-holes and **V** functions below except where confusion might arise.) We also define an 'update' relation $s[u_1, \ldots, u_n]s'$; however, we take it to be a relation between situations, and we allow for pidgeon holes of both types π_e and π_s.

- $s[u_1, \ldots, u_n]s'$ is short for $s \subseteq s' \wedge \mathrm{E}(\mathbf{V}_y(u_1,s'))(s') \wedge \ldots \mathrm{E}(\mathbf{V}(u_n,s'))(s') \wedge \forall v u_1 \neq v \wedge \ldots \wedge u_n \neq v \rightarrow \mathbf{V}_y(v,s) = \mathbf{V}_y(v,s')$, where $\mathbf{V}_y$ is the appropriate function given the type of v. Note that the values have to be defined in s'.
- $s[]s'$ is short for $\forall v_y\ \mathbf{V}_y(v_y,s) = \mathbf{V}_y(v_y,s)$.

We assume all axioms of TT_2^4 defined in [Muskens, 1989]. Of the axioms that specify the behavior of discourse referents in [Muskens, 1994], we maintain two: AX1 and AX3. Our version of AX1 is modified so as to take into account the possibility of having pidgeon holes that hold situations.

AX1-MOD $\forall s\, \forall v\, \forall x_y\ \ \exists s'\ \ s[v]s' \wedge \mathbf{V}_y(v)(s') = x_y$

AX3 $u_y \neq u'_y$ for each two different discourse referents of type y, where $y = e$ or s

We do not require two situations s and s' to be identical if for any discourse referent v, $\mathbf{V}_y(v)(s) = \mathbf{V}_y(v)(s')$.

We can now redefine the DRS constructs that we had introduced in Section 2.2 as follows:

$\mathbf{R}\{\tau_1,\ldots,\tau_n\}$	is short for	$\lambda s.\ \mathbf{R}(\tau_1)\ldots(\tau_n)(s)$
τ_1 **is** τ_2		$\lambda s.\ (\tau_1) = (\tau_2)$
not K		$\lambda s.\ \neg\exists s'\ \ K(s)(s')$
K **or** K'		$\lambda s.\ \ \exists s'\ \ K(s)(s') \vee K'(s)(s')$
$K \Rightarrow K'$		$\lambda s.\ \ \forall s'\ \ K(s)(s') \rightarrow \exists s'' K'(s')(s'')$
$[u_1, \ldots, u_n \mid \varphi_1, \ldots, \varphi_m]$		$\lambda s.\ \ \lambda s'.\ \ s[u_1, \ldots, u_n]s' \wedge$ $\varphi_1(s'), \ldots, \varphi_m(s')$
$K\,;K'$		$\lambda s.\ \ \lambda s'.\ \exists s''\ \ K(s)(s'') \wedge K'(s'')(s')$
$s : \varphi$		$\lambda s'.\ (\alpha\ s)$ (where α is of type $\langle s,t\rangle$ and s is of type s)

We have added to Muskens' set of conditions a new condition $s : \varphi$, stating that s is of type φ, in analogy to what done in DRT and Situation Theory.

We can show that the following lemmas from [Muskens, 1994] still hold:

Merging Lemma : If $v_1, \ldots, v_p$ do not occur in $\varphi_1, \ldots, \varphi_m$,
$[u_1, \ldots, u_n \mid \varphi_1, \ldots, \varphi_m]; [v_1, \ldots, v_p \mid \psi_1, \ldots, \psi_q] = [u_1, \ldots, u_n, v_1, \ldots, v_p \mid \varphi_1, \ldots, \varphi_m, \psi_1, \ldots, \psi_q]$

Unselective Binding Lemma : Let $u_1, \ldots, u_n$ be constants of type π_e or type π_s, let $x_1 \ldots x_n$ be distinct variables of type *e* or *s*, let φ be a formula that does not contain s' and write $[\mathbf{V}(u_1,s') / x_1, \ldots, \mathbf{V}(u_n,s') / x_n]\varphi$ for the formula obtained by simultaneous substitution of $\mathbf{V}(u_1,s')$ for $x_1, \ldots, \mathbf{V}(u_n,s')$ for x_n in φ. Then the following hold:

1. $\forall s\ \ ((\exists\, s'\ \ s[u_1, \ldots, u_n]s' \wedge [\mathbf{V}(u_1,s') / x_1, \ldots, \mathbf{V}(u_n,s') / x_n]\varphi)$
   $\leftrightarrow (\exists\, x_1\ \ \ldots \exists\, x_n\ \varphi))$
2. $\forall s\ \ ((\forall s'\ \ s[u_1, \ldots, u_n]s' \rightarrow [\mathbf{V}(u_1,s') / x_1, \ldots, \mathbf{V}(u_n,s') / x_n]\varphi)$
   $\leftrightarrow (\forall x_1\ \ \ldots \forall x_n\ \varphi))$

The Unselective Binding lemma gives us assurance that we get the right truth conditions. The Merging Lemma allows us to build our interpretation of the discourse situation incrementally, by merging together the interpretations of single utterances. Thus, (13) is equivalent to (12).