A Situation-Theoretic Formalization of Definite Description Interpretation in Plan Elaboration Dialogues*

Massimo Poesio
Department of Computer Science
University of Rochester
Rochester, NY 14627-0226, USA
poesio@cs.rochester.edu

1 Introduction

That pragmatic factors such as discourse segmentation (Grosz and Sidner 1986, Fox 1987), the current 'focus of attention' (Grosz 1977, Linde 1979), or what is mutually known by speaker and addressee (Hawkins 1978, Clark and Marshall 1981) play an important role in the interpretation of pronouns and definite descriptions has been repeatedly demonstrated. Yet, the formal study of referential expressions in natural language has so far concentrated on semantical questions, especially on the truth-conditional aspects of pronominal reference and on the dynamics of the information used in interpreting pronouns (Cooper 1979, Evans 1980, Kamp 1981, Heim 1982, Barwise 1987, Groenendijk and Stokhof 1991).

These authors usually leave the task of formalizing the pragmatic aspects of the interpretation of referring expressions for further research. This state of affairs is due in part to the fact that pragmatic phenomena are still poorly understood; in part to the number and diversity of phenomena that have been characterized at one time or the other as 'pragmatic;' and in part, finally, to the lack of appropriate formal tools. The only exception I am aware of is the work by Barwise and Perry in *Situations and Attitudes* (Barwise and Perry 1983). Barwise and Perry provide a framework in which the pragmatic aspects of interpretation can be formalized; they do not, however, provide either an account of pragmatic factors such as those mentioned above, or of the pragmatic processes involved in interpretation. The model of discourse interpretation proposed by Grosz and Sidner (1986), then, is at the moment the more comprehensive account of the structure of information in discourse and its use in interpreting referring expressions. This model is not, however, specified in enough detail that its predictions can be readily verified, and doesn't include a specification of the process by which the discourse structure used to interpret referring expressions is updated. Thus, it cannot be used as the basis for the implementation of the

*This paper appears in Aczel, P. and Israel, D. and Katagiri, Y. and Peters, S. (editors), Situation Theory and its Applications, vol. 3, 1993, p.339–374.

definite description interpretation module of a natural language processing system.

I report in this paper the current status of work whose ultimate goal is to arrive at a precise characterization of the inferential processes involved in interpreting definite descriptions in conversations. I present detailed proposals concerning both (i) the organization of the information used to assign definite descriptions their interpretation, and (ii) the 'dynamics' of this information —i.e., how utterances modify it. The resulting framework allows me to formalize the process of definite descriptions interpretation by means of defeasible axioms, along the lines of Perrault's (1990) account of intention recognition.

More precisely, I reconstruct in Situation-theoretic terms both the *Location theory of definite reference* (Hawkins 1978, Clark and Marshall 1981) and Grosz and Sidner's proposal about the organization of the information used to interpret referring expressions (that they call the *Attentional State*). These aspects of my proposal are discussed in §3 and §4, respectively. This reformulation provides the tools to formulate *principles for anchoring resource situations* that predict whether a definite description is going to be interpreted anaphorically or with respect to the 'visible situation.' My account of the reasoning process involved in interpreting 'visible situation' uses of definite descriptions is presented in §6.2.

As for the problem of providing rules specifying the temporal evolution of the attentional state, I propose to adopt the *action-based model of mental state update* developed in the literature on intention recognition (Cohen and Perrault 1979, Allen and Perrault 1980, Cohen and Levesque 1990, Perrault 1990). I show in §5 that a model in which the effect of an utterance is for the hearer to acquire beliefs about the occurrence of a communicative action does not simply afford a closer integration between the processes tracking attentional state change and those performing intention recognition.[1] Crucially, such a model also gives the opportunity to formalize a number of phenomena that are difficult or impossible to deal with otherwise. Among these phenomena, I consider here the organization of utterances into discourse segments and the way the 'focus of attention' shifts during the conversation (Grosz 1977, Linde 1979, Grosz and Sidner 1986).

In order to formulate principles of pragmatic interpretation it is essential to look at the way speakers behave in actual conversations. The data on which my analysis relies consist of transcripts of recorded conversations, collected and analyzed within the framework of the TRAINS project (Allen and Schubert 1991). These transcripts provide data about a restricted class of dialogues in a specific domain—conversations concerned with the development of plans to accomplish a task in a transportation

[1]For example, in the TRAINS project (Allen and Schubert 1991), in which this work is being applied, this model allows for the same representation to be used by both the module tracking the attentional state and the modules performing intention recognition.

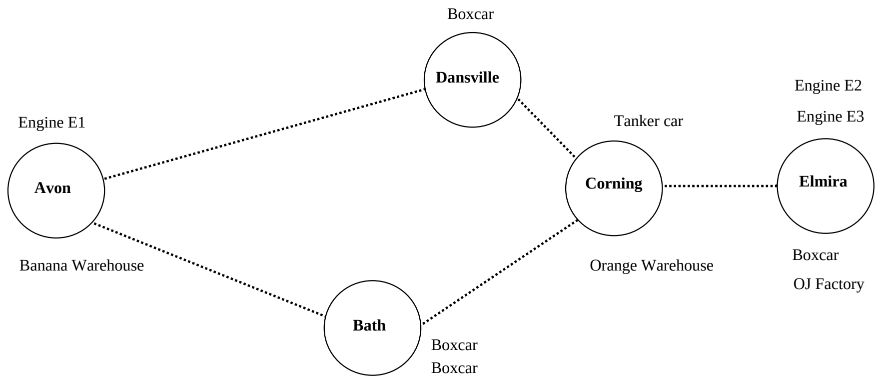

Figure 1 The map used by the participants in the conversation.

domain—which made it possible to limit the scope of my task. These data are described in §2. The work presented here can thus be summarized as a situation-theoretic analysis of the pragmatic inference processes involved in interpreting the two patterns of definite descriptions usage most common in our transcripts, the 'anaphoric use' and the 'visible situation use.' (This terminology will be discussed below.)

2 The Data

The aim of the TRAINS project (Allen and Schubert 1991) is to develop a natural language understanding system able to engage in conversations with a human *user* whose task is to develop plans for transporting goods by train. The role of the system in these conversations is to assist the manager in developing the plan. An example of the kind of tasks the manager has to develop plans for is given in (1).

(1) I have to get one tanker of orange juice to Avon, and a boxcar of bananas to Corning, by 3pm.

The development of the system is driven by the study of transcripts of recorded conversations.[2] The conversations we transcribe take place between two human speakers, one of which plays the role of the system while the other plays the role of 'user.' The 'user' and the 'system' are separated by a wall, and communicate via microphones and headphones. Each of them has a copy of a map like the one in Fig. 1.

[2]The transcripts of the conversations collected in 1991 (on which this paper relies) are presented in (Gross et al. 1992).

The discussion in this paper is concerned, for the most part, with the (edited) fragment in (2). The user's utterances are marked with 'U', the system's utterances with 'S.'[3]

(2) ...
13.1 U: not at the same time
13.2 okay
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
14.1 S: okay
14.2 We could use one engine to take both the tanker and the boxcar to Elmira
15.1 U: oh, we can do that?
16.1 S: yeah
17.1 U: then bag the whole thing with engine E3
17.2 and just hook up the tanker car with the boxcar that has oranges in it,
17.3 and take it back to Elmira
18.1 S: okay,
18.2 that's no problem.
...
29.1 U: okay,
29.2 great
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

Of the eight major usage types of the definite article in English (Hawkins 1978), two are especially common in our transcripts. The definite descriptions "the boxcar" in `13.3` and `29.5` are instances of *visible situation use* of definite NPs, which occurs when "...the object referred to is visible to both speaker and hearer in the situation of utterance, and is furthermore unique." (Hawkins 1978, p.110). In order to model the visible situation use, we need to represent the fact that the speaker's attention is focused at certain times on some objects, and that this *focus of attention* changes during a conversation (Grosz 1977, Linde 1979). The plan discussed in (2)

[3]Each utterance has an identification tag consisting of two numbers: the first number is the turn number, while the second indicates the utterance position within its turn. This notation is due to David Traum.

involves two boxcars, one in Elmira and one in Dansville. In `29.5` the focus of attention is apparently Dansville, since the reference to "the boxcar" is unambiguous even though three other boxcars are shown in the map. Yet, Dansville clearly isn't the focus of attention during the whole dialogue, since another boxcar is discussed in `13.3-16.1`, and at no moment in the discussion do the manager and the system seem to perceive an ambiguity, not even when "that boxcar" is used in `13.5`.

Grosz and Sidner (1986) propose that changes in the attentional state are the result of "...a set of transition rules that specify the conditions for adding and deleting spaces (p. 179)." Because Grosz and Sidner's intention is to develop an *abstract* rather than a *processing* model of discourse structure (p. 176), they do not address questions like: What form are the transition rules going to take? Under which conditions are they going to be applied—when, for example, does an utterance trigger a transition rule? How do the transition rules relate to the rules for modifying the intentional and linguistic structures? Without an answer to these questions it is hard to understand clearly what the theory predicts. A formal theory of discourse must provide an answer to these questions.

The definite descriptions "the boxcar" in `14.2` and "the boxcar" in `31.2` are cases of *anaphoric* use of definite NPs. According to Hawkins, we have an anaphoric use when the definite article is used to refer to an object explicitly 'entered into the memory store' by a previous NP (Hawkins 1978, p.86). The fragment in (2) illustrates another well-known fact about definite descriptions, namely, that when a definite description is used anaphorically, the only antecedents considered are those in the same *discourse segment* (Reichman 1985, Grosz and Sidner 1986, Fox 1987). For example, "the boxcar" in `31.2` is unambiguous, even though more than one boxcar has been mentioned in the dialogue.

The participants in our conversations refer to objects and events which are part of the plan as if they were actual objects and events which actually occurred. In the following fragment, for example, "a boxcar" is introduced into the plan by the user in sentence `3.1`, and then referred to in sentence `5.1`, without the user specifying which boxcar in the map he has in mind, if any.

(3)
1.1 U: okay, the problem is we better ship a boxcar of oranges to Bath by 8 AM.
2.1 S: okay.
3.1 U: now ... umm ... so we need to get a boxcar to Corning, where there are oranges.
3.2 there are oranges at Corning
3.3 right?
4.1 S: right.
5.1 U: so we need an engine to move the boxcar
5.2 right?

6.1 S: right.

Finally, (2) illustrates the need for interaction between the processes tracking attentional state and those performing intention recognition, recognized early on by Hobbs (1979). Consider `31.2`, for example. If the interpretation of the anaphoric definite "the boxcar" were to take place before intention recognition has been performed, the discourse segment which includes `31.2` would not have been determined yet, hence all potential referents ought to be considered. Conversely, if intention recognition were to take place before the referent for "the boxcar" has been identified, the plan reasoner ought to verify which action among all the actions involving boxcars in the plan is being discussed. The most crucial contribution of Grosz and Sidner was to provide an hypothesis about how discourse segmentation and intentional structure might be related.[4]

3 The Semantics and Pragmatics of Definite Descriptions

3.1 The Semantics of Definite Descriptions

The Russellian approach to the semantics of definite descriptions (Russell 1905) is enjoying renewed interest these days, thanks to the work of Grice (1969), Kripke (1977) and Neale (1990). According to Grice, it is necessary to distinguish between what the speaker *says* and what the speaker *means*. The truth conditions of an utterance of a sentence of the form 'the F is G' are thus strictly Russellian even when 'the F' is used referentially; "...the speaker may, however, wish to get it across to the hearer that a particular individual b is G, and may succeed in doing this by exploiting the fact that both speaker and hearer take b to be the F (Neale 1990, p.9)." So, while "...it is surely not open to dispute that a sentence of the form 'the F is G' may be used to *communicate* an object-dependent thought to someone, to the effect that some particular individual b is G (Neale 1990, p.7)," in our daily talk we very often convey things "...indirectly, relying on what we take to be our interlocutors' abilities ...to grasp ...what *we* mean by our utterances (Neale 1990, p.9)."

Neale implements this proposal by introducing in the object language a determiner THE defined as in (4), and by assuming that definite NPs such as "the boxcar" have a unique translation of the kind in (5) (Neale 1990, p. 45; to be refined below):

(4) '[THE x: F(x)]G(x)' is true iff $||\mathbf{F} - \mathbf{G}|| = 0$ and $||\mathbf{F}|| = 1$

(5) "the boxcar" $\rightsquigarrow \lambda$ Q [THE x: BOXCAR(x)] Q(x)

Whether this analysis is the definitive word on the semantics of definite

[4] An interesting phenomenon that I won't be able to discuss here is the fact that the participants in our conversations may refer to certain objects by means of descriptions which can only be interpreted if the hearer has 'kept track' of the plan—for example, the definite "the boxcar that has oranges in it" in `17.2` of (2).

descriptions remains to be seen; it does, however, account for an impressive array of facts. I have adopted the neo-Russellian approach in this paper, but my proposal would also be compatible with a treatment of definite descriptions along the lines of Heim's (1982).

3.2 The Location Theory

The *location theory of definite descriptions* (Hawkins 1978, Clark and Marshall 1981) is, arguably, the most widely accepted account of the processes resulting in an hearer's assigning a referential interpretation to a sentence containing a definite NP. According to Hawkins (1978, p.167), the defining aspects of this process are that:

1. "...the hearer is instructed to *locate* the referent in some *shared* set of objects." (Emphasis added.)
2. "the speaker refers to the *totality* of the objects/mass within this set that satisfy this restriction." (Emphasis added.)

This characterization includes both an hypothesis about the semantics of definite descriptions (point 2), and one about the process of pragmatic interpretation (point 1). According to the location theory, the aim of definite description interpretation is to identify the shared set of objects within which the referent of the definite is located. Clark and Marshall (1981) make the notion of 'shared set' used by Hawkins more precise. They argue that what is needed is the notion of *mutual knowledge* introduced by Lewis and Harman (Lewis 1969, Harman 1977) among others. Clark and Marshall adopt the 'recursive' definition of mutual knowledge proposed by Harman ((1977), cited by (Clark and Marshall 1981), p.17):

(6) A and B *mutually know* that p $=_{\text{def}}$ (q) A and B know that p and that q.

Clark and Marshall's main concern is to explain how people can infer mutual knowledge, which apparently requires checking an infinite amount of conditions of the form 'A knows that B knows that .... that p.' Their solution is based on Lewis': if A and B make certain assumptions about each other's rationality, they can use certain states of affairs (*grounds*) as a basis for inferring the infinity of conditions all at once. In the case of a 'visible situation' use of definite descriptions, for example, the grounds consist of two parts: direct visual evidence of copresence, and assumptions about the situation—that the other participant in the conversation is consciously attending, that he is rational, and so forth.

Additional interpretive processes are known to be involved in the interpretation of definite descriptions (Cohen 1984), but in order to interpret the definite descriptions found in the current set of TRAINS transcripts all that the hearer is required to do is to identify a suitable 'shared set.' Therefore, I only present an account of this process in this paper. In the rest of this section, I discuss my assumptions about the semantic treatment of definite descriptions, and propose a simple formalization of the location theory.

3.3 A Situation-Theoretic Reconstruction of the Location Theory

The Russellian approach to the semantics of definite descriptions is compatible with the semantic part of the location theory. The first part of my proposal is to formalize the aspects of the location theory related to pragmatic interpretation by modifying the translation for definite descriptions in (5) as follows:

(7) "the boxcar" $\rightsquigarrow \lambda$ P [THE x ([$\dot{S}$ $\models$ BOXCAR(x)] $\wedge$
SHARED(**spkr**,**hearer**,$\dot{S}$))]
P(x))

According to (7), "the boxcar" denotes the set of properties P such that the relation THE holds between P and the property of being a boxcar in a shared situation denoted by the *situational parameter* $\dot{S}$. The terms **spkr** and **hearer** in (7) are indexicals, referring to speaker and hearer, respectively. Following (Barwise and Perry 1983, p.145), I call the situation s which includes the objects quantified over by a determiner the *resource situation* of that determiner; $\dot{S}$ is the resource situation of the determiner THE in (7).

I make use in (7) of a relation $\models$ ('supports') to represent 'truth at a situation' (Devlin 1991). A statement of the form

$$[s_1 \models \text{DOG}(x)]$$

evaluates to true if the object—say, d—assigned to the variable x is a dog in the situation denoted by s_1. A *situation* is a set of objects and facts about these objects (Devlin 1991, Hwang 1992). I need a language that allows me to make statements about situations, and an ontology in which situations are objects in the universe.[5] *Episodic Logic* (Hwang 1992, Hwang and Schubert 1993) provides such a language and such an ontology; where not otherwise noted, the reader should assume that the expressions I use here have the semantics specified in (Hwang 1992). The main exception is the operator $\models$, that is similar to, but not identical with, the '*' operator used in Episodic Logic. The expression [s_1 * DOG(x)] of Episodic Logic denotes a function from situations to truth values[6] which returns true in a situation s if the object d assigned to the variable x is a dog in the situation denoted by s_1, as above, *and* the situation denoted by s_1 is *actual* in s. The

[5] I use variables with an 'e' suffix like e or ce to denote events, and variables suffixed with 's' like s or ds for situations and statives . I also adopt the convention of using square brackets for infix operators like $\models$. Finally, I use 'numbered variables' such as ce29.5, e29.5 etc. throughout the translation of the utterance numbered 29.5.

[6] An important characteristic of Episodic Logic is that formulas denote functions from situations to truth values, instead of truth values—the expression DOG(x), for example, is of type $\langle s,t \rangle$.

problem with this is that, for my description of the attentional state in §4 I'll need to be able to characterize *possible* situations, so I need a weaker operator than '*.' To avoid confusion, I call the logic used here $\mathcal{SEL}^{\mathcal{D}}$.[7]

The statement $[\dot{\mathtt{S}} \models \text{BOXCAR}(\mathtt{x})]$ in (7) contains a *parameter*, $\dot{\mathtt{S}}$. I use parameters to translate referential expressions. A parameter behaves semantically as an open variable, the value of which has to be provided by context.[8] Heim uses open variables for this purpose in (Heim 1982), ch. 2, and characterizes an expression as *felicitous* if context can provide a value for all of its open variables. By introducing a distinct syntactic sort, I make it explicit that the value of the expression has to be identified; my notation is otherwise semantically equivalent to Heim's.[9] A parameter can be *anchored* by logical statements of the form $\dot{\mathtt{x}} = y$ that constrain the evaluation function to assign to $\dot{\mathtt{x}}$ whatever value is assigned to y.

The location theory can be reformulated in terms of parameters as follows: the aim of the pragmatic processes involved in definite description interpretation is to find an appropriate anchor for the situational parameter denoting the resource situation of the determiner THE—i.e., to recognize an intention of the speaker's of the form:

$$\text{INTEND}(\mathbf{spkr}, \dot{\mathtt{S}} = \mathtt{s})$$

where $\dot{\mathtt{S}}$ is the resource situation of the definite and $\mathtt{s}$ is the situation the speaker intends the hearer to 'locate' the referent of the definite in. The task of a theory of definite description interpretation, then, is to provide *principles for anchoring resource situations* which generate hypotheses about the intended identity of the resource situation of the determiner.

A situation $\mathtt{s}$ is 'shared' between $\mathtt{x}$ and $\mathtt{y}$ if every fact Ψ supported by $\mathtt{s}$ is mutually known by $\mathtt{x}$ and $\mathtt{y}$, as from the following schema:

(8) $[\forall\ \mathtt{x,y,s}]\ \text{SHARED}(\mathtt{x,y,s}) \equiv ([\mathtt{s} \models \Psi] \supset \text{MK}(\mathtt{x,y},\Psi))$

The relation MK (for 'mutual knowledge') between two agents and a proposition is defined as in (6) (above).

3.4 Linguistic Meaning and Speaker's Meaning

The neo-Russellian, or Gricean, strategy presupposes the existence of two distinct levels of interpretation: a level at which semantic generalizations are stated (the *linguistic meaning*) and a level at which the intentions of the speaker are spelled out, called *speaker's meaning* by Grice ((1957; 1989, chap.14), quoted in (Chierchia and McConnell-Ginet 1990)). The task of

[7]For 'Simple Episodic Logic with Defaults.' The reason for this name will become clear in §5.

[8]The reader should be aware that while the notation and terminology I have adopted are borrowed from Situation Theory, parameters have a different semantic interpretation there (Devlin 1991).

[9]The idea is to add to the parameters of evaluation an *anchoring function a* that provides the values for parameters, thus plays the role of 'context' in Heim's proposal.

pragmatic interpretation is to identify the speaker's meaning on the basis of linguistic meaning and context.

I follow here the approach adopted in the work on intention recognition in Artificial Intelligence, in which recognizing the speaker's intentions involves reasoning about the motives of the speaker in performing a certain *conversational event*.[10] I propose to implement this approach by adopting a system of interpretation similar to the one used by Karttunen and Peters to deal with conversational implicatures and presuppositions (Karttunen and Peters 1979).

In the system of Karttunen and Peters, two translations of an utterance are computed: the first gives its extensional value, the second specifies its conversational implicatures. In the system presented here, the separation between the 'linguistic' and 'pragmatic' aspects of interpretation is also implemented by associating two distinct translations to each utterance. The first translation represents the utterance's truth conditions, and corresponds to Grice's linguistic meaning. Following Schubert and Hwang (1993), I call this interpretation the *Logical Form* (LF) of the utterance. The logical form will be discussed shortly. The second translation represents the interpretation of the utterance as the performance of an *action* by the speaker, who, in doing that action, intends to achieve certain effects. I call this interpretation the *conversational event* associated with the utterance. The conversational event generated by an utterance is obtained from the logical form by means of *conversational event generation rules*, although it would be possible to use a system closer to that of Karttunen and Peters. I discuss discourse interpretation and conversational event generation rules in more detail in §6, after having presented my hypotheses about the organization of information in our conversations and the role of conversational events.

3.5 The Logical Form

The logical form is generated by a GPSG grammar which uses $\mathcal{SEL}^{\mathcal{D}}$ as its target language, but is otherwise very similar to the grammars discussed in (Hwang and Schubert 1993, Hwang 1992), and therefore is not presented in this paper. The logical form for the imperative sentence `29.5` is the $\mathcal{SEL}^{\mathcal{D}}$ expression in (9).

---

[10] One of the assumptions of the work on TRAINS is that linguistic items other than complete sentences—for example, cue phrases—may also result in changes in the state of the discourse. For this reason, I use the term conversational event to indicate those units of a discourse which result in a belief update. For the purposes of this paper, 'conversational event' is synonymous with 'locutionary act.'

29.5 U: pick up the boxcar

(9) (IMPER
    (λ x
        (THE y [$\dot{S}$ $\models$ BOXCAR(y)] ∧ SHARED(**spkr**,**hearer**,$\dot{S}$))
        (PICKUP(y))(x))))

Imperative utterances are translated as tenseless VPs (denoting predicates over individuals) in the scope of a IMPER operator expressing the *sentential force* of the sentence (Chierchia and McConnell-Ginet 1990, p.164). The difficult problem of specifying truth conditions for all types of utterances is side-stepped by assuming that the truth conditions of an imperative sentence such as 29.5 can be captured by means of the IMPER operator. (Analogous operators for questions (QUES) and declarative utterances (DECL) are assumed to exist.) While this hypothesis is not implausible, and similar proposals have in fact been made in the past (Chierchia and McConnell-Ginet 1990, p.164–166), this approach leaves us with the task of specifying the denotation of an expression like (9). Some of the literature on the semantics of utterances other than declaratives (especially on the semantics of questions) is reviewed by Chierchia and McConnell-Ginet (1990). The strategy that we, as well as others, are pursuing, is to assign an utterance a value which reflects its potential for context change (Stalnaker 1979, Heim 1982, Groenendijk and Stokhof 1991). Different systems have been proposed; the possibility most closely related to the ideas about the common ground presented in §4 is to have utterances denote functions from sets of situations to sets of sets of situations (i.e., relations between sets of situations).

The logical form in (9) only represents a partial specification of the truth conditions of 29.5, since at this point in the interpretation the resource situation of the boxcar has not yet been determined. This is indicated by the presence of an unanchored parameter $\dot{S}$. The interpretation of 29.5 is going to be completed by processes of inference on the mental state, like those discussed in §6.2. The rest of the paper focuses on this aspect of interpretation.

4 The Information Used to Interpret Definite Descriptions and its Structure

The information used to interpret definite descriptions is part of what Stalnaker (1979) and Heim (1982) called *common ground*, which is "…the participants' mutually developed public view of what they are talking about" (Chierchia and McConnell-Ginet 1990, p. 166). In our conversations, this information includes:

1. Facts 'about the world:' in our case, information obtained from the map.

2. Generic information about the task, such as expectations about the intentions of each conversational participant, and information about the 'rules of the game' —e.g., causal information: after unloading a boxcar, that boxcar becomes empty.
3. What has been said (the 'discourse history').
4. The current status of the plan.

According to Grosz and Sidner (1986), the common ground is best seen as divided in three parts: information about the linguistic structure of the utterances in the dialog (the *Linguistic Structure*), information about the goals of the participants in the conversation (the *Intentional Structure*), and information about the objects introduced in the discourse and their relative saliency (the *Attentional State*). I am concerned here with that part of the common ground that Grosz and Sidner call attentional state.

The data concerning the use of referential expressions in our dialogues presented in §2 provide good evidence that the participants in our conversations assume that the information contained in the attentional state is 'carved' into 'chunks' of information. Each apparent violation of the uniqueness requirements for definites constitutes evidence for the existence of one such chunk; discourse segmentation is perhaps the best known example of this phenomenon at work. Moreover, because the chunks formed in this way are used as resource situations for definite descriptions, the participants must assume that these 'chunk formation principles' are mutually known.

Because situations are but chunks of information 'kept together' by some coherence factor, my proposal that the chunks formed by the conversational participants are situations should not come as a surprise. Each of these situations is predicted to be shared—in the sense discussed in §3.3—by Clark and Marshall's copresence heuristics. It also appears that some of these situations are defined *intensionally*—i.e., the agents are aware of their existence even though they may not know all of the information that these situations contain. An intensionally defined situation is characterized by a *situation forming principle* which states under which conditions a conversational participant will assume that a piece of information is part of that situation. These situation forming principles are also mutually known to the participants in our conversations.

In order to model the way the common ground is partitioned in situations, we need to allow not just for situations which are chunks of information about the actual world, but also for situations which consist of information about other possible worlds—for example, worlds in which the events which are part of the plan actually occurred.[11] I'll call the latter *possible*

[11]Note that Grosz and Sidner are implicitly assuming this when they propose that all the objects and events introduced in the common ground are pushed onto a 'focus space stack' used to interpret referring expressions.

situations. The need for possible situations is proved by the cases, seen in §2, of anaphoric reference to elements of the plan not related to any specific object in the map. The crucial property of possible situations is that truth at a possible situation `s` does not result in truth in the situation of evaluation even for persistent facts:

$$[\mathtt{s} \models \Phi] \not\models \Phi$$

The attentional state, then, consists of situations which may be classified along two different axes—actual vs. possible, intensional vs. extensional. In the rest of this section I will describe these situations in turn.[12]

In order to avoid confusion, I have adopted Devlin's term *Discourse Situation* to indicate that part of the common ground which consists of a record of what has been said (1991, p. 218).[13] Even though the discourse situation is one of the situations which are part of the common ground, I'll postpone talking about its organization until after introducing my formalization of conversational events.

4.1 The Information from the Map

One of the situations into which the common ground is partitioned consists of the information about the map. I call this situation *Map Situation* (`MapS`). `MapS` is used as a resource situation for definite descriptions like "the boxcar at Elmira" (`13.3`) or "the tanker car" (`13.8`), that are interpreted with respect to the 'visible situation' (which, in our case, is the world represented on the map). The information in `MapS` represents the 'visual field' of the agents.

`MapS` is an 'actual' situation (at least, it is interpreted as such), but it is defined intensionally. An agent may refer to objects contained in it without knowing whether the other agent is aware of the existence of that object; we have, for example, exchanges like:

(10) ...
1 A: You see the boxcar at Dansville?
2 B: wait ... got it.

The situation-forming principle for this situation can be informally characterized as follows:

Situation Forming Principle 1 *A piece of information Φ is part of* `MapS` *iff the source of information for that piece of information is the map that the conversational participants are looking at.*

The conversational participants mutually know this principle, and also that `MapS` is shared. (This is predicted by Clark and Marshall's 'physical copre-

[12]For simplicity, I will only be concerned here with those situations which appear to be part of the common ground during the whole conversation. From our conversations, it appears that the participants are also able to create additional, shared 'possible situations' dynamically. At the moment, it's not clear to me how this process works.
[13]This terminology originated with (Barwise and Perry 1983).

sence' heuristic.) I assume, that is, that (12) and the axiom schema (13) are part of the *a priori* knowledge about the task.[14]

(12) SHARED(**self**,user,MapS)

(13) MK(**self**,user,(SOURCE(map,Φ) $\equiv$ [MapS $\models$ Φ]))

4.2 Focus of Attention and Visual Attention

I discussed in §2 the relation between focus of attention and 'visible situation' use of definite descriptions first observed by Grosz: when an object is in the current *mutual* focus of attention, it can be felicitously referred to by means of a definite description even though other objects of the same type have been introduced in the discourse or are part of **MapS**. I propose that the 'focus of attention' studied in the literature on definite description interpretation is the object of *visual attention* (Allport 1987).

I assume, with the Situation Theory literature, that of all the objects in the visual field, only those within the current range of visual attention are actually 'seen,'[15] and that the object of visual attention is a situation; I call this situation *situation of attention.*[16] I assume a predicate SEE analogous to the one used by Devlin (ch.7) to talk about a relation between an agent and a situation the agent is actively 'looking at.' Finally, I introduce a relation MSEE between pairs of agents and situations to model the notion of current *mutual* situation of attention. Two agents a and b mutually see a situation s, written MSEE(a,b,s), if they mutually know that both of them see the situation. I assume that each situation of attention is shared.[17]

In our dialogues, the situation of attention is always a sub-situation of **MapS**. The participants in our conversations do not, however, group the information from the map on the basis of some random order of selection; the sub-situations used as situations of attention always consist of the information about a town in the map at a certain point in time. Presumably,

[14]Again, this is a simplification. Some speakers appear to assume the existence of *two* such situations. These speakers appear to assume the existence of a 'shared map' situation in addition to **MapS**; this situation consists of those facts about the map that have been explicitly mentioned in the conversation. Only this 'shared map' is actually shared. Evidence for this is the abundance in our dialogues of utterances of the form

(11) U: There is a boxcar at Avon.

whose purpose, at least in part, seems to be to make sure that certain information is actually shared. (These acts may also be interpreted as indirect requests for identification (Cohen 1984) and indirect ways to impose constraints on the plan (Traum 1993.))

[15]The distinction between 'seeing' and 'seeing that' has been repeatedly discussed in the Situation Semantics literature (Barwise and Perry 1983, Devlin 1991).

[16]The situation of attention plays a role similar to Barwise and Perry's 'object we are attending to' (1983, pag. 87). Assuming that the focus of attention is a situation leads to simpler axioms relating the current focus of attention to the resource situation and allows for more than one object to be in the current focus of attention.

[17]Note that in a logic of knowledge like S5 this would follow from the definition of 'mutual seeing,' the fact that each mutually seen situation is actual, and the veridicality axiom of Situation Theory.

this is because the conversational participants only refer to situations they may expect the other participant to be able to 'build' as well. One principle for 'building situations' that the participants clearly expect to be mutually known is the following:

Situation Forming Principle 2 *The collection of facts about a location at a certain point in time constitutes a situation.*

I use in the rest of the paper the function PLACE(p,s,t) to denote the situation characterized by the facts which are true at location p at time t in situation s.

4.3 The Information about the Plan

As seen in §2, our participants refer by means of definite descriptions to objects which have only been introduced in the plan. This observation can be reconciled with the location theory if we hypothesize that the participants in our conversations 'build' a situation out of all the objects and events introduced in the plan, and then use this situation as the resource situation when referring to objects only introduced in the plan.

The problem is how to reconcile this intuition with the hypotheses about the ontological status of plans currently prevailing in Artificial Intelligence. In the literature on planning, plans are usually seen as 'recipes' to perform kinds of actions, a recipe being a graph whose nodes represent 'operators' (action types) and whose arcs represent temporal or causal relations (Fikes and Nilsson 1971, Sacerdoti 1977). This view, however, is of limited use as a model of what people do when they talk about plans, as noted, e.g., by Pollack (Pollack 1990). Pollack argues that, in order to understand the way people talk about plans in conversations, we need to be able to talk about the mental attitudes the participants in our conversation have when they have a plan. She proposes the following definition for "having a plan" (the 'set of acts' Π of the definition is the 'recipe' used in standard planning formalisms):

Definition P0, **(Pollack 1990, p. 89)** An agent A has a plan to do β that consists in doing some set of acts Π (written PLAN(A,β,Π)) provided that

1. A believes that he can execute each act in Π.
2. A believes that executing the acts in Π will entail the performance of β.
3. A believes that each act in Π plays a role in his plan.
4. A intends to execute each act in Π.
5. A intends to execute Π as a way of doing β.
6. A intends each act in Π to play a role in his plan.

Pollack's observation that plans can be seen both as recipes and as mental attitudes (and that the two perspectives need to be separated) provides, I believe, a way to reconcile the facts about reference to objects in a plan with the view of plans common in the planning literature. I propose, first of all, to include 'recipes-for-action' among the objects of our domain; more precisely, I propose to treat them as *kinds of actions*. I use the predicate RECIPE to assert of an object that that object is a recipe-for-action.

Secondly, I propose that the participants to our dialogues use definite descriptions to refer to objects in the plan because they arrive at a specification of a recipe for their task by augmenting their shared characterization of an intensionally defined possible situation that I call *Plan Situation* (`PlanS`), and that they use this situation as a resource situation when they refer to an object that can only be identified by way of its role in the plan.

At every moment in the conversation, what the system and the user mutually know about `PlanS` defines a recipe, obtained by abstracting over each event and object in `PlanS`. Up to the moment in which both participants accept the plan, the events which are mutually known to be part of `PlanS` are only a partial specification of the plan. The recipe specified by `PlanS` becomes a plan for the task in the sense of Pollack only when condition 2 of the definition above is met.

The situation forming principle for `PlanS` (i.e., what makes the participants interpret the described event of an utterance as a part of it) can be obtained from a revision of Pollack's definition stating what it means for two agents to have a *shared* plan. Unfortunately, while there have been attempts at providing such a definition (see, e.g., (Grosz and Sidner 1990)), no completely satisfactory definition exists yet. I'll nevertheless assume that such a notion can be defined, and use MPLAN(A,B,β,Φ) to stand for 'Agents `A` and `B` have a mutual plan to do the kind of action β which consists of the recipe Φ.' The resulting situation forming principle can be tentatively stated as follows:

Situation Forming Principle 3 *An event described by an utterance is part of* `PlanS` *if the system and the user mutually believe that that event plays a role in the plan, mutually believe that the event can be executed, mutually intend to execute it as a way of performing the task required, and mutually intend it to play a role in the plan.*

Barwise and Perry introduced in (1983) the notion of *course of action* to characterize those situations whose defining characteristic is that they consist of a set of events ordered in a sequence and that 'go together' according to some 'forming principle.' What makes a sequence of events into a course of action may vary—it may be the agent's perception that they form a causal chain, or that some particular individual plays the agent's role in all of them, or some additional factor.[18] Because the events in a course of action form a sequence, we can define functions PRED`(e,coa)`

[18] The phenomenon of people forming 'stories' out of descriptions of events has been studied extensively in the work on understanding narratives (Nakhimovsky 1988,

and NEXT(e,coa) which return, for each event e in a course of action coa, the previous event and the next event in the sequence, respectively. Another feature of courses of action that I use below is that for any two successive events e_1, e_2 in a course of action, a function R(e_1,e_2) can be defined which returns the time interval between the culminations of the two events. I treat **PlanS** as a course of action.

Providing a completely satisfactory definition of what it means for a course of action to specify a recipe is rather tricky, and would require a separate paper. Here, I just introduce a one-to-one mapping RECIPE-OF-COA from courses of action to recipes, and assume that this mapping is defined on every course of action. We also want a course of action c to be an instance of the recipe RECIPE-OF-COA(c):

(14) (∀ r RECIPE(r)
        (∀ c COURSE-OF-ACTION(c)
           r = RECIPE-OF-COA(c) ⊃ INSTANCE-OF(c,r)))

The fact that **PlanS** is shared appears to be part of the common ground. This might be an application of Clark and Marshall's 'linguistic copresence' heuristic, since everything which is part of **PlanS** is introduced in the discourse. I also propose that part of the common ground is the fact that the task of the user and the system is to develop a recipe for the kind of action they have to accomplish.

4.4 Described Events and Common Ground Situations

According to the view of the organization of the common ground proposed in this section, part of the task of an agent interpreting an utterance is to identify that situation, among those contained in the common ground, of which the event described by the utterance (the *Described Event* of Barwise and Perry) is a part. I have also proposed that the information used by an agent in this process includes at least the situation-forming principles assumed to be mutually known. Once the agent has formulated an hypothesis about the intentions of the speaker, he may use the situation-forming principles to incorporate the described event in one of the situations, and verify the correctness of the hypothesis.

Discussing the process of identification of this situation would bring us too far afield; I'll just introduce here a representation for events which is consistent with this perspective. This representation will be used in later sections of the paper as well.

In Episodic Logic, events and situations are 'of the same type' in the sense that the $*$ operator ($\models$ in $\mathcal{SEL}^{\mathcal{D}}$) can be used to characterize both. The translation of "John left" in $\mathcal{SEL}^{\mathcal{D}}$ is shown in (16). This expression

Webber 1988, Hwang and Schubert 1992, Kameyama et al. 1993) and it is usually accepted that intensional factors are involved.

reads that e is an event of John leaving which takes place at location l which precedes the current moment in time **now**.

(15) John left.

(16) (∃ e AT-ABOUT(e,l) ∧ BEFORE(l,**now**)
[e ⊨ LEAVE(john)])

That events have temporal *locations* is described in Episodic Logic using the predicate AT-ABOUT: AT-ABOUT(e,l) reads "event e has temporal location l." (For simplicity, I consider only temporal locations here.) The set of logical operators of Episodic Logic that I will use also includes the kind-forming operator K (Hwang and Schubert 1993), a set of temporal predicates which includes BEFORE, a causal predicate CAUSE, and a SUBEPISODE-OF relation between situations analogous to the part-of $e_1 \subset e_2$ relation used by Barwise and Perry (1983).

It has been proposed in the literature on narrative understanding (Webber 1988, Kameyama et al. 1993) that one of the tasks of a reader is to identify for each described event the course of action that that event is a part of. I follow here (Kameyama et al. 1993), who propose to implement this is by including in the specification of each event a statement of the form SUBEPISODE-OF(e,ṡ), asserting that e is part of the situation denoted by the parameter ṡ. In the case of "John left,", for example, we get the translation in (17):

(17) (∃ e AT-ABOUT(e,l) ∧ BEFORE(l,**now**) ∧ SUBEPISODE-OF(e,ṡ)
[e ⊨ LEAVE(john)])

5 Conversational Events

In this section, I discuss in more detail my proposal that a *conversational event* be associated with each utterance, and how the dynamics of the common ground gets characterized.

5.1 Action-based Models of Mental State Update

The aim of the work on intention recognition of Allen, Cohen, Levesque, Perrault and others (Cohen and Perrault 1979, Allen and Perrault 1980, Grosz and Sidner 1986, Cohen and Levesque 1990, Perrault 1990) is to model the process by which the addressee comes to recognize the speaker's intentions in uttering a sentence. As the literal intention of an utterance can be rather different from the actual intention, recognizing these intentions may require complex reasoning. The representations of belief and action used in this literature have been developed to model this reasoning.

The unifying characteristic of these models is the assumption that by uttering a sentence, a speaker is performing a *speech act* (Austin 1962, Searle 1969). A declared goal of some recent research, most conspicuously the work of Cohen and Levesque (1990), is to derive the properties of speech acts observed in the earlier literature from general properties of actions, instead of stipulating the properties of illocutionary acts. Hence, part of the task of those engaged in this line of research is to develop a model of actions from which the general properties of conversations may be derived. For the purposes of this paper, the relevant aspect of the models of action proposed in this literature is that the occurrence of an event results in certain *states* holding at the time after the event. This enables us to write axioms specifying the effect of the occurrence of a conversational event on the intentions and/or beliefs of the addressee.[19]

The model proposed by Perrault in (Perrault 1990) allows for a fairly intuitive formalization of the effects of a conversational event. Perrault proposes a *persistence theory of belief*, in which the effects of speech acts are formalized by default inference rules in the sense of Reiter (Reiter 1980), and in which it is assumed that old beliefs persist and that new ones are adopted as a result of observing external facts, provided that they do not conflict with old ones. The main characteristics of Perrault's theory are as follows. A weak S5 formalization of belief is assumed. The occurrence of a conversational event causes a state of the hearer believing that a certain event occurred; Perrault calls this the *Observability* axiom (Perrault 1990, p.172):

Observability :
$$\vdash \mathrm{DO}_{x,t}\alpha \;\&\; \mathrm{DO}_{y,t}\mathrm{Obs(x)} \supset \mathrm{B}_{y,t+1}\mathrm{DO}_{x,t}\alpha$$

In this formula, $\mathrm{DO}_{x,t}\alpha$ reads "x did α at time t," while $\mathrm{B}_{x,t}p$ reads "x believes at time t that p." The following two (non-defeasible) axioms formalize the fact that belief persists:

Memory : $\vdash \mathrm{B}_{x,t}p \supset \mathrm{B}_{x,t+1}\mathrm{B}_{x,t}p$
Persistence : $\vdash \mathrm{B}_{x,t+1}\mathrm{B}_{x,t}p \supset \mathrm{B}_{x,t+1}p$

Finally, Perrault provides two default inference rules. The first one formalizes the effect of declarative sentences: if agent x produces a declarative utterance, and it is consistent to assume that x believes that utterance, do so. According to the second default inference rule, if agent x believes that agent y believes p, then by default agent x gets to believe that p. $\alpha \Rightarrow \beta$ stands for $\frac{\alpha:\beta}{\beta}$ in Reiter's notation.

Declarative : $\mathrm{DO}_{x,t}(p.) \Rightarrow \mathrm{B}_{x,t}p$
Belief Transfer : $\mathrm{B}_{x,t}\mathrm{B}_{y,t}p \Rightarrow \mathrm{B}_{x,t}p$

[19]It is important to keep in mind that Cohen, Levesque and Perrault are aiming at developing a *theorist's* logic, that is, a characterization of the agent's mental states and how utterances affect them 'from the outside' (this apt terminology has been proposed by Devlin). Ultimately, an *agent*'s logic —that is, an account of what goes on in an agent's mind when hearing an utterance—would be most desirable for someone trying to understand the actual reasoning processes of people; in this paper, however, I am developing an 'externalist' proposal. An attempt to formalize an agent's logic is in progress.

This formalization of belief transfer and of the effect of declarative sentences has well-known problems, as does the choice of weak S5 as the basis for a formalization of belief. Discussing these issues, however, would bring me too far beyond the scope of this paper. I will instead concentrate on whether Perrault's proposal gives us the right tools for my current purposes.

Perrault's formalism allows for a simple characterization of pragmatic interpretation. The focus shift rules that I propose in §6.2, for example, are, roughly speaking, of the form of the following default inference rule, whose intended interpretation is: If x utters α, and some other conditions Φ occur, then the mutual focus of attention will shift to β. (The idea is elaborated in §6.2.)

(18) $\vdash \mathrm{DO}_{x,t}\alpha \ \& \ \Phi \Rightarrow \mathrm{B}_{y,t+1}\ \mathrm{MSEE}_{x,y,t+1}\ \beta$

Another advantage of Perrault's theory is that making the occurrence of the conversational event a part of the common ground allows me to provide a situation-theoretic formalization of Grosz and Sidner' treatment of *discourse segmentation*, as discussed in §5.3. Finally, and most importantly, Perrault's theory (as well as Cohen and Levesque's) has been developed to formalize one form of 'practical reasoning' —the process of intention recognition—which, as I have discussed in §2, is closely related to the process of definite descriptions interpretation. It seems therefore reasonable to use this kind of model as a starting point when trying to come out with a formalization of the 'practical reasoning' involved in definite description interpretation.

Some work is needed, however, to adapt Perrault's proposal to my needs. Neither Perrault nor Cohen and Levesque provide formal rules of translation similar to, say, the GPSG system or the DRT construction rules, and anyway the language used by Perrault is not really suited to serve as a translation language for natural language, which makes it difficult for someone else to extend the proposal with such rules. This problem can be fixed by reconstructing the main ideas of Perrault's action based model in Episodic Logic; the language of Episodic Logic has been designed to be the target of a formal process of translation. I'll do that in the next section.

5.2 Conversational Events in a Situational Theory

In order to recast Perrault's treatment in Episodic Logic, I need a belief relation BELIEVE(a,p) between an agent a and a proposition p (I will be deliberately vague about the axiomatization of belief), a second relation INTEND(a,p) representing intention relations, and a **self** constant to stand for the agent whose mental state is being modeled—always the system, in our case. I'll follow Perrault in using Default Logic to formalize nonmonotonic inferences, even though I'm aware of the problems with this approach; I discuss the main problems at the end of the paper. I assume therefore a proof theory like that of Episodic Logic, but extended with default inference rules of the sort used by Perrault; I also adopt Perrault's notation, namely, I write

$$\alpha_1, \ldots \alpha_n \Rightarrow \beta_1, \ldots \beta_n$$

to indicate a default inference rule with premises $\alpha_1, \ldots \alpha_n$ and conclusions $\beta_1, \ldots \beta_n$. The resulting language has been called $\mathcal{SEL}^{\mathcal{D}}$, for "Simple Episodic Logic with Defaults."

Perrault's observability axiom can be rephrased in $\mathcal{SEL}^{\mathcal{D}}$ as stating that a (conversational) event ce of **spkr** telling **hearer** that Φ results in **hearer** acquiring a belief b characterized by **hearer** BELIEVING that **spkr** told **hearer** that Φ. TELL is the *surface speech act* into which declaratives are translated. The resulting belief is described by the following expression:

```
(19) (∃ b AT-ABOUT(b,l)
        (∃ ce AT-ABOUT(ce,l') ∧ CAUSE(ce,b)
           [b ⊨
            BELIEVE(hearer,
             [ce ⊨  TELL(spkr,hearer,Φ)])]))
```

The case of imperatives is illustrated by utterance 29.5 in (2), whose logical form was presented in (9). The conversational event associated with that utterance is characterized by the expression in (20), to be read: a belief of the system's bs29.5 holds at temporal location **now**, of the system (**self**) believing that the user instructed the system to pick up the boxcar. INSTRUCT is the 'surface speech act' used to translate imperatives. The expression K(λ e ...) in (20) is an *event kind*; event kinds are used in Episodic Logic as the uniform translation of infinitives like "to load the boxcar with bananas." I discuss how (20) is obtained from (9) in §6.1. (This translation for 29.5 is further revised below.)

29.5 U: pick up the boxcar,

```
(20) (∃ bs29.5 AT-ABOUT(bs29.5,now)
        (∃ ce29.5 AT-ABOUT(ce29.5,l3) ∧
               CAUSE(ce29.5,bs29.5)
           [bs29.5 ⊨
             BELIEVE(self,
              [ce29.5 ⊨
                INSTRUCT(user,self,
                       (K λ le29.5
                             SUBEPISODE-OF(le29.5,ċ) ∧
                             [le29.5 ⊨
                              (THE y ([Ṡ ⊨ BOXCAR(y)]
                                     ∧ SHARED(spkr,hearer,Ṡ))
                                     PICKUP(self,y))]
                                     ))])]))
```

In a temporal-based approach to mental state representation like the one presented here, one needs to specify which attitudes held by an hearer prior to a conversational event persist in time. Perrault assumes that all beliefs, once acquired, persist forever; he achieves this by means of the memory and persistence axioms seen above. Those axioms could easily be reformulated in $\mathcal{SEL}^{\mathcal{D}}$; instead, I have adopted a formulation based on a proposal by Hans Kamp (1990). Kamp proposes that a 'current' mental attitude is one that holds at the indexical **now** point. This way of achieving persistence eliminates the need for performing persistence reasoning, and is therefore appealing both from a conceptual and from a practical point of view. (This method does share with Perrault's the problem that older beliefs never get to be disbelieved.)

I assume the following as far as the persistence of attitudes other than belief is concerned. There is a great deal of debate about the characteristics of visual attention (Allport 1987), but it seems safe to assume that, unlike belief, mutual attention persists for a very short time, if at all. I'll simplify the issue by assuming that a state `s` characterized by the agents `x` and `y` paying mutual attention to the situation `ms` only holds for the time between one conversational event and the next event in the discourse situation `ds`.[20] The common ground contains, in these cases, a fact of the form:

(21) (∃ b AT-ABOUT(b,**now**)
(∃ s AT-ABOUT(s,R(ce,NEXT(ce,ds)))
[bs ⊨
BELIEVE(x,[s ⊨ MSEE(x,y,ms)])]))

As far as intentions are concerned, I simply assume for the moment that they are different both from beliefs (which always persist) and from attention (which doesn't persist) in that they persist by default.

The processes whose goal is to recognize the speaker's meaning operate on the basis of the information provided by the conversational event and the information which is part of the common ground. Of these processes, I only discuss below those whose goal is to anchor the resource situation of definite descriptions. I ignore the process of intention recognition which is the main interest of Cohen and Levesque, Perrault, and others. For the sake of concreteness, I assume here that this process can be modeled by means of defeasible inference rules similar to those proposed by Perrault. In the case of `29.5`, for example, I assume that such rules would result in the system acquiring by default a belief that it is the intention of the user that the system bring about an event which is an instance of the event type which is the argument of INSTRUCT in (20). (The $\mathcal{SEL}^{\mathcal{D}}$ expression describing this had to be omitted for space reasons, but the reader should be able to reconstruct it.)

[20] I.e., the next conversational event in absolute terms, not the next conversational event in the current course of action.

Some of the rules of interpretation presented in §6.2 make reference to the described event of an utterance; however, the translations proposed above make the described event a part of the linguistic meaning of the utterance only in the case of declarative sentences and questions—we can talk about the described event of imperatives only by referring to the intentions of the conversational participants. In order to simplify matters, I introduce a function DESCRIBED-EVENT-OF(ce) from conversational events to their described event, defined with respect to the intentions of the participants; I use this function in §6.2.

5.3 Conversational Threads and Discourse Segmentation

Grosz and Sidner account for the interaction between discourse segmentation and anaphoric uses of definite descriptions discussed in §2 by assuming that discourse segmentation is "parasitic upon the intentional structure" (1986, p.180). Whether a hearer interprets an utterance as being part of a particular discourse segment depends on whether the intention(s) expressed by that utterance (the *discourse purpose*) are related to the intentions expressed by the discourse segment.[21] Grosz and Sidner propose that intentions may be related in two different ways: when the discourse purpose is part of the satisfaction of another discourse purpose, the second purpose is said to *dominate* the first; if, instead, satisfying one intention is a prerequisite for satisfying a second one, the first intention is said to *satisfaction-precede* the second intention.

The effects of discourse segmentation on anaphoric accessibility are formalized by Grosz and Sidner by means of an abstract data structure called that *focus space stack*. As long as an utterance is part of the current discourse segment, the discourse referents evoked by that utterance are added to the 'focus space' on top of the stack, and the discourse referents already there are accessible for anaphoric reference. When an utterance introduces a discourse segment subordinate to the current one, a new focus space is pushed onto the stack. When an utterance completes the current discourse segment, the current focus space is popped from the stack.

I believe Grosz and Sidner's theory to be largely correct; I also believe, however, that much can gained by reformulating this account using the tools I have been introducing in the paper, one of the advantages being a better understanding of how this proposal fits into the larger picture of discourse interpretation. First of all, I propose that the hierarchical structure of discourse is reflected in the organization of conversational events, in the sense that, just as all other events, conversational events are arranged into courses of actions, that I call *conversational threads*. Whether a conversational event is perceived as part of a conversational thread depends

[21] Actually, Grosz and Sidner make a distinction between an utterance's intention(s) and the discourse segment purpose it carries. In simple cases like those discussed in this paper this distinction can be ignored.

on how the intentions expressed by that event are related to the intentions expressed by that thread. This gives us the following situation-forming principle which determines whether a conversational event is a member of a certain conversational thread:

Situation Forming Principle 4 *The hearer H achieves the belief that the speaker S intends a conversational event to be a subepisode of a conversational thread iff H achieves the belief that the intention expressed by S with that conversational event is subordinate to the intention associated with that conversational thread.*

Secondly, I assume that just as a described event is associated with each conversational event (see §4.4), a situation that I call *topic* is associated with each conversational thread. The events described by the conversational events in a thread are sub-situations of the topic of that thread. I propose that topics and their hierarchical organization do the work done in Grosz and Sidner's theory by the focus space stack.

Because topics are situations, they afford a natural reformulation of the focus space stack model. It's easy to see that each operation on the focus space stack can be reformulated as an operation on situations: 'adding to a focus space' corresponds to 'adding new constituents,' 'pushing' corresponds to 'create a new situation which informationally includes the previous one,' and 'popping' corresponds to 'selecting a situation informationally included in the previous one.' For each NP of the current utterance which introduces an anaphoric antecedent (and hence would result in the addition of an object to the focus space) a new constituent is added to the current topic. An utterance that opens a sub-segment results in a new conversational thread being open, whose topic informationally subsumes the topic of the previous conversational thread. An utterance that closes a discourse segment and pops to a previously current discourse segment results in the corresponding topic becoming current.

Replacing the focus space stack with a structure of situations results in three advantages: two of them are conceptual, one is related to the implementation. First of all, it is much simpler to understand the connection between this proposal and Grosz's earlier work on implicit focus (Grosz 1977). Second, we have a principled way to deal with the problem of choosing the discourse segment that the referent of a NP is part of in case nested segments are present (choosing the 'top of the stack' often results in undesirable pops). Finally, we get a quick way to check whether a conversational event is part of a conversational thread—namely, check whether the event described by that conversational event can be plausibly assumed to be part of the topic of the thread. Of course, if this test fails, a more complex process of intention recognition has to take place.

The system for introducing referents into discourse segments I assume here is inspired by Groenendijk and Stokhof's proposal in (1990).

The theory presented there is an example of so-called *dynamic* logic for language interpretation (Kamp 1981, Heim 1982, Barwise 1987, Groenendijk and Stokhof 1991, Groenendijk and Stokhof 1990). The defining characteristic of these systems is that the semantic translation of a sentence encodes its 'context change potential,' that is, it specifies which 'referents' are to be introduced into the discourse as the effect of the interpretation of that sentence. Each type of dynamic logic achieves this result in a different way. In Kamp's version of DRT, the context is a set of 'structures' called DRSs, each of which consists of a set of *discourse referents* and a set of *conditions*; a new discourse referent is added to the context by the DRS *construction rules* which interpret those classes of NP's which introduce pronominal antecedents (indefinites and definites). In the systems closer to Montague Grammar, like Episodic Logic and the system of Groenendijk and Stokhof, the context is a set of objects, and sentences map contexts into larger sets of objects.

Groenendijk and Stokhof call their contexts 'states,' and introduce a *state-switching operator* $\{\alpha/\mathrm{d}\}\beta$ whose semantics is specified as follows: "...The interpretation of $\{\alpha/\mathrm{d}\}\beta$ with respect to a state s is arrived at by interpreting β with respect to a state s' which differs at most from s in that the denotation of the discourse marker d in s' is the object that is the denotation of the expression α in s (p.7)." I propose to have topics do the work of Groenendijk and Stokhof's states, and to have discourse markers as constituents of topics. The translation of indefinite and definite NPs includes *topic update statements*, that are statements of the form $\{\alpha/\mathrm{d}\}\beta$ whose syntax is borrowed from Groenendijk and Stokhof's state-switching operator (e.g., α is the variable introduced in the translation of the NP). In addition, the translation of indefinite and definite NPs includes a statement to the effect that the discourse marker is part of the same topic as the event described by the conversational event; this is represented by means of statements of the form $\mathtt{d} \in \dot{\mathtt{c}}$.

Returning to our example 29.5, the conversational event that I propose to associate to 29.5 in order to incorporate the proposal that *each* event—whether 'described' or 'conversational'—is a subepisode of a course of events, and to take into account conversational update, is described by the $\mathcal{SEL}^{\mathcal{D}}$ expression in (22). $\dot{\mathtt{c}}'$ in (22) is a parameter to be resolved to a particular conversational thread by the process of intention recognition. The definite "the boxcar" 'updates' the topic $\dot{\mathtt{c}}$ that the described event is part of with a new discourse marker d29.5. (Compare (22) with the representation presented in §4.4.)

(22) (∃ bs29.5 AT-ABOUT(bs29.5,**now**)
      (∃ ce29.5 AT-ABOUT(ce29.5,$\mathtt{l}_3$) ∧ CAUSE(ce29.5,bs29.5)
            ∧ SUBEPISODE-OF(ce29.5,$\dot{\mathtt{c}}'$)
        [bs29.5 ⊨

```
BELIEVE(self,
 [ce29.5 ⊨
   INSTRUCT(user,self,
   (K λ le29.5
        SUBEPISODE-OF(le29.5,ċ) ∧
        [le29.5 ⊨
         (THE y ([Ṡ ⊨ BOXCAR(y)] ∧  d29.5 ∈ ċ'
                ∧ SHARED(spkr, hearer,Ṡ))
              {d29.5/y}(PICKUP(self,y)))]
              ))])]))
```

Grosz and Sidner imply that their focus space stack is a formalization of all the relevant aspects of the attentional state. They acknowledge, however, the need for additional mechanisms such as centering for the purpose of modeling pronoun resolution (p.191) and do not provide a specific proposal concerning the integration of the two mechanisms. I believe that maintaining a distinction between the structures used to interpret anaphoric and visible situation uses of definite descriptions leads to a clearer theory.[22] Topics provide the segmentation mechanism necessary for dealing with anaphoric uses of definite descriptions, that are clearly separated from the tools necessary for dealing with the visible situation uses introduced in §4.

6 Discourse Interpretation

As said in §3, I propose to conceptualize the process of pragmatic interpretation of an utterance as consisting of three parts. First of all, syntactic and semantic interpretation take place, and determine the truth conditions assigned to the sentence (which I called the *logical form* (LF) in §3). Secondly, the utterance is associated with a *conversational event*: the conversational event associated with an utterance is described by expressions like (20) discussed in §5. I propose to characterize the process that determines the conversational event by means of *conversational event generation rules*; some of these rules are briefly discussed below. Finally, there are processes of pragmatic interpretation that determine the speaker's meaning for the utterance from the conversational event and the current *mental state* of the agent. The result is a modified mental state. I propose to characterize this last part of discourse interpretation by means of *Mental State Change Axioms* (MSCA); Grosz and Sidner's 'Transition Rules' fall in this category.

According to this view, a theory about the process that determines the speaker's meaning for an utterance consists of a set of mental state change axioms. In this section, I provide a set of mental state change axioms which formalize the process of interpreting 'visible situation' de-

[22] Some recent work on attention also points out the need for a separation between the attentional components involved with different senses (Allport 1987).

finite descriptions—axioms that generate hypotheses about the anchor of the resource situation, together with axioms which formalize the process of attention shift.

6.1 Conversational Event Generation Rules

The input to the conversational event generation rules consists of a pair ⟨logical form,context⟩, where the context is a triple ⟨speaker, addressee, temporal location⟩. Expressions like (22) in §5 are the output of these rules. The rules for declaratives and imperatives are as follows:[23]

Imperatives

```
ce:⟨(IMPER (λ x Φ(x))),⟨spkr,hearer,t⟩⟩
   ⇝
(∃ bs AT-ABOUT(bs,now)
   (∃ ce AT-ABOUT(ce,l′) ∧ t = l′ ∧
         SUBEPISODE-OF(ce,ċ) ∧ CAUSE(ce,bs)
      [bs ⊨
       BELIEVE(hearer,
       [ce ⊨
         INSTRUCT(spkr,hearer,
          K(λ  e SUBEPISODE-OF(ce,ċ′) ∧
                [e ⊨  (λ x Φ(x))(hearer)]))]))])
```

Declaratives

```
ce:⟨(DECL (∃ e Φ(e) [e ⊨ Ψ])),⟨spkr,hearer,t⟩⟩
   ⇝
(∃ bs AT-ABOUT(bs,now)
   (∃ ce AT-ABOUT(ce,l′) ∧ t = l′ ∧
         SUBEPISODE-OF(ce,ċ) ∧  CAUSE(ce,bs)
      [bs ⊨
       BELIEVE(H,
       [ce ⊨
         TELL(spkr,hearer, (∃ e Φ(e) [e ⊨  Ψ]))]))])
```

6.2 Anchoring Resource Situations: the Interpretation of 'Visible Situation' Definite Descriptions

As an example of the kind of theory that the system discussed in the previous sections can be used to formalize, I'll present my current hypothesis about the process by which the 'visible situation' definite "the boxcar" in sentence `29.5` of the transcript (2) gets assigned its interpretation.

Two kinds of principles are involved in the interpretation of definite descriptions interpreted with respect to the visible situation. First of all,

[23] The system proposed by Hwang and Schubert includes a set of *deindexing rules*, some of which do the work of the conversational event generation rules proposed here. The outputs are similar, but not identical.

there are defeasible principles formulating hypotheses about ways for anchoring resource situations: I call these *principles for anchoring resource situations*. Second, there are principles governing visual attention shifts.

At least two principles for anchoring resource situations are at play in our dialogues: one predicting that when there is a situation of mutual visual attention, that situation may be used as the resource situation for definite descriptions; and one hypothesizing that the resource situation may be identified via the current topic. The first principle accounts for the visible situation uses of definite descriptions, the second for the anaphoric uses. The principle for anchoring resource situations of interest here, **PARS1**, says that if a speaker uses a referring expression "the P" as part of the description of an event e, and the speaker intends the mutual attention of the conversational participants to be focused on the situation s, then infer that s is the resource situation for "the P" if it is consistent to do so. This is formalized by the following defeasible axiom schema, where I use the function DESCRIBED-EVENT-OF introduced in §4.4:[24]

(PARS1)

$$
\begin{array}{l}
[\mathtt{bs}_1 \models \text{BELIEVE}(\mathtt{y},\ (\text{DESCRIBED-EVENT-OF}(\mathtt{ce}) = \mathtt{e}\ \wedge \\
\qquad\qquad [\mathtt{e} \models (\text{THE}\ \mathtt{z}\ [\dot{\mathtt{S}} \models \text{P}(\mathtt{z})]\ \text{Q}(\mathtt{z})))] \\
\qquad\qquad \wedge\ \text{AT-ABOUT}(\mathtt{ce},\mathtt{t}))] \\
\wedge\ \text{AT-ABOUT}(\mathtt{bs}_1,\mathbf{now})\ \wedge\ \mathtt{t} = \mathbf{now}\ \wedge \\
[\mathtt{bs}_2 \models \\
\ \text{BELIEVE}(\mathtt{y},\ \text{INTEND}(\mathtt{x},\ ([\mathtt{fs} \models \text{MSEE}(\mathtt{x},\mathtt{y},\mathtt{s})]\ \wedge \\
\qquad\qquad \text{AT-ABOUT}(\mathtt{fs},\text{R}(\text{PRED}(\mathtt{ce},\mathtt{coa}),\mathtt{ce})))))] \\
\wedge\ \text{AT-ABOUT}(\mathtt{bs}_2,\mathbf{now}) \\
\Rightarrow \\
\quad (\exists\ \mathtt{bs}_3\ \text{AT-ABOUT}(\mathtt{bs}_3,\mathbf{now}) \\
\qquad [\mathtt{bs}_3 \models \text{BELIEVE}(\mathtt{y},\text{INTEND}(\mathtt{x},\text{ANCHOR}(\dot{\mathtt{S}},\mathtt{s})))])
\end{array}
$$

As discussed in §2, the visible situation use of definite descriptions is affected by the current focus of attention: when an object is in the current focus of attention, it can be felicitously referred to by means of a definite description even when other objects of the same type have been introduced in the discourse or are part of the world described by the map. An agent's visual focus of attention changes continuosly (Allport 1987), yet not all of these shifts can be exploited to make the use of a definite reference felicitous: conditions on mutual knowledge have to be met (Clark and Marshall 1981). Thus, an attention shift can only be exploited when the participants in the conversation mutually know that the shift took place, on the grounds of some general fact about the conversation. In the case of the conversations studied by Grosz, the movement of the focus of attention was related to the

[24] All the unbound variables are to be taken as universally quantified.

structure of the task. In our transcripts, the most important conversational principle governing visual attention shifts appears to be the following:

Follow The Movement : Part of the intended effect of an utterance instructing an agent to move an object from one location to another is to make the terminal location of the movement the new mutual situation of attention.[25]

This principle can be formalized by the following default inference rule:

```
(23) [bs |=
      BELIEVE(y, [[ce |= INSTRUCT(x,y,
                                   (K λ e [e |= MOVE(y,z,p)]))]
                  ∧ AT-ABOUT(ce,t1)])]
     ∧ AT-ABOUT(bs,now)
     ⇒
       (∃ bs1 AT-ABOUT(bs1,now)
          (∃ fs1 AT-ABOUT(fs1,R(ce,NEXT(ce,ds)))
             [bs1 |=
               BELIEVE(y,
                INTEND(x,
                 [fs1 |=  MSEE(x,y,PLACE(p,MapS))]))]))
```

in (23), PLACE(p,MapS) is the 'place situation' consisting of the facts and the objects in MapS which describe the position p (see §4.2).[26]

Let us see now how these principles determine the interpretation of utterances 29.4 -29.5 in (2). The conversational event generation rules assign to 29.4 the translation in (24).

[25] This principle could perhaps be derived from a more general principle saying that in our dialogues it is expected of each conversational participant that he pays attention to what the other participant says, and 'paying attention' when the movement of an object in the map is concerned simply consists of following the movement of that object on the map. This, however, is not very important for my purposes.

[26] This is a simplification— actually, one should 'carve out' the place situation not from MapS, but from the situation 'resulting' from action of moving z to p. This is how the system actually works. Also, it's not clear whether a more general formulation in terms of intentions rather than the surface act would still hold.

29.4 U: take engine E1 to Dansville,

```
(24) (∃ bs29.4 AT-ABOUT(bs29.4,now)
       (∃ ce29.4 AT-ABOUT(ce29.4,l′) ∧
                 SUBEPISODE-OF(ce29.4,ċ1) ∧
                 CAUSE(ce29.4,bs29.4)
          [bs29.4 ⊨
            BELIEVE(self,
            [ce29.4 ⊨
              INSTRUCT(user,self,
              (K λ le29.4
                 SUBEPISODE-OF(le29.4,ċ2) ∧
                 [le29.4 ⊨
                  (TO(Dansville)
                   (TAKE(E1)))(self)]))])]))
```

The 'follow the movement' principle now applies (assuming a simple inference to the effect that each 'taking' event in our domain entails a move), with the effect that the system infers that the user intends the new situation of attention to be the 'place situation' consisting of the facts about Dansville—i.e., the following is hypothesized:

```
(∃ bs29.4b AT-ABOUT(bs29.4b,now)
   (∃ fs29.4b AT-ABOUT(fs29.4b, R(ce29.4,NEXT(ce29.4,ds)))
      [bs29.4b ⊨
       BELIEVE(self,
        INTEND(user,
         [fs29.4b ⊨
          MSEE(user,self,PLACE(dansville,MapS))]))]))
```

The next utterance, `29.5`, is interpreted in the common ground augmented by `bs29.4b`. I discussed the conversational event associated with `29.5` in detail in §5; the final characterization was described in (22). The principle for anchoring resource situations mentioned before, **PARS1**, now applies, with the result that the system hypothesizes that the user intends the resource situation for the definite "the boxcar" in `29.5` to be PLACE(`dansville`,`MapS`):

```
(25) (∃ bs4 AT-ABOUT(bs4,now)
       [bs4 ⊨
        BELIEVE(self,
                INTEND(user,Ṡ = PLACE(dansville,MapS)))])
```

Because there is only one boxcar in PLACE(`dansville`,`MapS`), `b1`, it is consistent for the system to infer (25). This entails (26).

```
(26) (∃ bs5 AT-ABOUT(bs5,now)
       [bs5 ⊨ BELIEVE(self,INTEND(user, PICKUP(self,b1)))])
```

7 Discussion

7.1 Formal Issues

My main reason for choosing Default Logic to formalize defeasible inference is that its properties are relatively well-known. It is also well-known, however, that this logic is not completely appropriate for the job of formalizing pragmatic reasoning. The main problem is that if conflicting defaults apply, nothing can be concluded; in fact, the very strong form of belief persistence adopted by Perrault and here is technically necessary to avoid conflicts between new and old beliefs. These characteristics make it difficult to deal with cases in which two principles for anchoring resource situations apply, one suggesting an anaphoric interpretation of a definite description, the other a 'visible situation' interpretation, and the hypotheses that they suggest are in conflict. There are several ways out of the problem. One way is to adopt a logic in which preferences can be expressed, e.g., Shoham's logic of preferred models (Shoham 1988). The properties of these logics are not, however, as well-known as those of Default Logic. A second alternative is to use a probabilistic logic in which beliefs are augmented with probabilities. Schubert and Hwang are working on a version of Episodic Logic that includes a probabilistic version of material implication. A third alternative is to adopt a more 'representational' approach to mental state representation that allows the formulation of (extra-logical) procedures for hypothesis comparison. I have done some preliminary work in this direction.

7.2 Implementation

These ideas are embodied in a system called SAD-92, a module of the TRAINS-92 system. The task of SAD-92 is to record the occurrence of a conversational event in the representation of the mental state, and to perform scope disambiguation and reference interpretation. The input to the system is a logical form of the kind discussed in §3 and §6, obtained by the module of TRAINS-92 first called on an input sentence, the (GPSG) parser. Conversational event generation rules are first recursively applied to the logical form, yielding the conversational event associated with the sentence. The occurrence of a conversational event is then recorded in the representation of the mental state; this causes the activation of the mental state change axioms 'triggered' by the occurrence of that kind of event. These axioms implement, in addition to the definite description interpretation procedures described in this paper, procedures for scope disambiguation (Poesio 1993) and intention recognition (Heeman 1993). The inference engine which computes the consequences of the mental state change axioms is, at the moment, a simple version of EPILOG, the inference engine for Episodic Logic. The result is an hypothesis about the intentions expressed by the user in uttering the sentence; the consistency of this hypothesis is verified by a module which attempts to relate this intention to the current plan (Ferguson 1992). If the hypothesis is found to be consistent, the system starts planning its response (Traum 1993).

8 Concluding Remarks

The first contribution of this paper is the proposal to formalize Hawkins' location theory of definite descriptions in situation-theoretic terms as follows: upon hearing a definite description, a participant in the conversation tries to 'anchor' its 'resource situation' to a situation he or she believes to be shared. I also proposed that whether a definite description is interpreted anaphorically or with respect to the currently visible situation depends on which one among several competing principles for anchoring resource situations is found to yield a consistent hypothesis, and I gave a formalization of these axioms as default rules like those proposed by Reiter.

Secondly, I hypothesized that the 'discourse structuring principles' presented in the literature are best seen as principles for organizing the information contained in the common ground in a set of *possible situations*. This hypothesis is supported by a number of facts about definite reference, and allows an immediate integration of the location theory with the literature on discourse structure, and especially with the model proposed by Grosz and Sidner. I adopted a finer-grained structure for the attentional state than the one proposed by Grosz and Sidner; in particular, I proposed to separate visual attention from the phenomenon of verbal attention, realized in discourse segmentation.

Last, but not least, I proposed that the changes to the common ground brought about by an utterance are caused by the hearer's believing that a *conversational event* occurred and by his/her attempt of placing this conversational event in one of the *conversational threads* which compose the discourse situation. The reasoning processes involved in definite description interpretation are integrated with the processes involved in intention recognition, as usually assumed in the literature. I presented an hypothesis about the mechanisms which determine shifts in visual attention in our dialogues that seems to account for a large part of the attentional phenomena related to the 'visual' situation.

Unfortunately, I don't have much to say at the moment about a number of aspects of definite description interpretation which, some may argue, are the really difficult ones to model. Among these, the processes involved in interpreting 'generic' descriptions such as "the engines" in "the engines can carry up to three boxcars;" the processes involved in the interpretation of infelicitous descriptions (i.e., descriptions which have no referent); or those involved in the interpretation of descriptions which refer rather vaguely, such as "the whole thing" in `17.1` in (2) or "the path" in "use the path through Bath to Avon." I believe, however, that one cannot deal with these problems without first gaining the kind of understanding of the process of interpretation of definite descriptions which one can acquire by spelling out its details in the way I have done it in this paper.

Acknowledgments

I wish to thank my advisor Len Schubert and James Allen, Breck Baldwin, George Ferguson, Peter Heeman, Janet Hitzeman, Megumi Kameyama, Marc Light, Rebecca Passonneau, Stuart Shapiro, David Traum, Marc Vilain, and William Woods for useful discussions. This work was supported by the Air Force-Rome Air Development Center Research contract no.F30602-91-C-0010.

References

Allen, J. F., and C. Perrault. 1980. Analyzing Intention in Utterances. *Artificial Intelligence* 15(3):143–178.

Allen, J. F., and L.K. Schubert. 1991. The TRAINS Project. TRAINS Technical Note 91-1, University of Rochester, Department of Computer Science.

Allport, A. 1987. Visual Attention. In *An Introduction to Cognitive Science*, ed. M. I. Posner, chapter 16, 631–682. The MIT Press.

Appelt, D., and K. Konolige. 1988. A Practical Nonmonotonic Theory of Reasoning about Speech Acts. In *Proc. ACL-88*. Buffalo.

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

Barwise, J. 1987. Noun Phrases, Generalized Quantifiers and Anaphora. In Gärdenfors (1987), 1–30.

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

Chierchia, G., and S. McConnell-Ginet. 1990. *Meaning and Grammar: An Introduction to Semantics*. The MIT Press.

Clark, H. H., and C. R. Marshall. 1981. Definite Reference and Mutual Knowledge. In *Elements of Discourse Understanding*. Cambridge University Press.

Cohen, P. R. 1984. The Pragmatics of Referring and the Modality of Communication. *Computational Linguistics* 10(2):97–146.

Cohen, P. R., and H. J. Levesque. 1990. Rational Interaction as the Basis for Communication. In *Intentions in Communication*, ed. P. Cohen, J.Morgan, and M. Pollack, chapter 12, 221–256. Morgan Kaufmann.

Cohen, P. R., and C. R. Perrault. 1979. Elements of a Plan Based Theory of Speech Acts. *Cognitive Science* 3(3):177–212.

Cooper, R. 1979. The Interpretation of Pronouns. In *Syntax and Semantics 10*, ed. F. Heny and H. Schnelle, 61–92. Academic Press.

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

Evans, G. 1980. Pronouns. *Linguistic Inquiry* 11(2):337–362.

Ferguson, G. M. 1992. Explicit Representation of Events, Actions, and Plans for Assumption-Based Plan Reasoning. TR 428, University of Rochester.

Fikes, Richard E., and Nils J. Nilsson. 1971. STRIPS: A New Approach to the Application of Theorem Proving to Problem Solving. *Artificial Intelligence* 2:189–208.

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

Gärdenfors, P. (ed.). 1987. *Generalized Quantifiers.* D. Reidel.

Grice, H. P. 1957. Meaning. *Philosophical Review* 66:377–388. also in D.A. Steinberg and L. A. Jakobovits (eds.), *Semantics*, Cambridge University Press, 1971.

Grice, H. P. 1969. Vacuous Names. In *Words and Objections*, ed. D. Davidson and J. Hintikka, 118–145. D. Reidel.

Grice, H. P. 1989. *Studies in the Way of Words.* Harvard University Press.

Groenendijk, J.A.G., and M.J.B. Stokhof. 1990. Dynamic Montague Grammar. In *Quantification and Anaphora I*, ed. J. Groenendijk, M. Stokhof, G. Chierchia, and P. Dekker. DYANA Deliverable R2.2.A.

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

Gross, D., J. Allen, and D. Traum. 1992. The TRAINS 91 Dialogues. TRAINS Technical Note 92-1, Computer Science Dept. University of Rochester. to appear.

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

Grosz, B. J., and C. L. Sidner. 1986. Attention, Intention, and the Structure of Discourse. *Computational Linguistics* 12(3):175–204.

Grosz, B. J., and C. L. Sidner. 1990. Plans for Discourse. In *Intentions in Communication*, ed. P. R. Cohen, J. Morgan, and M. E. Pollack, chapter 20, 417–444. The MIT Press.

Harman, G. 1977. Review of *Linguistic Behavior* by Jonathan Bennett. *Language* 53:417–424.

Hawkins, J. A. 1978. *Definiteness and Indefiniteness.* Croom Helm.

Heeman, Peter A. 1993. Speech Actions and Mental States in Task-Oriented Dialogues. In *Working Notes AAAI Spring Symposium on Reasoning about Mental States: Formal Theories & Applications.* Stanford, March.

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

Hobbs, J. R. 1979. Coherence and Coreference. *Cognitive Science* 3:67–90.

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

Hwang, C. H., and L. K. Schubert. 1992. Tense Trees as the "Fine Structure" of Discourse. In *Proc. ACL-92*, 232–240. Newark, DE, June.

Hwang, C. H., and L. K. Schubert. 1993. Episodic Logic: A Situational Logic for Natural Language Processing. This volume.

Kadmon, N. 1987. *On Unique and Non-Unique Reference and Asymmetric Quantification.* PhD thesis, University of Massachusetts at Amherst.

Kameyama, M., B. Passonneau, and M. Poesio. 1993. Temporal Centering. In *Proc. ACL-93.* Columbus, Ohio.

Kamp, H. 1981. A Theory of Truth and Semantic Representation. In *Formal Methods in the Study of Language*, ed. J. Groenendijk, T. Janssen, and M. Stokhof. Amsterdam: Mathematical Centre.

Kamp, H. 1990. Prolegomena to a Structural Account of Belief and Other Attitudes. In *Propositional Attitudes—The Role of Content in Logic, Language, and Mind*, ed. C. A. Anderson and J. Owens, chapter 2, 27–90. CSLI.

Karttunen, L., and S. Peters. 1979. Conventional Implicature. In *Syntax and Semantics 11: Presupposition*, 1–56. Academic Press.

Kripke, S. A. 1977. Speaker Reference and Semantic Reference. In *Contemporary Perspectives in the Philosophy of Language*, ed. P. A. French, T. E. Uehling, and H. K. Wettstein, 6–27. University of Minnesota Press.

Lewis, D. K. 1969. *Convention: A Philosophical Study*. Harvard University Press.

Linde, C. 1979. Focus of attention and the choice of pronouns in discourse. In *Syntax and Semantics 12*, ed. T. Givon. Academic Press.

Nakhimovsky, A. 1988. Aspect, Aspectual Class, and the Temporal Structure of Narratives. *Computational Linguistics* 14(2):29–43.

Neale, S. 1990. *Descriptions*. The MIT Press.

Perrault, C. R. 1990. An Application of Default Logic to Speech Act Theory. In *Intentions in Communication*, ed. P. R. Cohen, J. Morgan, and M. E. Pollack, chapter 9, 161–185. The MIT Press.

Poesio, M. 1992. Conversational Events and Discourse State Change: A Preliminary Report. In *Proc. KR-92*. Boston, MA, October.

Poesio, M. 1993. The Scope and Deindexing Module of TRAINS-92. Technical note, University of Rochester, Department of Computer Science. To appear.

Pollack, M. E. 1990. Plans as Complex Mental Attitudes. In *Intentions in Communication*, ed. P. R. Cohen, J. Morgan, and M. E. Pollack, chapter 5, 77–104. The MIT Press.

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

Reiter, R. 1980. A Logic for Default Reasoning. *Artificial Intelligence* 13(1–2):81–132.

Russell, B. 1905. On Denoting. *Mind* 14:479–493. Reprinted in *Logic and Knowledge*, ed. R. C. Marsh. George Allen and Unwin.

Sacerdoti, Earl D. 1977. *A Structure for Plans and Behavior*. Elsevier.

Searle, J. R. 1969. *Speech Acts*. Cambridge University Press.

Shoham, Y. 1988. *Reasoning About Change*. The MIT Press.

Stalnaker, R. 1979. Assertion. In *Syntax and Semantics*, ed. P. Cole, Vol. 9, 315–332. Academic Press.

Traum, David R. 1993. Mental State in the TRAINS-92 Dialogue Manager. In *Working Notes AAAI Spring Symposium on Reasoning about Mental States: Formal Theories & Applications*. Stanford, March.

Webber, B. L. 1988. Tense as Discourse Anaphor. *Computational Linguistics* 14(2):61–73.