A Multi-Purpose Model of Conversational Context

Massimo Poesio
University of Edinburgh
Centre for Cognitive Science
2 Buccleuch Place
Edinburgh EH8 9LW, Scotland, UK
poesio@cogsci.ed.ac.uk

David Traum
TECFA, FPSE
Universite de Geneve
9, Route De Drize, BAT D
CH-1227 Carouge, Switzerland
traum@divsun.unige.ch

Abstract

Natural language dialogue systems require contextual information for a variety of processing functions, including reference resolution, speech act recognition, and dialogue management. While much has been written about individual contextual problems, many of the proposed representations are mutually incompatible, unusable by the agents involved in a conversation, or both.

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

1 Motivations

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

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

The TRAINS prototype must perform at least three kinds of linguistic activities that depend on a context: reference resolution, conversation action recognition and dialogue management. The models of context for reference resolution are typically developed with the aim of making available the antecedents of various forms of anaphoric reference [Grosz, 1977; Webber, 1979; Grosz and Sidner, 1986; Kamp and Reyle, 1993], whereas the emphasis in the models of context proposed for action recognition and dialogue management has been on representing the occurrence of speech acts and information about the beliefs, intentions, and obligations of the participating agents [Cohen and Levesque, 1990; Perrault, 1990; Traum and Allen, 1994]. This difference in emphasis has resulted in conceptual differences between the models of context proposed within the two traditions.

For example, ever since Grosz's dissertation [Grosz, 1977] it has been known that subordination relations between discourse goals affect the structure of global focus, hence the 'pragmatic' accessibility of referents. A model of context that contains information about the intentions of the conversational participants is therefore advantageous for reference resolution purposes. But the models developed for intention recognition and dialogue control purposes do not provide enough information about which referents are available at each point in the conversation. On the other hand, the discourse models proposed for reference resolution purposes typically 'abstract away' from aspects of the input utterance that are not relevant for reference resolution: thus utterances of acknowledgments, such as *okay*, and other discourse particles are typically ignored. But these aspects of the utterance are crucial for speech act recognition and for tracking the state of the conversation for dialogue management purposes: thus if the model of context used for reference resolution does not provide this information, the dialogue manager either has to do without it, or it has to merge the information about context from the reference resolution module with the information which has been discarded.[1] All modules that deal with context could therefore gain from a unified discourse model.

Currently, there are very detailed proposals concerning context for reference resolution (e.g., [Kamp and Reyle, 1993]) and about the effect of speech acts on the mental state of agents, for intention recognition purposes (e.g., [Cohen and Levesque, 1990]) but only programmatic proposals for a unified model of context, too underspecified for use in an implemented system (e.g., [Grosz and Sidner, 1986]).

We outline a model of context that tries to reconcile in a general and systematic fashion the differences between the discourse models used for reference resolution, conversation act recognition, and dialogue management in a system dealing with conversations. This new theory is meant to be one that an agent can use as an internal on-line representation of context while engaging in conversation. This work is based on merging some of the ideas from two recent works within these traditions, namely [Poesio, 1994] and [Traum, 1994]. As the ideas from both of these models of context were implemented as modules within the same conversation system (using conversion languages for communication), we have reasons for optimism, both for the soundness of the contextual model, and for the utility of the result. In the rest of this paper, we gradually develop this model, starting from traditional basics of the reference resolution tradition and progressively adding in other features of context from other traditions.

2 A Minimal Representation of Context: Discourse Representation Theory

A discourse model for reference resolution purposes must minimally include a list of accessible DISCOURSE REFERENTS in addition to some sort of representation of the propositional content of utterances [Karttunen, 1976; Webber, 1979]. Such discourse models must also include a specification of how a new utterance 'updates' the previous discourse model, i.e., how it produces a new discourse model that include new discourse referents and/or propositional information. Discourse Representation Theory (DRT) [Kamp, 1981; Kamp and Reyle, 1993] and related 'dynamic' theories such as Dynamic Predicate Logic [Groenendijk and Stokhof, 1991] provide a formal way of capturing the information about discourse referents and their properties, and include a detailed update algorithm.

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

(1) Engine E2 is at Avon.

(2)

x w
engine E2(x) **Avon**(w) **at**(x,w)

A DRS is best thought of as a database including discourse referents in addition to facts, and that can include additional, subordinate databases. At the same time, DRSs can be thought of as logical expressions: (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.

We use the term ROOT DRS to indicate a DRS that represents the whole common ground, i.e., what is usually called the 'model of discourse'. The root DRS is updated after every sentence by a DRT CONSTRUCTION ALGORITHM whose first step is to add to the previous discourse model an 'uninterpreted condition' that provides information about the syntactic structure of the sentence and the interpretation of its lexical items, followed by the application of rewrite rules called CONSTRUCTION RULES which may add new conditions or new referents, or 'flesh out' the interpretation of the uninterpreted condition. Updating the model of discourse in (2) with the content of sentence *The boxcar is also at Avon and is hooked to it* is the 'extended' DRS in (4). Note that two new discourse referents have been introduced, y and u, for the definite *the boxcar* and the pronoun *it* respectively. The 'box' obtained by the construction algorithm also contains the discourse referents introduced by the first sentence, which are thus 'accessible' for reference purposes.[2]

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

(4)

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

[1]This approach was adopted in the TRAINS-93 demo system.

[2]Space prevents a detailed discussion of the algorithm. See [Kamp and Reyle, 1993].

Many recent theories of pronoun interpretation, definite description interpretation, and VP ellipsis resolution have found it convenient to adopt a formal discourse model such as DRT, because of its explicitness. The basic version of DRT, however, concentrates on representing semantic information, and abstracts away from all information of a pragmatic nature, including information that is needed for reference resolution purposes such as the fact that a certain speech act occurred at time t, or that agent a has certain beliefs and intentions.[3]

3 Context and the Discourse Situation

At the very least, we want a unified model of context to include, in addition to information about discourse referents, information about which speech acts have occurred in a conversation and the subordination relationships between them. This information is needed both by all modules of a system like TRAINS that make use of context: the reference resolution module (to build up segments of discourse and implement pragmatic constraints on accessibility), the speech act recognition module of course, and the dialogue manager (to decide what to do next).

Including speech acts in the discourse model entails, however, assuming a different position with respect to the discourse model from the one adopted in DRT. In DRT, the discourse model captures the truth conditions of a text; but the fact that speaker A told B that P is not part of the truth conditions of a sentence. The proposal of including speech acts in the discourse model amounts to a shift from a model of sentence *meaning* to a model of sentence *use*. We will show that the tools developed in DRT can be put to this use without altering their interpretation.

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

[3] See [Asher, 1986; Kamp, 1990] for examples of work in which this latter concern is addressed.

As we still want to be able to handle reference problems, we would like to preserve as much of DRT as possible and simply reinterpret what the DRT representation we are constructing says. Basically, we want to use the root DRS as a representation of the discourse situation, whereas the content of speech acts is a representation of one or more described situations. Issues of both a conceptual and a technical nature have to be addressed when doing so. At the very least, it is necessary to make sure the accessibility of discourse referents is still preserved in the new representation. For example, a representation as in (6) for the made-up mini-dialog in (5) between speakers A and B, under the standard semantics for semantics of DRT, would not make the discourse referent x 'evoked' by the NP *engine E2* accessible to the pronoun *it* in the next sentence, represented in (6) as the discourse referent u.

(5)
A: Engine E2 is at Avon.
B: The boxcar is also at Avon and is hooked to *it*.

(6)

tell(A,B, [x w | **engine**(x) **Avon**(w) **at**(x,w)])

tell(B,A, [y u | **boxcar**(y) **at**(y,w) **hooked-to**(y,u) $u = x$])

A more general problem with (6) is that it doesn't say that both assertions are 'about' the same situation. 'Vanilla' DRT allows us to specify the semantics of predicates one of whose arguments is a DRS in two ways. We could treat **tell** as an extensional predicate, i.e., we could let the embedded DRS be evaluated with respect to the same situation in which the condition asserting the occurrence of the telling event is evaluated. But in this way we would 'mix' facts about what's going on in the situation in which A and B are speaking, with facts about the situation they are talking about. Under this interpretation, in fact, (6) would assert of a single situation that in that situation, x is an engine and is at Avon, y is a boxcar, is at Avon and is hooked to x, and also that in that same situation, A tells B that x is at Avon, and B tells A that y is at Avon and hooked to x. This interpretation is often unproblematic, but in conversations the speakers may be talking about situations distinct from the one that includes the discourse situation. In the TRAINS conversations,[4] for example, assertions such as those in (5) might not be describing

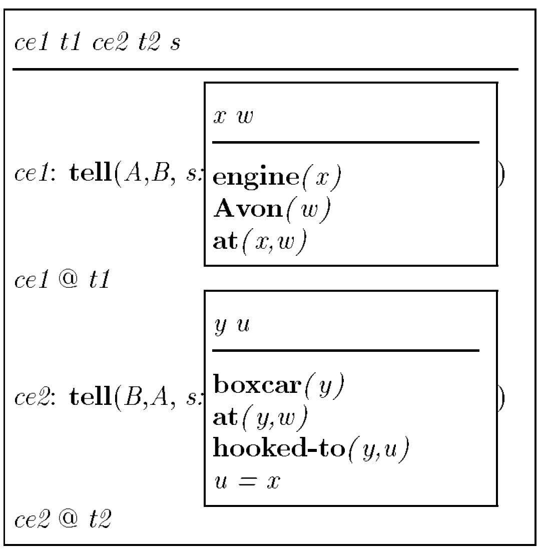
[4] Transcribed in [Gross *et al.*, 1993].

the state of the world which A and B are currently in, but, rather, a possible state of the world in the future, resulting from a few steps of the plan that A and B are trying to agree upon.

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

A version of DRT that addresses these issues and is based on the hypotheses about the common ground made in Situation Semantics, called CONVERSATION REPRESENTATION THEORY (CRT), was introduced in [Poesio, 1994]. In CRT much of the formal machinery of DRT is preserved, including the update algorithm, but the Root DRS is taken to consist of information about the occurrence of speech acts or, as we will call them here, CONVERSATIONAL EVENTS, and each utterance introduces two situations: a conversational event[5] and a described situation. The dialog in (5) is represented in CRT as in (7).

(7)



The DRS in (7) represents a discourse situation in which two conversational events occurred, *ce1* and *ce2*, both of which are about the described situation *s*. The first two conditions assert that *ce1* is an event of A telling B that the described situation *s* includes an engine located at Avon, and that *ce1* took place at time *t1*. The third and fourth condition assert that *ce2* is an event, taking place at time *t2*, of B telling A that boxcar *y* is also in Avon, and is hooked to *u*.

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

The problems with (6) discussed above are addressed by modifying the interpretation of DRSs and by adding a new type of condition, indicated in (7) by expressions of the form $s{:}\phi$. A DRS is treated in CRT as a *situation type*, i.e., as denoting the set of situations which can be made to verify the conditions contained in the DRS once appropriate values for the discourse referents have been found. Conditions like *s*:K, then, assert that a situation *s* is of the type specified by the DRS K. More precisely: Expressions in CRT are assigned a value with respect to a situation, a variable assignment, and a set of CASES, one for each situation. An expression of the form *s*:K asserts that *s* is of the type specified by K, and furthermore, it shifts the parameters of evaluation so that the value of the discourse markers occurring in K is provided by the case associated with the value of *s*. This ensures that the DRS in the complement of *ce2* is evaluated with respect to the same case that is used to evaluate the expressions in the complement of *ce1*, *e.g.*, that *x* is accessible from within the complement of *ce2*. Conditions of the form *s*:K resolve both the problem of specifying the situation of evaluation for the propositional complement of locutionary acts, and the problem of guaranteeing that discourse referents introduced by one conversational event are accessible from the next.[6]

The update algorithm of DRT is (minimally) modified as follows. The process of updating the existing discourse model when a new utterance takes place is initiated by CONVERSATIONAL EVENT GENERATION RULES that update the existing model of the discourse situation by adding two new discourse referents (a conversational event and a time) and two conditions, one asserting the occurrence of the appropriate locutionary act[7] (one of **tell**, **ask**, and **instruct**), the other recording the time at which it occurred. The process of DRS construction can then proceed much as in [Kamp and Reyle, 1993], except that the whole process takes place within the DRS that is the complement of the locutionary act. DRSs like (7) are the final result of this process.[8]

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

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

[8]Whereas the third argument of the locutionary act **tell** is a proposition of the form *s*:K, the third argument of the locutionary act **ask** is a QUESTION–the denotation of expressions of the form (s?K)—and the third argument of the locutionary act **instruct** is a situation type. Several ways of specifying what kind of semantic object a question is have been proposed in the literature, among which the best known are the proposals of Karttunen [Karttunen, 1977] and Groenendijk and Stokhof [Groenendijk and Stokhof, 1984].

4 Mental States, Speech Acts, and Discourse Structure

All the pragmatic facts that play a role in speech act recognition, reference resolution, and dialog management—i.e., facts about the (mutually known) intentions, beliefs, and obligations of the participants in a conversation—can also be characterized as states that are part of the discourse situation, and can be represented by conditions of the form s:K. For example, the fact that A intends engine x to be at Bath in some situation s' which extends s (and therefore 'inherits' all the individuals that occur in s) can be thought of as a state *i1* holding at time *tj*. The occurrence of this state in a discourse situation can be represented by including in the Root DRS a condition that expresses the presence in the common ground of an intention *i1* of A, as follows:

(8)

ce1 t1 ce2 t2 s ... *s′ i1 tj*
...
...
$s \sqsubseteq s'$
i1: **intend**(*A*, *s′*: [**at**(*x,Bath*)])
i1 @ *tj*

Once the discourse situation has been updated through use of the conversational event generation rules, there follows a process of illocutionary act recognition: an attempt to determine what was actually done by the speaker in performing the utterance. Context plays a crucial role in this process. As described by Austin [Austin, 1962] and others, utterance of the same sentence in different contexts may be a performance of radically different actions. Relationships to previous utterances, and hypotheses about the speaker's mental state (including beliefs, and local and global goals and intentions) will be crucial in forming and evaluating hypotheses about the illocutionary act that has been performed. For instance, in the example developed above, the conversational event *ce1* might be any of **inform** of the engine's location, a **check** to make sure that the agents agree, a **suggestion** to use the engine in a developing domain plan, etc.

Illocutionary act recognition algorithms based on those discussed in [Allen and Perrault, 1980; Hinkelman, 1990; Traum and Hinkelman, 1992] have been developed in the TRAINS System [Allen *et al.*, 1995]. Following the implemented system, we adopt here the multistratal 'Conversation Acts' theory, presented in [Traum and Hinkelman, 1992]. This theory maintains the classical illocutionary acts (e.g., **suggestion**) as one level (called CORE SPEECH ACTS) which are now, however multi-agent collaborative achievements, taking on their full effect only after they have been *grounded* (see below). The theory also adds levels of action for discourse relations (called argumentation acts), grounding, and turn-taking.[9]

The core speech act(s) performed by a speaker by means of an utterance are also added to the discourse situation. We think of these core speech act(s) as being *generated* by the utterance in the sense of Goldman [Goldman, 1970]. Assuming the **suggestion** interpretation of the locutionary act *ce1* in (7), the information acquired via the speech act recognition process can be represented as in (9).

(9)

... *sug1*
sug1: **suggest**(*A*, [**use**(*{A,B},x,pl*)])
generate(*ce1,sug1*)

This means that (one of) the actions performed in the utterance of *ce1* is a suggestion that the agents use x in their ongoing domain plan. This will have the further (perlocutionary) effect of B trying to incorporate the use of this object into his idea of the already developing plan, which might necessitate further inference of what might have been implicated by *ce1*. These further implicatures will also be added into the DRS along with the suggestion.

An additional result of speech act recognition is the establishment of relations of dominance or precedence between the new illocutionary act and the existing set of acts. Facts encoding Grosz and Sidner's intentional structure, or more general rhetorical relations [Mann and Thompson, 1987], can then be added to the discourse situation (in the form of ARGUMENTATION ACTS [Traum and Hinkelman, 1992]).

The second aspect of Grosz and Sidner's model of discourse, the grouping of utterances into 'discourse segments' determined by the intentional structure, can be modeled in terms of relations between situations. It is assumed in CRT that the grouping of illocutionary acts into 'discourse segments' is analogous to the grouping of events into 'threads' or 'stories' in narratives. We call the threads of conversational events CONVERSATIONAL THREADS. Threads are also situations of a special type ('plural' situations), therefore the relation between the conversational event *sug1* in (9) and its conversational thread (let's call it *ct1*) can be represented as a subsituation relation, as follows:

(10)

...
...
sug1 $\sqsubseteq$ *ct1*

The illocutionary act(s) performed in an utterance may either become part of one of the existing conversational

[9] A similar theory was presented in [Novick, 1988].

threads, or originate a new one. DISCOURSE SEGMENTS are a a type of conversational threads, that consist of core speech acts. (We will see another case of conversational threads below.) We assume that all core speech acts performed by an utterance become part of the same minimal discourse segment.

Finally, the third argument of a conversational event (the expression s:K) specifies how a conversational event is going to affect the FOCUS SPACE STACK in Grosz and Sidner's sense. In the formalism used here, the kind of 'pragmatic accessibility' information that, in Grosz and Sidner's formalism, is specified by the focus stack, is given by the relations of inclusion between situations: a situation s that extends another situation s' includes all the information of s', as well as additional information. A conversational event may either introduce a new described situation (e.g. situation s, in *ce1*), which is analogous to the case in which a new stack is created and a new focus space added on top of it; or extend the situation described by previous conversational events, which corresponds to the case in which an utterance 'adds' material to an existing focus space on the stack (e.g., the re-use of s in *ce2*); or introduce a new described situation which extends an existing one (as in the case in which a subplan is being discussed), which corresponds in Grosz and Sidner's terms to the case in which a new focus space is pushed on top of the focus space stack, allowing access to the previous focus spaces. E.g., after introducing a (possible) situation s by saying something like *We need to get the boxcar to Avon by 5pm*, a speaker might continue by saying *We need an engine to move the boxcar*. The third argument of the conversational event associated with this new utterance describes a situation s' such that $s \sqsubseteq s'$.

5 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. 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 immediately becomes a part of the common ground and, hence, is available for reference.[10] As shown, e.g., in [Clark and Wilkes-Gibbs, 1986; Clark and Schaefer, 1989], however, this assumption is simply an idealization. Utterances by one party must be recognized and acknowledged by the other before entering common ground. The collaborative process of adding to the common ground (called GROUNDING) must include installments by each of the conversants. These installments may be explicit —after the first utterance of (5), B may have acknowledged by uttering something like *okay* or *right*—or tacit —the second utterance in (5) 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:[11]

(11)
```
3.1  M: now
3.3   : so
3.4   : need to get a boxcar
3.5   : to Corning
3.6   : where there are oranges
3.7   : there are oranges at Corning
3.8   : right
4.1  S: right
5.1  M: so we need an engine to move the
        boxcar
5.2   : right
6.1  S: right
7.1  M: so there's an engine
7.2   : at Avon
7.3   : right
8.1  S: right
```

Clark and Schaefer presented an off-line model of grounding in [Clark and Schaefer, 1989], while a computational account was provided in [Traum, 1994] and implemented within the TRAINS system. In [Traum, 1994], participation in the grounding process is seen as the performance of GROUNDING ACTS, one of the levels of conversation acts from [Traum and Hinkelman, 1992]. When an agent utters something, in addition to performing core speech acts like those discussed in the previous section—such as **suggest**, **inform**, etc.—the agent is also performing one or more grounding acts. Presented material that could be acknowledged together (e.g., with a single *okay*) is grouped into a DISCOURSE UNIT (DU). Grounding acts include **init**, which opens a new DU, **continue**, which adds more material to an already open DU, and **ack**, which makes the contents of the DU enter the common ground. Each DU has its own state, representing whether or not the DU has been grounded and which kinds of actions (e.g., acknowledgements or repairs) 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 stack-like structure of accessible DUs is maintained in the model, as context for recognizing and performing grounding acts.

The model of context discussed in the previous section already includes two of the features needed to represent the grounding process, namely, the occurrence of conversational events, and the possibility of organizing conversational events into threads. A fairly minor augmentation is the addition of a new relation, **ga**, which holds between a locutionary act and a grounding act if the grounding act has been generated by the perfor-

[10] Most work on reference resolution makes the further assumption that felicitous reference requires mutual knowledge [Clark and Marshall, 1981].

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

mance of the locutionary act. For example, in the exchange in (5), 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 material that it contains in its own right. We can represent this by augmenting the Root DRS with the conditions shown in (12). We have used *du1* for the DU acknowledged by B's utterance (containing *sug1* and other material from *ce1*), and *du2* for the DU that *ce2* initiates. Both *du1* and *du2* are conversational threads.[12]

(12)

... du1 ... ack1 du2 init2
...
ack1: **ack**(*B*,*du1*)
init2: **init**(*B*,*du2*)
ga(*ce2*,*ack1*)
ga(*ce2*,*init2*)

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 participants to a conversation mutually believe (or, perhaps, what one participant believes that is mutually believed) without worrying about 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.

We propose to model the process as follows. We modify the DRS construction algorithm so that the result of an update due to a conversational event generating an **init**, **continue**, or **repair** grounding act is always an ungrounded DRS. We then define acknowledgments as moves ungrounded to grounded states.

The discourse situation will now consist of two parts, both accessible for reference purposes, though in subtly different ways. First, as before, we have the root DRS, representing the common ground. Information within is assumed to be shared by both dialogue participants. In addition, we have an EXTENDED ROOT DRS (ER-DRS), consisting of those conversational threads which represent ungrounded DUs. Each DU thread within the ER-DRS represents a particular agent's view of what should be in the common ground. Items in the ER-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 a DU initiated by the other agent can be a means of acknowledging that thread. In this manner, we can also model the collaborative nature of the referring process itself, in a manner similar to [Clark and Wilkes-Gibbs, 1986; Heeman, 1991]. [13]

[12] Although DUs are represented using the same technical device as Discourse Segments (namely, conversational threads), the two concepts should not be confused. Grounding and intentional discourse structure are two orthogonal phenomena. Any given core speech act will be part of two conversational threads, one representing its groundedness, and another representing its topical content. E.g., *sug1* is a part of both *ct1* and *du1*.

6 Other Issues

Space precludes a detailed explanation of how other contextual features are represented and used by other system modules, but we briefly note some of the issues here.

Dialogue Management

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

Repair and Attitude Revision

Many of the processes described above, including speech act interpretation, reference resolution, and decisions on grounding, are provisional assumptions, not guaranteed to be accurate, when performed by one conversational participant in an on-line manner. We are

[13] The formalization of this distinction between root DRS and ER-DRS, and of the effect of grounding acts, are based on the hypotheses about mental states and their connection with discourse interpretation developed in work by Asher and Kamp, among others, [Asher, 1993; Kamp, 1990]. Asher and Kamp see discourse interpretation as a process of operations on 'mental states', collections of objects of type ⟨M,K⟩, where M is a 'mental attitude' and K is a DRS, where DRSs also denote objects that correspond fairly closely to Frege's 'senses'; the statements about the world that these objects represent are specified by a separate, 'referential' interpretation, which is just the semantics seen in the previous sections. Space prevents a full discussion of the matter.

currently working on including a model of repair phenomena to the unified model. For some preliminary ideas in a related situation theory, see [Traum, 1994].

Incrementality

The actual spoken input to the TRAINS system does not consist of complete utterances. In order to process repairs, for example, a conversational system has to be able to understand and respond to partial utterances as well, as shown in the following example:[14]

(13)
```
9.1  M: so we should
9.2   : move the engine
9.3   : at Avon
9.4   : engine E
9.5   : to
10.1 S: engine E1
11.1 M: E1
12.1 S: okay
13.1 M: engine E1
13.2  : to Bath
13.3  : to /
13.4  : or
13.5  : we could actually move it to
        Dansville to pick up the
        boxcar there
14.1 S: okay
```

The need to account for this kind of input played in fact a major role in the development of the model of discourse discussed in this paper—for example, in adopting a speech act-based model of segmentation rather than a model along the lines of [Asher, 1993]. We believe in fact that the discourse model introduced above can be extended to accomodate this kind of input simply by allowing for MICRO CONVERSATIONAL EVENTS in the discourse situation—events associated with utterances of elements smaller than sentences—in addition to the conversational events that correspond to 'traditional' speech acts. In other words, the discourse model will contain not only facts of the form 'a locutionary event of type **tell** occurred', but also facts of the form 'an event of A uttering the word-form *to* occurred'. The current status of this work is described in [Poesio, 1995].

7 Summary

As we have shown above, there are a large number of contextual issues that confront the designer of a system to engage in natural language conversation. Often the most natural representation of context for performing one type of processing makes it more difficult or impossible to do other types of processing. Based on experiences in confronting and overcoming just this type of issue within the TRAINS system, we have presented the beginning stages of a unified general-purpose model of context. The model already can represent several types of discourse structure, as needed for reference resolution, speech act recognition, and dialogue management, in a way that is accessible for on-line processing algorithms. Current work includes extending this representation to include models of repair and incremental interpretation, and implementation of new algorithms working within the unified framework.

[14] From the same dialog discussed before.

Acknowledgements

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

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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