Definite Descriptions and the Dynamics of Mental States

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

Abstract

I propose a situation-theoretic formalization of the organization of information in the common ground which makes it possible to formulate principles about (i) the way an interpretation for definite descriptions is chosen, and (ii) how the information used for interpreting definite descriptions changes during the conversation.

1 INTRODUCTION

Interpreting definite descriptions requires using information about different mental attitudes—in particular, attention, intention, and (shared) belief —and keeping track of the evolution of this 'mental state' [Hobbs, 1979; Heim, 1982; Cohen, 1984; Grosz and Sidner, 1986]. I propose to characterize the reasoning processes involved in the interpretation of definite descriptions using a Situation Theoretic logic called Episodic Logic [Hwang and Schubert, 1993; Hwang, 1992] and adapting Perrault's action-based model of belief update [Perrault, 1990]. The main ideas I discuss here are:

- Following Hawkins [Hawkins, 1978] and Barwise and Perry [Barwise and Perry, 1983] I assume that a large part of the task of interpreting definite descriptions consists in the identification of a *resource situation* in which the referent of the definite description is located. (This is known as the *location theory*.)
- I hypothesize that the 'discourse structuring principles' presented in the literature are best seen as principles for organizing the information contained in the mental state in a set of *possible situations*, including a *situation of visual attention*. 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, especially with the model proposed by Grosz and Sidner.
- I propose that Cohen and Levesque's program of deriving from general properties of actions facts about the pragmatics of natural language that previously seemed to require independent stipulations [Cohen and Levesque, 1990] can find an application in the study of reference phenomena. For example, I propose that the phenomenon of restricted anaphoric access to discourse antecedents known as 'discourse segmentation' results from the general tendency of agents of organizing events which 'fit together' into *courses of events* or *threads*, which are themselves situations.
- I propose that whether a definite description is interpreted anaphorically or with respect to the currently visible situation depends on which among several competing *Principles for Anchoring Resource Situations* is found to yield a consistent hypothesis, and I give a formalization of these axioms as default rules.
- I propose 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.

2 THE PROBLEM

The facts about definite descriptions I discuss are illustrated by the transcript fragment in (1). This transcript is part of a corpus collected for the TRAINS project [Allen and Schubert, 1991], whose aim is to develop a natural language understanding system able to engage in conversations with a human user (the *manager*) 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. The corpus used for my analysis consists of twelve transcripts of conversations between two speakers, in which one speaker plays the role of the system and the second speaker plays the role of the manager. The manager's utterances are marked with 'M', the system's utterances with 'S'.

(1) 13.1 M: ...not at the same time
13.2 ok
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 M: oh, we can do that?
16.1 S: yeah
...
29.1 M: 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 M: okay
31.2 then load the boxcar with bananas

The definite descriptions "the boxcar" in 13.3 and 29.5 are instances of *visible situation use* of definite NP's, 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). The plan discussed in (1) 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 present 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. 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 or situations, and that this *focus of attention* changes during a conversation [Grosz, 1977].

The fragment in (1) displays an instance of another well-known phenomenon, namely, the fact that when a definite description is used anaphorically,[1] the only antecedents considered are those in the same 'discourse segment' [Reichman, 1985; Grosz and Sidner, 1986; Fox, 1987]. Both "the boxcar" in 14.2 and "the boxcar" in 31.2 are cases of anaphoric use, yet "the boxcar" in 31.2 is unambiguous.

Finally, (1) 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 31.2 is a part of 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. That this interaction takes place is a rather compelling arguments against having the algorithm for belief update depend on syntactic information only.

[1] 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).

3 EPISODIC LOGIC

Due to the fact that it is being developed as the translation language for an English GPSG grammar [Schubert and Pelletier, 1982; Hwang, 1992], Episodic Logic is a fairly conservative version of Situation Theory[Barwise and Perry, 1983; Devlin, 1991], both in its syntax and in its semantics. The language of Episodic Logic is a conventional typed language with restricted quantification, lambda abstraction and adverbial operators; its semantics is closer to the logic proposed by Kratzer [Kratzer, 1989] than to the mainstream version of Situation Theory as exemplified, say, in [Devlin, 1991].

The crucial feature of Episodic Logic, for the purposes of this work, is that its language allows reference to *situations*. Situations are collections of facts reflecting not so much how the world is organized, but rather how we organize our information about it. More specifically, the language of Episodic Logic includes an operator for talking about *truth at a situation*, the * operator. The fact [f * s], to be read f *characterizes* s, is true in a model iff the fact f is true in the situation denoted by s. In part for the sake of understandability, in part for technical reasons that need not be discussed here, I use here the more standard notation [s $\models$ f] introduced by Devlin [Devlin, 1991]. To avoid confusions, I called the logic used here SEL^D (for 'Simple Episodic Logic with Defaults'). The $\models$ operator can be used to write axioms describing reasoning about situations.

All versions of Situation Theory include constructs for talking about *events*. In Episodic Logic, events and situations are 'of the same type' in the sense that the $\models$ operator can be used to characterize both. Thus, the translation of "John left" in Episodic Logic is as in (2). This expression reads that e is an event of John leaving which takes place at location l which precedes **now**.[2]

(2) ($\exists$ e at-about(e,l) $\wedge$ before(l,**now**)
[e $\models$ leave(john)])

That events have temporal *locations* is described in Episodic Logic using the predicate AT-ABOUT: at-about(e,l) reads "the event e has temporal location l." (For simplicity, I consider only temporal

[2] In (2), as in the rest of the paper, I use variables with an 'e' suffix like e or ce to denote events, and variables suffixed with 's' like s or is for statives and situations. I also adopt the Episodic Logic convention of using square brackets for infix operators like $\models$. Finally, I use throughout 'numbered variables' ce31.2, e31.2 etc. when presenting the translation of utterance 31.2.

locations here.) The set of logical operators of Episodic Logic that I will use also includes the kind-forming operator K [Schubert and Pelletier, 1982] (derived from the μ operator used in [Schubert and Pelletier, 1988]), 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 [Barwise and Perry, 1983].

Hawkins [1978] proposed that the uses of definite NP's he observed, as well as the contrast between definites, indefinites, and demonstratives, can be accounted for by stipulating that a speaker, when using a definite article,

1. instructs the hearer to locate the referent in some shared set of objects, and
2. refers to the *totality* of the objects/mass within this set that satisfy the restriction.

I propose to formalize this idea by associating the following translation to definite descriptions:

(3) "the boxcar" $\leadsto$

$$\lambda\ P\ [\text{the}\ x\ ([\dot{S} \models \text{boxcar}(x)] \wedge \text{shared}(\mathbf{spkr}, \mathbf{hearer}, \dot{S}))]\ P(x))$$

According to (3), "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 *parametric situation* $\dot{S}$. The terms **spkr** and **hearer** in (3) 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 (3). For the purpose of this paper, parameters like $\dot{S}$ can be thought of as open variables. The task of the process of definite descriptions interpretation is to identify the resource situation of each definite.

4 THE STRUCTURE OF THE COMMON GROUND

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,' which, in this case, means information obtained from the map.
2. Generic information about the task.
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 to 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 called attentional state.

The data about the use of referential expressions in our dialogues presented in §2 provide good evidence that the participants to our conversations assume that the information contained in the attentional state is 'carved' in '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, assuming that the chunks formed by the conversational participants are situations is almost a truism. Each of these situations is predicted to be shared by Clark and Marshall's [1981] copresence heuristics. Finally, I propose that some of these situations are defined *intensionally* as opposed to extensionally—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 will a conversational participant assume that a piece of information is part of that situation. These Situation Forming Principles are also mutually known to the participants of 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. I'll call the latter *possible* situations. The crucial property of possible situations is that, for s a possible situation and Φ a proposition,

$$[s \models \Phi] \not\models \Phi$$

The attentional state, then, consists of situations which may be classified along two different axes—actual vs. possible, intensional vs. extensional. I assume here the existence of two such situations. One is characterized by the information contained in the map. I call this 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), when they 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. The current *situation of attention* is a subset of MapS which plays a role similar to Barwise and Perry's 'object we are attending to' ([1983], pag. 87). I use 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 is they mutually know that both of them see the situation. This is represented in $\mathcal{SEL}^D$ as follows:

(4) `msee(a,b,s)`

A second situation the participants to our conversations appear to be referring to consists of the objects and actions assumed to be part of the plan. Again, the argument for assuming the existence of such a situation is that our participants refer by means of definite descriptions to objects which have only been introduced in the plan, as seen in §2; this observation can be reconciled with the location theory if we hypothesize that the participants to 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 into the plan. This situation, that I call PlanS, is intensionally defined and 'possible' in the sense discussed above.[3]

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 an course of action may vary—it is either 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. I propose that PlanS is a course of action, its situation forming principle being the one just described.

Because the events in a course of action form a sequence, we can define functions `pred(e,coa)` 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 each two successive events in a course of action, a function R can be defined which returns the time interval between the culminations of the two events.

5 ACTION-BASED MODELS OF BELIEF UPDATE

The aim of the work on intention recognition of Allen, Cohen, Levesque, Perrault and others is to model the process by which a hearer gets 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. I propose to use a representation of this sort to model the process of mental state change.

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]. For the purposes of this paper, the relevant aspect of these models of action is that the occurrence of an event results in certain *states* holding at the time after the event. This allows to write axioms specifying what the occurrence of a *conversational event* [4] tells about the intentions and/or beliefs of the speaker—for example, that the use by the speaker of a declarative sentence indicates his/her intention of informing the hearer about some fact.

The model proposed by Perrault in [Perrault, 1990] allows for a fairly intuitive formalization of the effects of a conversational event. Perrault's main concern is to formulate axioms for declarative sentences which predict when the hearer gets to believe p as the result of hearing an assertion that p. Perrault proposes a *persistence theory of belief*, according to which the effects of speech acts are formulated as defaults 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 (P) : $\vdash \mathrm{B}_{x,t+1}\mathrm{B}_{x,t}p \supset \mathrm{B}_{x,t+1}p$

Finally, Perrault provides two default inference rules, one for declaratives and one for Belief Transfer ($\alpha \Rightarrow \beta$ stands for $\frac{\alpha:\beta}{\beta}$ in Reiter's notation):

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

[3] In the longer version of this paper, [Poesio, 1993b], I discuss in detail the Situation Forming Principles involved in the definition of the situations used in our dialogues. Here, I only discuss informally the situation forming principle for the *discourse situation*, the situation which consists of the conversational events which occurred in the dialogue.

[4] It is one of the assumptions of the work on TRAINS that linguistic items other than complete sentences —for example, cue phrases—may also result in changes in the mental state. For this reason, I use in the paper the term conversational event to indicate those components of a discourse which result in a belief update. Here I use 'conversational events' as synonymous with 'locutionary acts.'

Belief Transfer : $B_{x,t}B_{y,t}p \Rightarrow B_{x,t}p$

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 model allows for a simple formulation of the principles of pragmatic interpretation that I propose in §9. For example, the focus shift rules I propose 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 §9.)

(5) $\vdash DO_{x,t}\alpha \ \& \ \Phi \ \Rightarrow B_{y,t+1}\ MSEE_{x,y,t+1}\ \beta$

In addition, making the occurrence of the conversational event a part of the information state allows us to provide a situation-theoretic formalization of Grosz and Sidner's treatment of discourse segmentation. I also need an improved event ontology to do this, however, so I'll postpone this issue until §7. Last, but not least, this theory (as well as the competing theories, like Cohen and Levesque's) has been developed to formalize one form of 'practical reasoning' —the process of intention recognition—which, as we have seen 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 descriptions interpretation.

Some work is needed, however, to adapt Perrault's proposal to my needs.[5] In the previous two sections, I have discussed how situations can be used to formalize the goal of the process of definite description interpretation. Also needed are formal rules of translation similar to, say, the GPSG system or the DRT construction rules; however, the language used by Perrault is not really suited to serve as a translation language for natural language, which makes it difficult for somebody 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.

6 CONVERSATIONAL EVENTS IN SITUATION 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. Finally, I follow Perrault in using Default Logic to formalize non-monotonic inferences, being aware of the problems with this approach.[6] I assume therefore a proof theory like that of Episodic Logic, but extended with default inference rules of the sort used by Perrault.

If we were to directly rephrase Perrault's observability axiom for the case of imperatives as stating that a (conversational) event ce of **spkr** instructing **hearer** to do Φ results in **hearer** acquiring a belief b characterized by **hearer** BELIEVING that **spkr** instructed **hearer** to Φ , utterance 29.5 in (1) would translate into the SEL^D expression in (6), 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 $K(\lambda$ e ...) in (6) is an *event kind*; event kinds are used in Episodic Logic as the uniform translation of infinitives like "to pick up the boxcar." I will discuss how (6) is obtained from 29.5 in §8. (This translation for 29.5 is further revised below.)

29.5 U: pick up the boxcar,

(6)
```
(∃ 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,ĉ ) ∧
                [le29.5 ⊨
                 (the y [Ṡ  ⊨ boxcar(y)]
                        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 SEL^D, but instead I have adopted a formulation based on a proposal by Hans Kamp [1990]. Kamp proposes that mental attitudes are always asserted to hold at an indexical **now** point (**now** stands for 'now'). This way of achieving persistence eliminates the need for performing persistence reasoning, and is therefore appeal-

[5]The problems of Perrault's theory as a theory of the reasoning involved in discourse interpretation have been discussed in the literature [Appelt and Konolige, 1988].

[6]An alternative would be using a probabilistic logic in which beliefs are augmented with probabilities. Schubert and Hwang are working on formalizing a version of Episodic Logic which includes a probabilistic version of material implication.

ing both from a conceptual and from a practical point of view. However, it shares with Perrault's the problem that older beliefs never get to be disbelieved.

7 DISCOURSE SEGMENTATION

The organization of conversational events into discourse segments is, I propose, the result of a situation-forming principle, according to which agents tend to group events into courses of actions. I call the courses of conversational events *conversational threads.*

Conversational threads can be used to do the work done in Grosz and Sidner's theory by the *focus space stack.* The focus space stack models the effects of discourse segmentation on anaphoric reference as follows: 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 on the stack. When an utterance completes the current discourse segment, the current focus space is popped from the stack.

Each of the operations proposed by Grosz and Sidner on the focus space stack corresponds to a basic 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'. Because conversational threads are situations, they afford a natural reformulation of the focus space stack model: for each NP of the current utterance which introduces an anaphoric antecedent (hence would result in the addition of an object to the focus space) a new constituent is added to the current conversational thread. (The constituents of a conversational thread are also called *discourse referents.*) An utterance which opens a new discourse segment results in a new conversational thread being open, which informationally subsumes the previously current conversational thread. An utterance which closes a discourse segment and pops to a previously current focus space results in the corresponding conversational thread becoming current.

Grosz and Sidner propose that discourse segmentation is "parasitic upon the intentional structure" (p.180), in the following sense: whether a hearer inserts an utterance into a certain 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. According to Grosz and Sidner, 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. We may use the dominance and satisfaction-precedes relations between the discourse purposes expressed by conversational events to define derived relations among those conversational events, that I will call *dominance** and *satisfaction-precedes**. In this way we can formulate the following principle governing the process by which a hearer achieves the belief that a conversational event is part of a conversational thread:

CT Membership Principle (CTMP):
A conversational participant achieves the belief that a conversational event is part of a conversational thread iff the belief that that conversational event is dominated* or satisfaction-preceded* by another conversational event in that conversational thread is part of the information state of that participant.

Returning to our example 29.5, the following SEL^D expression translating the speaker's meaning inferred from the occurrence of 29.5 revises (6) to incorporate the proposal that *each* event—whether 'described' or 'conversational'—is a sub-episode of a course of events. $\dot{c}'$ in (7) is a parameter to be resolved to a particular conversational thread by the process of intention recognition.

```
(7)  (∃ bs29.5 at-about(bs29.5,now)
        (∃ ce29.5 at-about(ce29.5,l_3) ∧
                  cause(ce29.5,bs29.5)
                  ∧ subepisode-of(ce29.5,ċ′ )
          [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)]
                          pickup(self,y))]
                          ))]))])))
```

8 DISCOURSE INTERPRETATION

I assume here a roughly Gricean strategy, with 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], quoted in [Chierchia and McConnell-Ginet, 1990]).

In the system I have developed, as in the work of Hwang and Schubert [1993], the separation between 'linguistic' and 'pragmatic' aspects of definite phrases interpretation is 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 Pelletier [1982], I call this interpretation the *Logical Form* of the utterance. The Logical Form is generated by a GPSG grammar which uses SEL^D as its target language, but is otherwise very similar to the grammars discussed in [Schubert and Pelletier, 1982;

Hwang and Schubert, 1993], therefore I'll omit presenting it here. The Logical Form for an imperative sentence such as 29.5 is the SEL^D expression in (8). (Imperatives are translated as tenseless VPs in the scope of an operator IMPER expressing the sentential force of the sentence.)

29.5 U: pick up the boxcar

(8)
```
(imper
  (λ x
     (the y [Ṡ ⊨ boxcar(y)]
      (pickup(y))(x))))
```

The second translation represents the interpretation of the utterance as the performance of an *action* by the speaker, who, in doing that, intends to achieve certain effects. I call this interpretation the *Conversational Event* associated with the utterance, and we have seen examples above (e.g., (7)). The conversational event associated with an utterance corresponds roughly to Austin's *locutionary act*, and is obtained by the logical form by means of *Conversational Events Generation Rules*. The input to the Conversational Event Generation Rules consist of a pair ⟨logical form,context⟩, where the context is a triple ⟨speaker, addressee, temporal location⟩. Expressions like (7) in §6 are the output of these rules. The rule for imperatives is as follows:[7]

Imperatives

```
ce:⟨(imper (λ x Φ (x))),⟨spkr,hearer,t⟩⟩
   ⇝
(∃ bs at-about(bs,now)
   (∃ ce at-about(ce,l′) ∧ cause(ce,bs) ∧
       subepisode-of(ce,ċ ) ∧ T = l′
     [bs ⊨
      believe(hearer,
       [ce ⊨
         instruct(spkr,hearer,
          K(λ  e subepisode-of(ce,ċ′ ) ∧
                [e ⊨  (λ x Φ (x))(hearer)]
                ))]))]))
```

9 THE 'VISIBLE SITUATION' USE OF DEFINITE DESCRIPTIONS

As an example of the kind of theory that the system presented here can be used to formalize, I'll present my current hypothesis about the interpretation of 'visible situation' definite descriptions. As discussed in §2, a relation can be observed in our dialogues between the visible situation use of definite descriptions and 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 though other objects of the same type have been introduced in the discourse or are part of the world described by the map.

I propose that two kinds of principles are involved in the interpretation of definite descriptions interpreted with respect to the visible situation. First of all, there are (usually, defeasible) principles formulating hypotheses about ways for anchoring resource situations: I call these *Principles for Anchoring Resource Situations* One such principle hypothesizes that, if there is a current situation of mutual attention, then that situation may be the resource situation for a definite description. Second, there are principles governing visual attention shifts.

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 mutual attention of the conversational participants is currently 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 a function DESCRIBED-EVENT-OF from conversational events to the event described by the utterance the conversational event is associated with:

(PARS1)
```
[bs₁ ⊨ believe(y,
        (described-event-of(ce) = e ∧
        [e ⊨  (the z [Ṡ  ⊨ p(z)]
                    q(z)))]
       ∧ at-about(ce,t))]
∧ at-about(bs₁,now) ∧ t = now ∧
[bs₂ ⊨
 believe(y,
  intend(x,
   ([fs ⊨ msee(x,y,s)] ∧
    at-about(fs,r(pred(ce,coa),ce)))))]
∧ at-about(bs₂,now)
⇒
  (∃ bs₃ at-about(bs₃,now)
     [bs₃ ⊨
      believe(y,intend(x,anchor(Ṡ ,s)))])
```

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

Follow The Movement : part of the intended effect of an utterance instructing an agent to move an object from a location to another is to make the terminal location of the movement the new mutual situation of attention.

The machinery I have been developing so far allows us to formalize this principle by means of the following default inference rule:[8]

(9) [bs ⊨

[7]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.

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

```
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 (9), place(p,MapS) is the 'place situation' consisting of the facts and the objects in MapS which describe the position p.

Let us see now how these principles are applied when interpreting utterances 29.4-29.5 in (1). The conversational event generation rules produce the following translation for 29.4:

29.4 U: take engine E1 to Dansville,

```
(10) (∃ 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 simpoe 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 thus modified. The conversational event associated with 29.5 is described in (7) seen before. 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 to be place(dansville,MapS):

```
(11) (∃ bs4 at-about(bs4,now)
        [bs4 ⊨
          believe(self,
           intend(user,
            anchor(Ṡ ,place(dansville,MapS))))])
```

(The statement anchor(Ṡ ,f) 'fixes' the interpretation of the parameter Ṡ to f.) Because there is only one boxcar in place(dansville,MapS), b1, it is consistent for the system to infer (11). In turn, this entails that

```
(∃ bs5 at-about(bs5,now)
   [bs5 ⊨
     believe(self,
      intend(user, pickup(self,b1)))])
```

10 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 system's representation of the mental state and to perform scope disambiguation and reference interpretation. The input to SAD-92 is a logical form of the kind discussed in §8, obtained by the (GPSG) parser of TRAINS-92. 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, 1993a] and intention recognition [Heeman, 1993]. Currently, the inference engine which computes the consequences of the mental state change axioms is a simple version of EPILOG (the inference engine for Episodic Logic). An hypothesis about the intentions expressed by the user in uttering the sentence is obtained; the consistency of this hypothesis is verified by trying 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].

Acknowledgements

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

References

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

[Appelt and Konolige, 1988] Appelt, D. and Konolige, K. 1988. A practical nonmonotonic theory of reasoning about speech acts. In *Proc. ACL-88*, Buffalo.

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

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

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

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

[Cohen and Levesque, 1990] Cohen, P. R. and Levesque, H. J. 1990. Rational interaction as the basis for communication. In Cohen, P.R.; J.Morgan, ; and Pollack, M., editors 1990, *Intentions in Communication*. Morgan Kaufmann. chapter 12, 221–256.

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

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

[Ferguson, 1992] Ferguson, G. M. 1992. Explicit Representation of Events, Actions, and Plans for Assumption-based Plan Reasoning. Technical Report 428, University of Rochester.

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

[Grice, 1957] 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.

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

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

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

[Heeman, 1993] 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..*

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

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

[Hwang and Schubert, 1993] Hwang, C. H. and Schubert, L. K. 1993. Episodic logic: A situational logic for natural langage processing. In Israel, D. and Peters, S., editors 1993, *Situation Theory and its Applications, v.3*. CSLI. To appear.

[Hwang, 1992] Hwang, C. H. 1992. *A Logical Approach to Narrative Understanding*. Ph.D. Dissertation, University of Alberta, Department of Computing Science.

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

[Kratzer, 1989] Kratzer, Angelica 1989. An investigation of the lumps of thought. *Linguistics and Philosophy* 12:607–653.

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

[Poesio, 1993a] Poesio, M. 1993a. The scope and deindexing module of TRAINS-92. TRAINS Technical note, University of Rochester, Department of Computer Science. To appear.

[Poesio, 1993b] Poesio, M. 1993b. A situation-theoretic formalization of definite descriptions interpretation in plan elaboration dialogues. In Israel, D. and Peters, S., editors 1993b, *Situations Theory and its Applications, v.3*. CSLI. To appear.

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

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

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

[Schubert and Pelletier, 1988] Schubert, L. K. and Pelletier, F. J. 1988. Generically speaking. In Chierchia, G.; Partee, B. H.; and Turner, R., editors 1988, *Properties, Types and Meaning, vol. II*. Kluwer. 193–268.

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

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

[Traum, 1993] Traum, David R. 1993. The TRAINS-92 dialogue manager. In *Working Notes AAAI Spring Symposium on Reasoning about Mental States: Formal Theories & Applications..*