

Planning English Referring Expressions

Douglas E. Appelt

*Artificial Intelligence Center, SRI International, Menlo Park,
CA 94025, U.S.A.*

Recommended by B. Webber

ABSTRACT

This paper describes a theory of language generation based on planning. To illustrate the theory, the problem of planning referring expressions is examined in detail. A theory based on planning makes it possible for one to account for noun phrases that refer, that inform the hearer of additional information, and that are coordinated with the speaker's physical actions to clarify his communicative intent. The theory is embodied in a computer system called KAMP, which plans both physical and linguistic actions, given a high-level description of the speaker's goals.

I. Introduction

One major goal of artificial intelligence research in natural-language generation is to develop a means of producing utterances that are natural and as close as possible to what humans would produce, given a similar domain and a similar need to communicate.

To use language with the competence of a human, it is not sufficient to have only a description of the syntactic, semantic, and discourse rules of a language: Human language behavior is part of a coherent plan of action directed toward satisfying a speaker's goals. Furthermore, sentences are not straightforward actions that satisfy only a single goal. When people produce utterances, their utterances are crafted with great sophistication to satisfy multiple goals at different communicative levels.

Fig. 1 illustrates a typical situation arising when two people cooperate on a common task, in which a speaker plans an utterance that has multiple effects on the intended hearer. The speaker moves the wheelpuller in the direction of the hearer and says "Use the wheelpuller to remove the flywheel." The hearer, who is observing the speaker while he makes the request and knows the speaker is referring to the particular tool, thinks to himself, "Ah, so *that's* a wheelpuller." The speaker's gesture with the wheelpuller is an instance of

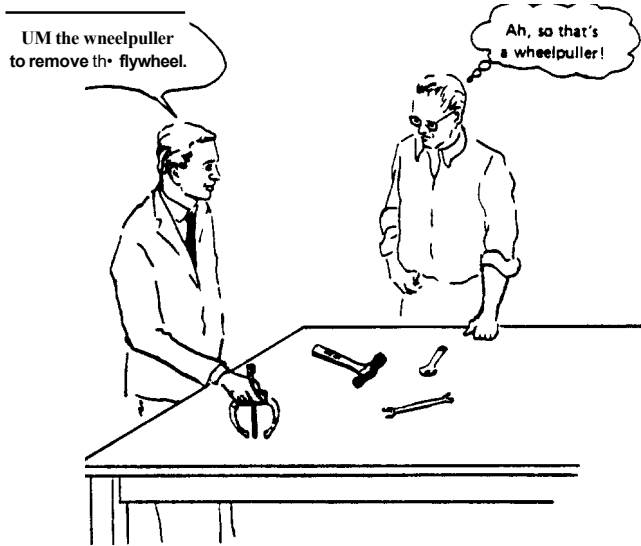


FIG. 1. Satisfying multiple goals in an utterance.

communicative pointing, i.e., using a physical action to communicate an intention to refer to something.

In this situation, most obviously, the speaker is requesting that the hearer carry out a particular action, because the use of the imperative strongly suggests that a request is intended. Notice, however, that the speaker includes using the wheelpuller as part of his request. If he knew that the hearer did not know that he was supposed to use the wheelpuller to remove the flywheel, then his utterance was also intended to communicate the knowledge of what tool to use for the task. Also, the fact that he is *pointing* to a particular object and doing so in an obvious way communicates the speaker's intention to refer to it with the noun phrase "the wheelpuller". Because the intention to refer has been communicated, the noun phrase also communicates the fact that the intended referent is called a wheelpuller. The speaker could have said "Remove the flywheel", if he thought the hearer knew how to do it, and he could have said "Use *that thing* to remove the flywheel", if he had no goal of informing the hearer that the tool is a wheelpuller. In this case, the speaker has probably reasoned that future communication involving the wheelpuller will be necessary, and if the speaker and hearer mutually believe some description of the object other than its physical properties, it will be easier to refer to it in the future. The speaker and hearer mutually know that the phrase "the wheelpuller", rather than "the flywheel", is intended to refer to the object of the gesture because they share sufficient knowledge of the domain to determine that the intended referent of "the flywheel" is some other object.

The satisfaction of multiple goals in utterances is more the rule than the exception in communication. There are many different combinations of goals that can be simultaneously satisfied, and utterances with such combinations are common in everyday conversation. Here are a few examples:

- *Referring and communicating additional information.* A rock climber says to a friend, "Joe and I are going to climb the Snake Dike route on Half Dome next weekend." The speaker does not use the prepositional phrase "on Half Dome" to pick out a particular Snake Dike route from several that he and the hearer mutually know about, but rather assumes the hearer has never heard of the route, and provides additional information to inform him where it is located.
- *Referring and communicating an emotional attitude.* A speaker points to an accused spy and says "That scoundrel is the one who betrayed us!" The speaker is not trying to distinguish a particular scoundrel from some set of scoundrels. Although the speaker may be attempting to inform his audience that the intended referent is a scoundrel, quite possibly he is using a pejorative expression to convey his emotional attitude toward the intended referent.
- *Requesting and being polite.* Multiple-goal satisfaction even plays a role in such conventional utterances as "Could you tell me what time it is?" In this case the speaker chooses the indirect speech act to satisfy a goal of demonstrating politeness toward the hearer, while the more direct but less polite "What time is it?" would convey the request equally well.

These examples illustrate how a great deal of sophisticated reasoning about the effects of an utterance can be required to produce a seemingly simple sentence. A speaker capable of producing such utterances can be modeled by the process illustrated in Fig. 2. The speaker, modeled by the planner in the center of the diagram, plans to satisfy physical, knowledge-state, discourse and social goals, using knowledge about the world, his own mental states and those of the hearer, and knowledge about the language. The speaker's plans can ultimately entail both physical and linguistic action.

A planning system such as the one in Fig. 2 has been implemented as part of this research and is called KAMP, which stands for Knowledge And Modalities Planner. KAMP is a hierarchical planning system that uses a nonlinear representation of plans called a *procedural network* [1]. It is capable of taking a set of axioms about the state of the world, the preconditions and effects of actions, the beliefs of different agents and a description of an agent's high-level goal, and producing a plan from the perspective of the agent that involves the performance of both physical and linguistic actions. The linguistic actions are refined until an English sentence is completely specified.

A primary consideration in the design of KAMP was to avoid dividing the language planning task into two independent processes of deciding *what* to say and deciding *how* to say it. In this view, the process of deciding how to say

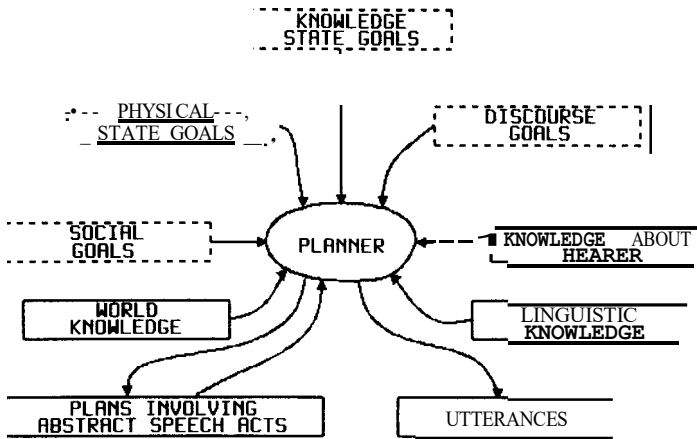


FIG. 2. Overview of an utterance planner.

something is essentially a process of translation from a specification of the propositional content to the actual utterance. The propositional content remains unchanged by this process. Previous language-generation systems have always made this distinction (e.g. Mann [2]; McKeown [3]), because it allows one to separate the linguistic part of the system from everything else. Intuitively, this modularity is both theoretically and practically attractive.

In spite of its appeal, the 'what-how' distinction has less merit when examined in the light of a theory of language generation based on planning. Such a theory views communication as actions directed toward satisfying goals. There are many decisions at every level of the language planning process that can be described in terms of action and goal satisfaction. At the highest level, there is planning illocutionary acts, at lower levels there is deciding how to communicate an intention to refer, and deciding how to communicate intentions about the discourse. The actions that satisfy these goals depend to different degrees on what linguistic choices are available in the current context. Thus, the planning at each level involves consideration of both linguistic rules and goal satisfaction. The distinction between 'what' and 'how' then becomes merely two points on a continuum between goal-satisfaction and rule-satisfaction processes, and no modularization based on the distinction is obvious. This does not, of course, imply that modularization per se is undesirable, but that the dimension along which the modules are separated is not one of 'deciding what to say' and 'deciding how to say it'. For example the TELEGRAM grammar [23] separates knowledge about the grammar and its semantics from the rest of the knowledge the system needs, but integrates the processes that uses these knowledge sources so that planning plays a role in all aspects of utterance design.

Criticism of other research efforts may be somewhat unfair because the resulting limitations on what they can do, for example, the lack of ability to coordinate physical and linguistic actions, were not problems that these researchers have sought to overcome. However, designers of natural-language generation systems will ultimately have to confront these issues.

A hierarchical planner was selected as the design of KAMP because it provides for a separation between the planning of domain-level goals and actions and low-level linguistic actions as well as intermediate levels of abstraction that facilitate the integration of multiple goals into utterances. The abstraction hierarchy coincides with the spectrum of goal satisfaction versus rule satisfaction, which makes this design well suited to planning without a sharp distinction between 'what' and 'how'.

This work is closely related to other recent work in artificial intelligence and linguistics. The idea of using speech acts as operators in a planning system originates with Bruce [4], and Cohen and Perrault [5]. Evans [7] independently developed a theory of speech acts based on situation semantics [8] that shares many of the fundamental assumptions of the research reported in this article. An initial description of KAMP and utterance planning was given in two previous papers by Appelt [9, 10]. This paper describes the system and its underlying theory as it has been fully developed and implemented.

This paper is organized in three main sections. Section 2 describes the problem that is being addressed. Section 3 describes a formal approach to axiomatizing referring actions, and Section 4 illustrates the KAMP planner with an example.

2. English Referring Expressions

Speakers use noun phrases for many different purposes, and this research has not attempted to deal with all of them. This paper will be concerned only with planning singular, specific, definite noun phrases that do not contain explicit quantifiers. Even within this narrow domain of phenomena, there exist a number of different referring intentions that a speaker may have when he utters a noun phrase, and part of a hearer's task in understanding an utterance is figuring out which one the speaker has in mind. This intention-recognition process is necessary because identical noun phrases may be used in different contexts with different intentions. When a speaker utters a definite noun phrase, the hearer must decide whether the speaker intends that the hearer actually *identify* some person or object in the world to which the speaker intends to refer. The following intentions underly uses of the noun phrases under consideration:

- *Speaker intends to refer to a mutually known object.* In this case there is assumed to be a mutually known object to which the speaker intends to refer. A mutually known object is an object of which the speaker and hearer

mutually know some properties. This mutual knowledge arises out of the immediate context either by being (to use Prince's terminology [11]) *situationally evoked* (part of the shared context) or *textually evoked* (previously introduced to the discourse). Pronominal and anaphoric noun phrases are used to communicate intentions of this type.

- *Speaker intends that the hearer identify a referent.* The speaker may intend to refer to an object (of which he and the hearer may or may not share mutual knowledge), and intends that the hearer, based on his understanding of the speaker's intentions, perform whatever actions are necessary to identify the individual to which the speaker intends to refer. On occasion, speakers may convey this intention explicitly by means of a request to find the referent. However, this intention often underlies referring actions for which no explicit identification request is made. For example, the italicized noun phrase in "Turn left at *the third light past the large pink house on the left.*" is an example of a noun phrase that could be uttered with this intention. What distinguishes this case from the previous case is that there is not necessarily any mutual knowledge of the intended referent at the time of the utterance; in the above example, the hearer may have no prior knowledge at all of the large pink house. The speaker's implicit intention that the hearer identify the -referent may require the hearer to form and execute a complex plan to make the identification. Instead of planning a description with respect to the speaker and hearer's mutual knowledge, he tries to plan a description that is *useful* for the hearer to plan an identification action. For example, a speaker does not give the hearer a useful description when he tells the hearer (whom he has just met on the bus) which bus stop to get off by saying "Get off one stop before I do," because there is no way the hearer can use the description to form an effective plan to identify the referent, even though the noun phrase "one stop before I do" semantically denotes the right thing.
- *Speaker intends that the hearer not identify a referent.* When a definite noun phrase is used with this intention it is commonly called *attributive*. The speaker may have a description of an individual that relates the individual to the mutual knowledge of the speaker and hearer, and the speaker may or may not know which individual satisfies that description. The hearer must realize that the speaker does not intend to say anything about a particular individual, but rather of whatever individual it is that satisfies a description. The noun phrase *John's children* (where John does not have any children yet) in the sentence "*John's children* will be very rich", is used in this manner.

According to KAMP's model of language production and understanding, a speaker uttering something implies he intends for the hearer to recognize the intended propositional content. This entails recognizing predicates and the arguments to which they are applied. These arguments are drawn from a set of so called 'active' concepts. The speaker introduces concepts to this set through the performance of *concept activation actions*. These actions are frequently

performed by means of *referring expressions*, which are expressions that bear a semantic denotation relationship to objects in the world. All of a speaker's physical and linguistic actions may be instrumental to conveying the intentions behind a concept activation, so while referring expressions are an important form of realization, they are not the only means that must be considered. Also, a concept can be active by means of inferential connection to active concepts even though it is not explicitly the object of a concept activation.

This paper will consider only the planning of concept activation actions in which the speaker and hearer mutually know some facts about the object of the concept activation at the time of the speaker's utterance. An adequate formal theory of concept activation actions should account for the following phenomena: (1) how speakers reason about mutual knowledge to arrive at a description of the object; (2) how speakers use nonlinguistic means (e.g. pointing) to contribute toward the satisfaction of a goal to activate a concept; and (3) how speakers plan noun phrases that satisfy goals in addition to reference. The next section examines these questions.

3. A Formal Theory of Reference Planning

The logic that is used for the formal description of referring actions is based on Moore's logic of knowledge and action [12]. This logic consists of an *object language* whose terms denote actions and objects in the world, and a *meta-language* whose terms also denote individuals, but which (in addition) can denote object-language terms and *possible worlds*. The object language includes intensional operators such as **Know** and **Intend**, the semantics of which are axiomatized in the first-order meta-language in terms of possible worlds. The semantics of action-terms in the object language are axiomatized as relations on possible worlds, where actions map one world into another, with possible worlds playing a role similar to the familiar states of a conventional logic of action [14]. The reader familiar with model-theoretic semantics will notice that the notion of possible world used here is quite different from the familiar notion of a world as a complete course of events, and should bear this distinction in mind throughout the article.

In this article, the following notational conventions will be adopted: Intensional operators (e.g. **Know**) are written in boldface type. Predicates, functions, and constants appear in lower-case type with initial capital letter (e.g. Kernel). Variables appear in lower-case italic type (e.g. *act*). Schema variables are upper-case italic type (e.g. *PS*). Most of the predicate naming conventions are taken directly from Moore [12] to facilitate cross-reference by the reader desiring more information.

A detailed description of Moore's scheme for reasoning about knowledge and action would be much too long for this article. This paper will only describe extensions to that theory that were necessary for utterance planning.

The reader should bear in mind that the following predicates are used in most of the examples, and have the following meanings:

- $T(w, P)$ means object-language formula P is true in world w .
- $D(w, x)$ is the denotation of object-language term x in world w .
- $\mathbf{K}(\mathbf{A}, w_1, w_2)$ means world w_2 is consistent with \mathbf{A} 's knowledge in world w_1 .
- $\mathbf{R}(e, w_1, w_2)$ means that world w_2 is the result of event e happening in w_1 .
- $\mathbf{@}(x)$ is the *standard name* of x , i.e., a term whose denotation is x in every possible world.

For the sake of simplicity, this article uses a loose notation to represent object-language logical connectives and quantifiers. The same symbols will be employed to represent similar operations in both the object and the meta-language. Moore [12] discusses the correspondence between object language and meta-language, including some problems associated with quantifying into opaque contexts.

One may argue that an adequate theory of language planning must be based on a theory of *belief* rather than a theory of knowledge. Although this is a valid point, an adequate theory of belief is difficult to formalize, because once one admits the possibility of holding beliefs that are not true of the world, a theory of belief revision and truth maintenance is required. It is true that Moore's logic of knowledge can be transformed into a logic of belief by appropriately weakening the axioms (viz. the axiom that asserts reflexivity of the accessibility relation on possible worlds) so that it is no longer possible to infer \mathbf{P} from $\mathbf{Know}(\mathbf{A}, \mathbf{P})$. However, without addressing all the problems associated with belief and justification, one has really accomplished little else besides changing the word **Know** to **Believe**. Because a detailed study of reasoning about belief is beyond the scope of this research, the axioms are presented using **Know** as a first approximation to the best theory.

3.1. Reference and concept activation

KAMP is based on a theory of speech acts that is similar to that described by Searle [13]. According to Searle, the speaker performs some utterance acts from which the hearer recognizes the proposition that the hearer intends to convey. The speaker conveys these components of the propositional content to the hearer by means of what Searle calls *propositional acts* of referring and predicating. Then the hearer infers what the speaker wants him to do with the proposition—for example, whether it is to affect his beliefs or his future actions. This constitutes the recognition of the *illocutionary force* of the utterance.

The problem with Searle's theory is that it is too strongly sentence-oriented. **With** the exception of conjunctions and conditionals, uttering a referring expression constitutes the performance of exactly one propositional act, and uttering a sentence constitutes the performance of exactly one illocutionary act.

The KAMP theory is an attempt to move beyond the sentence-oriented perspective and account for sentences that realize multiple illocutionary acts and illocutionary acts that appear to be realized over the course of several sentences. To achieve this goal it is necessary to formally decouple illocutionary acts and propositional acts from sentences and noun phrases. Toward that end, a hierarchy of linguistic actions is defined and axiomatized, and is employed by KAMP. At the top of the abstraction hierarchy are illocutionary acts. These actions are performed by means of surface speech acts. A surface speech act is realized by means of uttering a sentence.

KAMP plans concept activations, which are analogous to Searle's propositional acts, and these actions are performed by means of describing actions (realized directly as noun phrases) and communicative pointing actions. As with illocutionary acts and surface speech acts, it is possible to separate the intention communication action from its linguistic realization.

Concept activation actions and their realization will be the primary focus of this article. The verb 'refer' is often used in a sense that is very close to what is meant by activating a concept. If the subject of the verb 'refer' is a linguistic expression, then what is being described is a semantic relationship similar to what one means by 'denote'. When the subject of 'refer' is an agent, then the intended interpretation is "to perform a concept activation action by means of uttering a referring expression".

KAMP represents concepts as intensional object-language expressions that denote different objects in different possible worlds. In general, it is possible for object-language terms to denote different individuals with respect to each possible world. If an object-language term denotes the same individual in every possible world, it is called a *rigid designator*. A rigid designator is like a proper name, and it is often convenient to use such names in the theory because they simplify the axiomatization of the domain. Under such assumptions, it is easy to show that any agent can decide whether two rigid designators denote the same individual, and they are therefore useful for describing the process of an agent reasoning about who or what something is. Because of the simplification that results, the example in this article assumes that objects have standard names despite the implausibility of people having names for every individual (including, for example, each of the 200 identical screws in some parts bin).

3.2. Reasoning about mutual knowledge

The planning of concept activation actions requires not only the ability to reason about what different agents know, but also what they *mutually know*. Agents *A* and *B* mutually know *P* if and only if *A* knows *P*, *B* knows *P*, *A* knows that *B* knows *P*, *B* knows that *A* knows *P*, *A* knows that *B* knows that *A* knows *P*, and so on, ad infinitum. It is insufficient in planning a concept activation for *A* to consider only his own knowledge and the knowledge of *B*,

because failure to do so can result in an infelicitous reference. Clark and Marshall [15] demonstrate that it is possible to construct examples where the definite description fails to identify the right concept for the hearer if only a finite number of assertions about the speaker's knowledge of the hearer's knowledge are considered.

Since mutual knowledge and mutual belief follow from an infinite number of facts, it is impossible to deduce mutual knowledge directly from its definition. Cohen and Perrault [16] demonstrate that the mutual-knowledge condition on referring expressions is in fact too strong. They demonstrate that *A* can use a definite description to refer to an object in speaking to agent *B* if it can be concluded that the nested assertions about *A*'s knowledge about *B*'s knowledge hold in all but a finite number of cases. Unfortunately, this analysis still leaves the problem of verifying an infinite number of conditions.

Nadathur and Joshi [17] circumvent this problem by adopting a weaker condition for the use of a definite description: that the speaker believes that the hearer believes that the description denotes the object, and the speaker has no reason to believe that the chosen description is not mutual knowledge.

The KAMP theory maintains that a description used in a referring expression must be mutually believed, but admits heuristics by which mutual belief can be plausibly inferred without recourse to verifying an infinite number of conditions. Clark and Marshall call such assumptions *copresence heuristics*. According to their copresence heuristics, mutual knowledge results from three sources: (1) common membership in a community of speakers; (2) sharing physical proximity that enables the agents to observe, and observe each other observing; (3) linguistic exchange of information.

In KAMP, copresence heuristics (1) and (2) above are stated directly in the axioms describing a particular communicative situation. Heuristic (3) is captured by a suitable axiomatization of illocutionary acts, described fully by Appelt [5], namely that the successful performance of an illocutionary act produces mutual knowledge of its performance.

In addition to copresence heuristics, KAMP requires a logical representation of mutual knowledge from which it is possible to derive any one of the infinite consequences of an assertion of mutual knowledge. The mutual knowledge of two agents *A* and *B* is everything that is true in the *union* of the possible worlds compatible with *A*'s knowledge and *B*'s knowledge. Notice that the *intersection* of the propositions believed by two agents is represented by the *union* of possible worlds compatible with their knowledge. For the purpose of stating this fact formally, a pseudo-individual called the *kernel* of *A* and *B* is defined such that the set of possible worlds consistent with the kernel's knowledge is the set of all worlds consistent with either *A*'s knowledge or *B*'s knowledge. This leads to the following definition of mutual knowledge:

$$\begin{aligned} \forall w_1 T(w_1, \text{MutuallyKnow}(A, B, P)) = \\ \forall w_2 K(\text{Kernel}(A, B), w_1, w_2) \Rightarrow T(w_2, P). \end{aligned} \quad (1)$$

The second axiom that is needed is:

$$\forall x, w_1, w_2 K(x, w_1, w_2) : J \forall y K(\text{Kernel}(x, y), w_1, w_2). \quad (2)$$

Axiom (2) states that the possible worlds consistent with any agent's knowledge is a subset of the possible worlds consistent with the kernel of that agent and any other agent. Note that because the knowledge axioms that allow one to conclude that if an agent knows P , he knows that he knows P , apply to the kernel individual as well as ordinary individuals, it follows that if A and B mutually know P , then they mutually know that they mutually know P . This allows statements of the form

$$\mathbf{Know}(A, \mathbf{Know}(B, \mathbf{Know}(A, \dots)))$$

to be derived to any arbitrary depth. Axioms (1) and (2) can be used efficiently by a first-order logic theorem-prover that handles equational theories, such as that of Stickel [18].

3.3. Reasoning about intention

KAMP uses a possible-worlds semantics for intention that is similar to the possible-worlds semantics for knowledge and action. Reasoning about intention is crucial to utterance planning at several stages, because all actions with communicative intent (viz. illocutionary acts and concept activation actions) depends on the hearer's recognition of the speaker's intention for successful performance.

There are two levels at which KAMP describes an agent's intentions. First, an agent can intend to make a proposition true, or he can intend to perform an action. Thus,

$$\mathbf{Intend}(A, P)$$

means that agent A intends to bring P about, and

$$\mathbf{IntendToDo}(A, act)$$

means that A intends to perform act in the immediate future.

The semantics of **Intend** are that there is some set of possible worlds, called a *preference set* of an agent, such that for every world w in that preference set, P is true in w . This is expressed by Axiom (3).

$$T(w_1, \mathbf{Intend}(A, P)) = \exists s PS(A, w_1, s) \wedge \forall w_2 (w_2 \in s) \supset T(w_2, P) \quad (3)$$

in which $PS(A, w, s)$ is true if and only if s is a preference set in world w of agent A

The semantics of **IntendToDo** are similar. In that case, there is some preference set s such that for every world $w \in s$, w is the result of A performing action act . This is expressed by Axiom (4).

$$T(w_1, \text{IntendToDo}(A, act)) = \exists s \text{PS}(A, w_1, s) \wedge \forall w_2 (w_2 \text{Es}) : \exists R(Do(A, act), w_1, w_2). \quad (4)$$

It follows directly from (3) and (4) that if an agent intends to perform an action, then he intends to bring about the effects of the action.

These two axioms give KAMP a rudimentary ability to reason about what an agent intends to bring about and what he intends to do next which is adequate for KAMP to make simple multiple-agent plans. These axioms are not claimed to come close to an adequate theory of desire or intention.

3.4. An overview of KAMP

KAMP differs in a number of important ways from planning systems that are restricted to physical domains. The most fundamental difference is that the utterance-planning system is necessarily always planning in an environment with at least one other agent, and this introduces problems of cooperating with the other agent, or thwarting its goals [19]. The necessity of reasoning about different agents requires some means of reasoning about their knowledge, beliefs, and intentions, and how their mental states are affected by the actions that they and others perform.

The necessity of reasoning about propositional attitudes led to the adoption of the possible-worlds semantics representation described by Moore [12]. This formalism is best suited to proving that certain facts hold in a state of the world, and is not well suited to planning. Because knowledge states are represented as sets of possible worlds, straightforward application of a backward chaining algorithm to these sets in search of a plan is cumbersome at best. The design of the KAMP planner differs from other planning systems because of the need to overcome this disadvantage.

Fig. 3 illustrates the operation of KAMP. KAMP solves problems by using a heuristic problem-solving method, which is successful at finding a good plan with minimal effort most of the time while preserving the option to rely on brute-force search if heuristic methods fail. KAMP has two descriptions of the actions available to the planner. One description is in the form of axioms relating possible worlds, as described above. The other description is an *action summary*, which summarizes the preconditions and effects of actions in a STRIPS-like formalism (see [20]) involving preconditions and add and delete lists. The planner uses the action summaries as a heuristic guide to the selection of actions that are likely to result in a good plan. They are not intended to be complete descriptions of all the consequences of performing the action. The possible-worlds-semantics axiomatization is used to reason about whether the proposed plan is actually going to work. If the action summaries are well designed, stating the effects that the action is expected to have in the most common situations in which it is performed, the planner will produce correct

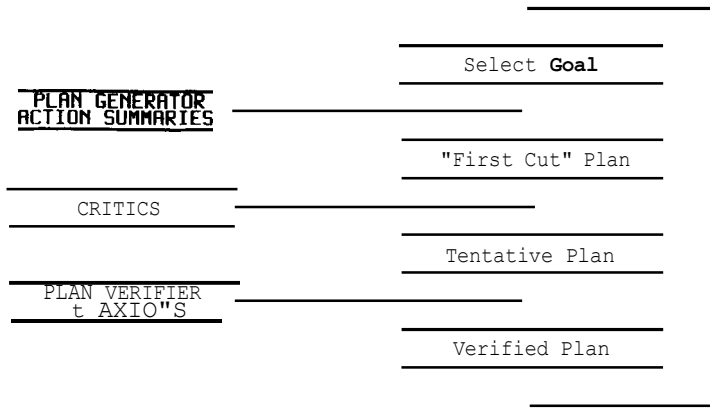


FIG. 3. The operation of KAMP.

plans most of the time, and the search required for finding a correct plan will be significantly reduced.

The search for a plan is facilitated by the simplifications introduced by the action summaries. For example, an implicit assumption in the action summaries is that all agents know what the effects of the actions are. In some instances this assumption may not hold, and any plan that depends on this assumption will fail the verification step. The process that uses the action summaries can be viewed as a 'plausible move generator' that proposes actions that are likely to succeed in achieving the goal.

KAMP uses a *procedural network* to represent the plan as it is being constructed and refined. A procedural network can be thought of as a two-dimensional data structure. The horizontal dimension is a temporal one, reflecting the partial ordering among the actions; the vertical dimension is one of abstraction, where goals and abstract actions are refined into sequences of low-level actions that can be performed directly by the agent. The connection between the planning data structure and the possible-worlds-semantics formalism is made by associating with each node in the procedural network a world that represents the state of affairs at that point. Whenever a fact must be proved to hold in the situation resulting from the execution of a series of actions, it is proved using the world associated with the appropriate node in the procedural network as the current real world.

KAMP's database contains assertions about what each agent knows, and what each agent knows what the other agents know. KAMP is not actually one of the agents doing the planning, but rather simulates how the agents would plan, given certain information about them. When KAMP plans, it 'identifies' with one of the agents and makes plans from the perspective of the agent it identifies with. This perspective makes an important difference when the planner con-

siders the intentions of other agents. Assuming that an agent A_1 doing the planning has a particular goal to achieve, it is possible for the planner to assume that A_1 will intend to do any action that A_1 knows will contribute to achieving the goal. However, if it is necessary to incorporate the actions of another agent, A_2 , into the plan, A_1 must be able to assume that A_2 will actually do the actions required of him. This amounts to showing that A_2 intends to do the action. Guaranteeing that this condition holds can lead to planning requests and commands. Once it has been established that A_2 will intend to do a high-level action, then the planner assumes that A_2 will intend to do any action that he knows will contribute toward the realization of the high-level action. Although A_2 may not have the knowledge necessary to carry out the action, it can be assumed that A_2 will execute a plan he can figure out. A_2 can ask questions of A_1 , however, if A_1 can anticipate this need for information and furnish it at the time of the request, the overall plan may be simplified, and the resulting dialogue will be more natural.

When the planner is given an initial goal, it first creates a procedural network consisting of a single plan step containing the goal. The following process is then executed repeatedly until either the planner concludes that the goal is unachievable, or some sequence of executable (i.e. low-level) actions is found that achieves the goal: First, a world (serving in its role as a situation) is assigned to each node in the procedural net reflecting the state of the world at that time (i.e. at the time *before* the action or goal named in the node is performed or achieved). The initial node is assigned W_0 , the initial actual world. Then, iteratively, when the planner proposes that a subsequent action is performed in a world to reach a new world, a name is generated for the new world and the relation of the new world to its predecessor is explicitly asserted. All goal nodes that have worlds assigned are then evaluated, i.e. the planner calls on the deduction system to attempt to prove that the goal is satisfied in the assigned world. Any goal for which the proof succeeds is marked as a phantom goal, (i.e. a goal that is already satisfied), but is kept in the plan so that if actions planned at a later stage should make it no longer hold, corrective action can be taken to preserve or re-achieve it.

Next, all the unexpanded nodes in the network that have been assigned worlds and that are not phantoms are examined. Some may be high-level actions for which a procedure exists to determine the appropriate expansion. These procedures are invoked if they exist, otherwise the action generator is invoked that uses the action summaries to propose a set of actions that might be performed to achieve the goal. If an action is found, it is inserted into the procedural network along with its preconditions.

Like Sacerdoti's system, KAMP uses procedures called *critics* to examine the plan globally and to determine interactions between proposed actions. A critic is a modular procedure that examines the plan for specific kinds of interactions between the effects of the actions. At the end of each cycle of expansion, each

critic looks for a particular type of interaction. If the interactions occur, the critic reorganizes the structure of the plan in some way.

There is an important distinction between modifications to the plan made by critics and modifications made during the process of expanding an action to a lower level of abstraction. The process of expansion is local to an action and concerned with determining what actions can be used to achieve a given goal. The process considers only the state of the world as it is assumed to be at the time of performing an action, and what actions are available. Critics examine interactions between actions in the plan and propose changes, but do not propose new plans.

The result of separating expansion and criticism is an overall simplification of the planning process. The process of expanding actions is simpler because the many possible interactions do not have to be considered at the time of expansion. Obtaining a rough plan and refining it reduces the amount of blind search the planner has to do. The process of discovering interactions is also simpler because it does not have to be concerned with what actions to perform, only with the interactions between actions that have already been selected.

If the expansion of the plan to the next level of abstraction is complete, then the planner invokes the deduction system to prove that the proposed sequence of actions actually achieves the goal. If the proof is successful and the plan is not yet fully expanded to the level of executable actions, the process of world assignment is carried out again and the entire procedure is repeated.

If the proof fails, the reason is probably that the simplifying assumptions made by the action summaries were incorrect; in this case, the planner must resort to a more detailed search of the space to find a plan. Finding the best strategy for plan modification when the correctness proof fails is a good area for future research.

3.5. Planning concept activation actions

As described so far, KAMP has a general ability to plan actions that affect the knowledge and intentions of agents; however for KAMP to produce language, it needs axioms and critics that capture information about linguistic actions.

Fig. 4 depicts the hierarchy of linguistic actions used by KAMP. At the top of the hierarchy are such illocutionary acts as informing and requesting. These are the highest-level communicative actions. A correctly performed illocutionary act has the effect of making the speaker and hearer mutually aware that the speaker intended to perform the illocutionary act. For example, if the speaker performs the illocutionary action of informing the hearer that the box-end wrench is in the toolbox, then (as a result of the action) they mutually know that the speaker intended to do so. From this knowledge and from other knowledge that the hearer has about the speaker (e.g. whether or not he is willing to believe something the speaker says) and knowledge about the

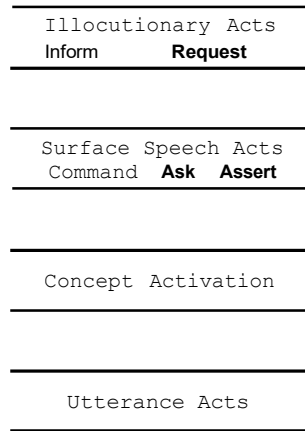


FIG. 4. KAMP's hierarchy of linguistic actions.

conventions of communication (e.g. the Gricean maxims [21]), he may come to believe that a particular wrench is in a particular location, thus realizing the *perlocutionary effects* of the action. A detailed description of the axioms for illocutionary acts and their relationship to perlocutionary effects is given elsewhere [6] and is beyond the scope of this paper.

The next level of abstraction is that of surface speech acts. Performing a surface speech act entails uttering a particular sentence. The distinction between these two levels of actions can be described as follows: Saying *that* the box-end wrench is on the table is an illocutionary act. A surface speech act realizes the illocutionary act by a *particular* utterance, in this case "The box-end wrench is on the table" (or perhaps, "The green tool is next to the platform", if the concept activation actions are realized differently). There is a one-to-one correspondence between surface speech acts and utterances, because the former are merely abstract representations of the latter.

It is impossible to state simple axioms that describe the effects of surface speech acts in the same manner as has been done for illocutionary acts for two reasons: The same surface speech act can realize different illocutionary acts, depending on the context in which it is performed, and it is possible for a single surface speech act to realize several illocutionary acts simultaneously. Because the effects of a surface speech act depend on what illocutionary act it is being used to realize, it is impossible to describe its effects directly. The axioms need to state (1) that the hearer realizes what surface speech act has been performed, and (2) how the hearer deduces the illocutionary act from the surface speech act.

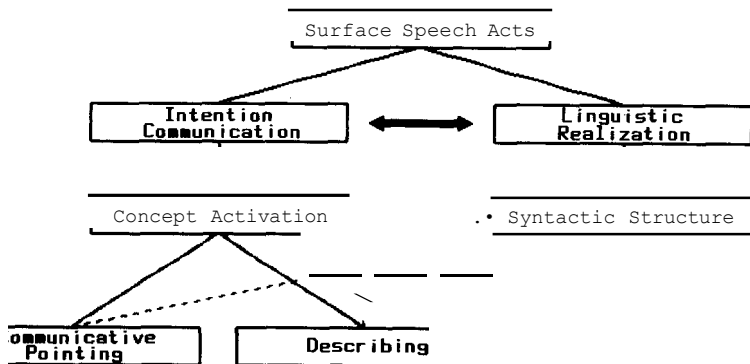
The relationship between surface speech acts and illocutionary acts can be quite complex, although frequently the connection is quite straightforward. For

example, a speaker can perform the surface speech act of uttering a declarative sentence with propositional content *P* and intend that the hearer recognize the intention to inform him that *P*. Such speech acts are called *direct speech acts*. In some cases, the inferential connection between the surface speech act and the intended illocutionary act is more complex, for example, when a speaker makes an utterance like "The door, please" and intends that the hearer recognize a request to open the door. Such actions are referred to as *indirect speech acts*, and Allen and Perrault [22] present a detailed analysis of how such intention recognition takes place.

KAMP does not currently plan surface speech acts that require the hearer to make indirect interpretations, not because it is inherently incapable of doing so, but rather because indirect speech acts are frequently planned to satisfy multiple goals, often along social dimensions and others that are very difficult to formalize. However, the ability to plan indirect speech acts is important for the generation of plausible utterances, and the incorporation of Allen and Perrault's intention-recognition conditions into the axioms for surface speech acts is an important area for further investigation.

Fig. 5 illustrates the two components of a surface speech act: the *intention communication* component and the *linguistic realization* component. The intention communication component is concerned with how the intentions of the speaker get communicated to the hearer. This includes communicating the intention to refer to objects or to get the hearer to identify objects. The linguistic realization component is concerned with taking the actions specified by the intention communication component and realizing them in a valid syntactic structure. A two-way channel of communication between these two components exists, because the means of communication of intention determines what linguistic structures must be chosen; in addition, the grammatical choices available constrain the possible means of intention communication.

Associated with each surface speech act node is the syntactic structure of the



Flo. 5. Components of a surface speech act.

sentence that constitutes its linguistic realization. Initially, these structures are only partially specified; as the plan is expanded to lower levels, the actions in the expansion contribute to the syntactic structure of the sentence associated with the surface speech act of which they are a part. When the plan is complete, each surface speech act node specifies a sentence, the utterance of which demonstrably satisfies the speaker's goals.

KAMP's grammar originally consisted of a relatively small number of context-free rules distributed throughout the system. A grammar of basic clause structure was available to the planner when surface speech acts were planned, in expanding surface speech acts with concept activation actions, a grammar of noun phrase constituent structure was used. Obviously, the linguistic coverage was severely limited. This was not viewed as a serious limitation, however, because the purpose of KAMP was originally to explore and develop some of the most basic relationships between utterances and their relationship to communicative intentions, and the interesting problems concerned how utterances were used in multiple-goal satisfaction and integrated into a general plan of action rather than in the internal complexity of the utterances themselves.

The inadequacy of the initial approach led to the development of the TELEGRAM grammar formalism. Currently work is in progress on constructing a larger, more robust grammar based on function unification grammar [27] that provides a cleaner integration of the grammar with the planning mechanism. This work is described elsewhere [23] and discussion of the details of the grammar is beyond the scope of this article.

3.6. A formal theory of concept activation

The level of abstraction below surface speech acts in Fig. 4 is that of concept activation actions. Currently KAMP only plans concept activation actions for which the concept is part of the speaker's and hearer's mutual knowledge. Concept activations are part of the expansion of a surface speech act. We will consider only concept activation actions that have realizations as noun phrases. As discussed earlier, the types of actions planned by KAMP are only a subset of the types of actions that can be realized by noun phrases, but nevertheless represents a significant subset.

As explained in Section 2, the process of understanding a speech act involves first constructing a proposition from a set of available concepts. This set of available concepts are called 'active', and the predicate $Active(C)$ means that C is an object-language term that belongs to the available set of terms for constructing the proposition the speaker is conveying by way of the surface speech act. Terms can enter the active set explicitly through the performance of a concept activation with the term as its object, or implicitly by way of inference. For example, in requesting a hearer to remove the pump from the platform in a domain involving the repair and assembly of an air compressor, if

the hearer knows that the pump is attached to the platform and nothing else, the platform need not be mentioned because it is implicitly active. Therefore, it is only necessary to say "Remove the pump" for the hearer to recognize

Intend(S, Do(H, Remove(pumpl, platform!))).

The performance of a communicative pointing action directed toward an object causes the standard name of that object to enter the set of active concepts.

The actual process of recognizing the proposition from the active concepts is not axiomatized. As a simplification, the axioms for surface speech acts state as a precondition that the terms that are part of the propositional content of the act must be active.

Axioms (5)-(7) describe what happens when an agent performs a concept activation action (Cact). Axiom (5) describes the preconditions that the speaker and the hearer are at the same location and that the speaker intends that the hearer know that that concept is active. Axiom (6) describes the change in the set of active concepts and axiom (7) states that the speaker and hearer mutually know that the concept activation action has taken place by specifying that the only worlds that are consistent with what the two agents mutually know are those that are the result of performing the action in some world consistent with their mutual knowledge before the action. The mutual knowledge of the action's occurrence results in the mutual knowledge of the preconditions and effects holding in the appropriate states, and thereby provides a means of stating the knowledge effects of actions without the necessity of listing every consequence explicitly. The axioms are expressed in the notation of Moore [12]. The function @ constructs a rigid designator that denotes its argument in all possible worlds, meaning that it functions like an inverse QUOTE for embedding meta-language terms inside of object-language formulas.

$$\begin{aligned} & \forall A, B, C, w_1, w_2 R(Do(A, Cact(B, C)), w_1, w_2) :: J \\ & D(w_1, Location(A)) = D(w_1, Location(B)) \text{ \textcircled{1}} \\ & T(w_1, \mathbf{Intend}(@\mathbf{A}), \mathbf{MutuallyKnow}(@\mathbf{A}), @\mathbf{B}), \mathbf{Active}(@\mathbf{C})), \end{aligned} \quad (5)$$

$$\forall A, B, C, w_1, w_2 R(Do(A, Cact(B, C)), w_1, w_2) :: J T(w_2, \mathbf{Active}(C)), \quad (6)$$

$$\begin{aligned} & \forall A, B, C, w_1, w_2 R(Do(A, Cact(B, C)), w_1, w_2) :: J \\ & \forall w_3 K(\mathbf{Kernel}(A, \mathbf{B}), w_2, w_3) :: J \\ & \exists w_4 K(\mathbf{Kernel}(A, \mathbf{B}), w_1, w_4) \text{ \textcircled{1}} R(Do(A, Cact(B, C)), w_4, w_3). \end{aligned} \quad (7)$$

When KAMP is expanding a concept activation action to lower-level actions, it takes into account both the intention communication and linguistic realization components of the action. The intention communication component may be

realized by a plan involving either physical or linguistic actions. KAMP relies on *description* as a linguistic means of communicating an intention to refer, and on *pointing* as a nonlinguistic means.

The following schema defines the preconditions of the *describe* action in a manner similar to axiom (5):

$$\begin{aligned}
 & \forall A, B, w_1, w_2 R(\text{Do}(A, \text{Describe}(B, 9fl)), w_1, w_2) :J \\
 & \quad 3x \ T(w_1, \text{Know}(A, 9fl(@x))) \ \wedge \\
 & \quad \quad T(w_1, \sim \text{MutuallyKnow}(A, B, \sim 9fl*(@x))) \ \wedge \\
 & \quad \quad T(w_1, \forall y \sim \text{MutuallyKnow}(A, B, *(y)) \Rightarrow y, i@x). \quad (8)
 \end{aligned}$$

Axiom (8) states the preconditions for an agent A planning to use a description $9fl$ to refer to an object to an agent B . The description is an object-language predicate composed of a number of individual descriptors. It is defined as

$$9fl = \text{defAx}(D_1(x) \ \wedge \ \dots \ \wedge \ D_n(x)),$$

where the $D_i(x)$ are the individual descriptors that comprise the description. The symbol $9fl^*$ denotes a similar expression that includes all the descriptors of $9fl$ composed with the predicates that define the center of the discourse. These predicates restrict the possible individuals to which the description can apply to be only those that are relevant in the current discourse context. The general idea is that a single concept is identified as the *center* (Ch) of the discourse. As a simple approximation to a correct theory,

$$9fl^* = \text{defAy}((9fl(Cb) \ \wedge \ Y = Cb) \vee (\sim 9fl(Cb) \ \wedge \ 9fl(y))).$$

This axiom says that if a description is true of the center, then $*$ applies only to the center, otherwise, $9fl^*$ applies to the other objects, if any, of which $9fl$ is true. With the addition of axioms to describe how Cb moves as the discourse progresses, KAMP can plan referring expressions that take centering into account. A full discussion of centering is beyond the scope of this article. A more detailed account of the role of centering in KAMP's description planning strategy is discussed by Appelt (6).

The first clause in the conclusion of axiom (8) states that the speaker must believe the description that he attributes to the intended referent actually holds. The second and third clauses state that the intended referent is the only one that is not ruled out according to the mutual knowledge of the two agents. The reason for this indirect approach is that one must allow for the case in which a speaker plans a description that serves to both identify a referent and inform the hearer of properties of that referent. In that case, descriptors $D_1(x)$ through $D_i(x)$ (called the basic descriptors) will be mutually known to identify a single object, while descriptors $D_{i+1}(x)$ through $D_n(x)$ will be believed only by the speaker. Therefore, when the hearer interprets the speaker's utterance, the

description *£ig* is not known to apply to anything at all. However, if the speaker planned the description so that the basic descriptors are mutually known to pick out a single referent in context, then there will be only one object that is not *ruled out* by the description. Because the hearer knows that the speaker believes the *entire description* he then can then decide to believe the information conveyed by the additional descriptors.

KAMP chooses a set of basic descriptors when planning a describe action to minimize both the number of descriptors chosen, and the amount of effort required to plan the description. Choosing a provably minimal description requires an inordinate amount of effort and contributes nothing to the success of the action. KAMP chooses a set of descriptors by first choosing a *basic category* descriptor (see [26]) for the intended concept, and then adding descriptors from those facts about the object that are mutually known by the speaker and the hearer, subject to the constraint that they are all linguistically realizable in the current noun phrase, until the concept has been uniquely identified.

Some psychological evidence suggests the validity of the minimal description strategy; however, one does not have to examine very many dialogues to find counter-examples to the hypothesis that people always produce minimal descriptions. According to the language-generation theory embodied in KAMP, people do choose minimal descriptions for concept activations; however these descriptions can be augmented for a variety of reasons, (e.g., to realize additional informing actions, as in the example in the next section, or to make it easier for a speaker to identify an object when an identification is requested, see [28]).

The other action that can be used to expand a concept activation action is *communicative pointing*. The following axiom describes pointing as one way of performing a concept activation action, which directly activates the concept of the object of the pointing action.

$$\forall A, B, w_1, w_2 R(\text{Do}(A, \text{Point}(B, X), w_1, w_2)) :J \\ T(w_2, \text{MutuallyKnow}(A, B, \text{Active}(@X))). \quad (9)$$

Axiom (9) says that if an agent performs a pointing action, the standard name of the object he is pointing at becomes active.

A problem with pointing actions, not dealt with here, is how it is possible to decide whether such an action has communicative intent. It is a convention of language use that utterances almost universally have communicative intent (the exceptions being actions like muttering to oneself). However, a particular physical action may or may not have communicative intent, and KAMP does not attempt to describe how a particular perceived gesture is interpreted as communicative pointing.

Axioms (5)-(7) work together with (8) and (9) to produce the desired communicative effects. When a speaker utters a description, or points, he

communicates his intention to activate a concept, provided that the action has a suitable corresponding linguistic realization in the surface speech act.

3.7. Satisfying multiple goals in a referring expression

KAMP attempts to take advantage of opportunities to achieve multiple goals in a single utterance by recognizing situations in which *action subsumption* is possible. An important application of this principle is in the planning of referring expressions.

An action A_1 *subsumes* an action A_2 if A_1 and A_2 are part of the same plan, and action A_1 (in addition to producing the effects for which it was planned), also produces the effects for which A_2 was intended. Therefore, the resulting plan need only include action A_1 (and its expansion) to achieve all the goals.

The concept of action subsumption is particularly useful for planning linguistic actions because many options are typically available for expanding an illocutionary act into a surface utterance. Frequently, the planner can detect situations in which minor alterations to the expansion of an action will allow an action in another part of the plan to be subsumed. Although the term 'minor alterations' is somewhat vague, the general idea is clear. When planning surface speech acts, it means making a change localized to only one of the constituents of the sentence. Changes can be made to a surface speech act during the planning that do not alter the overall structure of the sentence, but are sufficient to subsume other actions in the plan. An example of such a change that is relevant to this article is adding a descriptor to a referring expression.

Axiom (8) provides the justification subsuming informing actions by concept activation actions. For efficient planning, a good strategy is needed to recognize when action subsumption is possible. Action subsumption is an example of a global interaction between actions in a plan. Such interactions are detected by the *critics* discussed in Section 3.4. The action-subsumption critic works by first applying a set of rules to see if action subsumption may be possible. If so, it then tries several action-subsumption strategies that specify the exact modification to the plan that must be made. If the strategy is successful, then the plan is altered, and the subsumed action marked so that no further work is done by the planner to expand it.

An example of the action-subsumption test rules would be "look for a situation in which a concept activation of a concept C is being performed, and where in the same plan there is a goal that the hearer know some property holds of C." The planner attempts to (1) expand the concept activation with a describe action, and (2) incorporate the property as one of the descriptors in the describe action. Axiom (8) enables KAMP to verify the correctness of the plan modifications proposed by the critics.

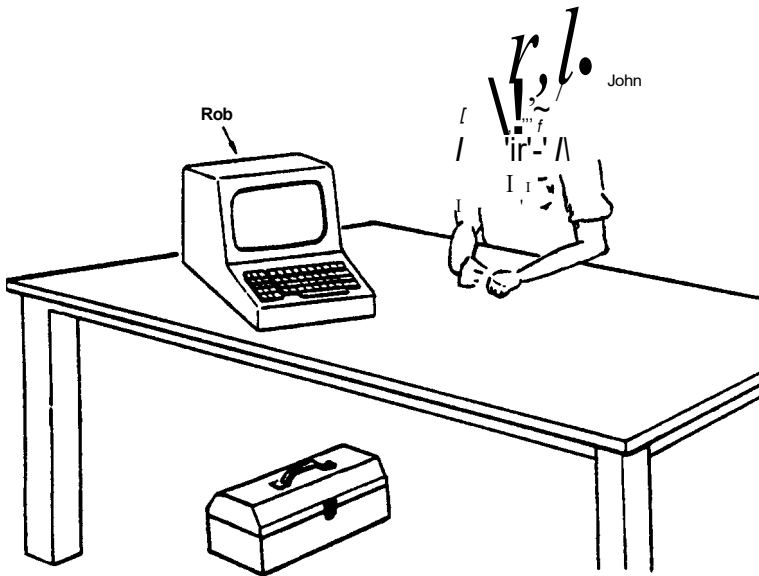
This type of reasoning is what enables the speaker in the example of Fig. 1 to

conclude that he is informing the hearer that the object he is pointing to is a wheelpuller. The example example presented in the next section illustrates how KAMP plans multi-purpose referring expressions.

4. An Example of Planning Referring Expressions

KAMP's initial domain is the information required by an expert system that knows about the assembly and repair of a particular piece of equipment, and in which the user is a novice seeking assistance. There are two reasons for choosing this particular domain: First, dialogue protocols have been collected [29] that provide a body of linguistic data raising interesting issues and examples of phenomena that can be explained by the theory on which KAMP is based. Second, the domain provides an ideal situation for multiple-agent planning in which communicative actions arise naturally.

Fig. 6 illustrates a typical situation in which KAMP operates. This domain has two agents called Rob and John. Rob is a robot that incorporates KAMP for planning and deduction. Rob's only connection with the world is the computer terminal; thus, he is capable of performing speech acts, but no actions that directly affect the physical state of the world. John is assumed to be a person who is capable of performing both speech acts and physical actions. The particular situation for this example includes a piece of equipment to be repaired (in this case, an air compressor) and tools that are necessary for the task. The tools can be out in plain view on the table (in which case Rob and John mutually know properties such as their color, shape, size and location) or



Flo. 6. KAMP's domain.

they can be stored away out of sight in the toolbox (in which case Rob may know where they are, but not necessarily John). In general, Rob is the expert and he knows almost everything about the situation. For example, Rob knows how to assemble the compressor; specifically, he knows how the parts fit together, what tools are needed for the various assembly operations, and where the tools are located.

This domain provides an ideal setting for studying multiple-agent planning as it relates to the production of utterances. Communication arises naturally in this domain because of the difference in knowledge and capabilities of the agents. Because Rob is incapable of performing physical actions, he must make requests of John whenever he intends to change the physical state of the world. Because Rob knows all there is to know about the task (and John knows this) John must ask questions to get the information he knows he needs when he requests John to do something. Therefore, the need for communication arises in order for either agent to satisfy his goals.

The example presented in this section does not touch on every issue that arises in utterance planning, however, it touches on enough of them to give the reader a sense of how the different components of the KAMP planner fit together. The following notation is used for the illustrations in this section: Each node in the plan has some sort of boldface label (**P1**, **P2**, ... , **Pn**) to make it easier to refer to. Dotted boxes are used to represent phantom goals. The successor relation between actions is represented by solid connecting lines and hierarchical relationships by dotted lines. Each node has an associated world. For goal nodes, the world is written inside parentheses (e.g. (*W*)), to represent that the planner is to start in world *W*; and find some actions to reach a world in which the goal is satisfied. For phantom nodes, the world name is not in parentheses to indicate the goal is actually satisfied within the indicated world. Action nodes have a label like '*W*; ' to indicate that the action is a transformation relating worlds *W*; and - Actions will often be planned without knowing precisely what worlds they will be performed in, or precisely what world will be the result of the action. This is particularly true of actions that are represented at a high level of abstraction. Worlds are represented in the diagram as '?' if the planner has not yet assigned a definite world. (Note that KAMP can often reason about what is true at a given point in the plan, even though it has not assigned a world to the node, because frame axioms can be stated for high-level actions that describe some changes and leave others unspecified.) A notation like '*W*₁ ?' is assigned to a high-level action that may be expanded to several actions at a lower level. The planner knows the action sequence will begin in *W*₁, but it will not **know** the resulting world until the action is expanded. A notation such as '? ?' is used when the planner knows where in a sequence a high-level action must fall in relation to other actions in the plan, but cannot assign either an initial or final world.

KAMP requires a fairly rich description of its domain to plan communication

acts. KAMP needs knowledge in five areas: Basic common-sense knowledge (e.g., wrenches are tools, a compressor pump can only be attached to one thing at a time), basic knowledge about the objects in the domain (e.g., there is a wrench, it has an end-type of box-end, it is located in the toolbox), knowledge and mutual knowledge of agents in the domain (e.g., Rob knows the box-end wrench is located in the tool box, John does not know where the wrench is, Rob and John mutually know the pump is located on top of the table, it is universally known that all agents know their own location at all times), and descriptions of actions and their physical and knowledge-state effects, (e.g., if an agent performs an unfastening action, then he knows that he has just performed it, and the two objects that were fastened together are now now longer connected, and the agent knows this), and the basic axioms about knowledge and communication actions discussed earlier.

In the example discussed here the agents are Rob and John, the domain objects are a pump, PU, and platform, PL, mutually known to be on a table, T1, an object WRI mutually known to be a box-end wrench, TBI mutually known to be a toolbox and mutually known to be located under the table. The pump is mutually known to be fastened to the platform by a bolt, B1. Rob and John are initially in the same location. Because they are always in close proximity, they will always mutually know each other's location as well as their own. It is explicitly stated that John does not know what tool to use for unfastening B1, and that John does not know the location of WRI. Rob begins with the initial goal that the pump be removed from the platform:

True(~ Attached(PU, PL)).

Given the above goal, KAMP begins planning as described in Section 3.4 by creating a procedural network and refining the plan to successively lower levels of abstraction. Refinement to the first level of abstraction results in the plan shown in Fig. 7, nodes **P1 to P6**. KAMP has decided that Rob should request that John perform the action of removing the pump because according to the action summaries, a request is the only possible action one agent can perform to affect another's intentions. Whenever complete expansion to one level of abstraction is completed, KAMP uses the axioms to prove that the plan proposed according to the action summaries is successful.

Because the request is the only illocutionary act that has been planned so far, there is no more linguistic planning to be done at this stage. KAMP now turns its attention to expanding the REMOVE action. KAMP's axioms for remove (included as part of the general description of the domain) specify that removing is performed by unfastening any fasteners that are connecting one object to the other. This leads KAMP to include unfastening bolt B1 using the appropriate tool as part of the plan he must execute to remove the pump. This leads to the plan illustrated in Fig. 8. The preconditions for John performing this action are represented by nodes **PIO, PII**, and P12-that John knows what the right tool

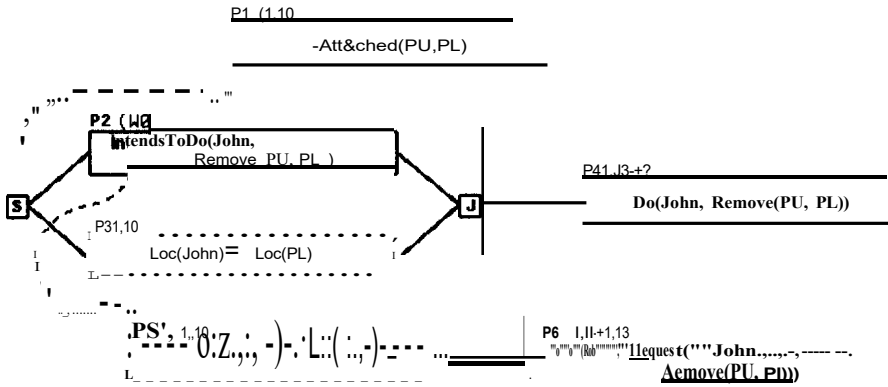


FIG. 7. Rob requests that John remove the pump.

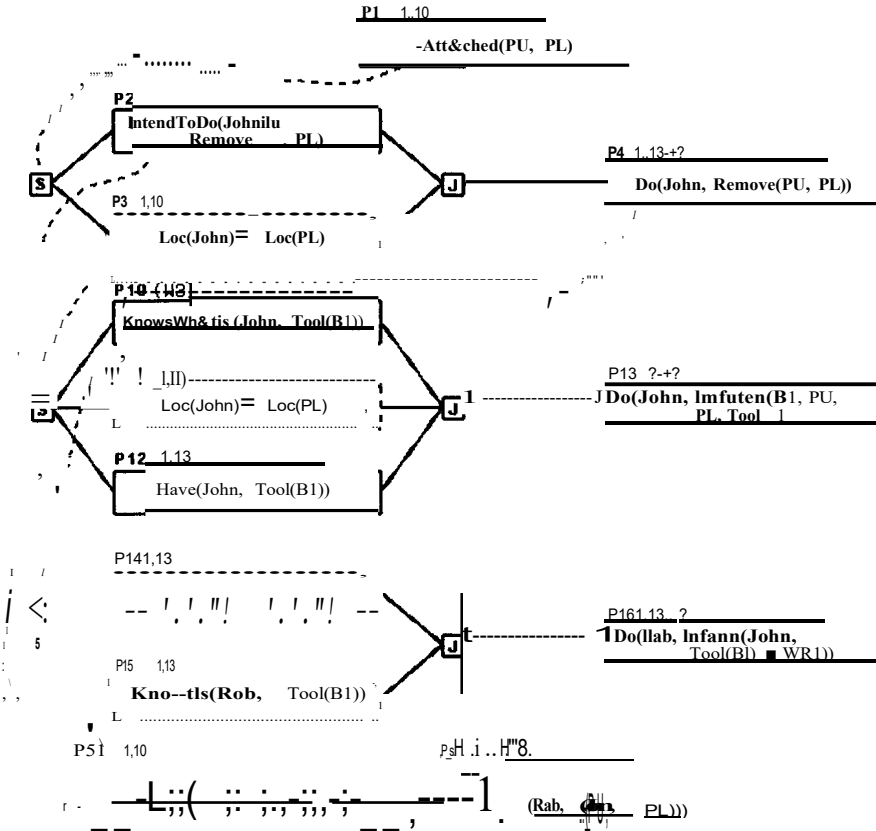


FIG. 8. Rob plans for John to remove the pump.

is, that John is in the same place as the platform, and that John has the tool. Because John is already assumed to be in the same location as PL, the location goal, **P11** is a phantom (i.e. it is already true, but must be considered by the planner, in the event that some other action causes it to no longer hold). Rob does not know whether John has the tool, or even that John knows where the tool is located; therefore, KAMP plans for Rob to inform John that the tool for removing bolt B1 is wrench W1 (Node **P16**).

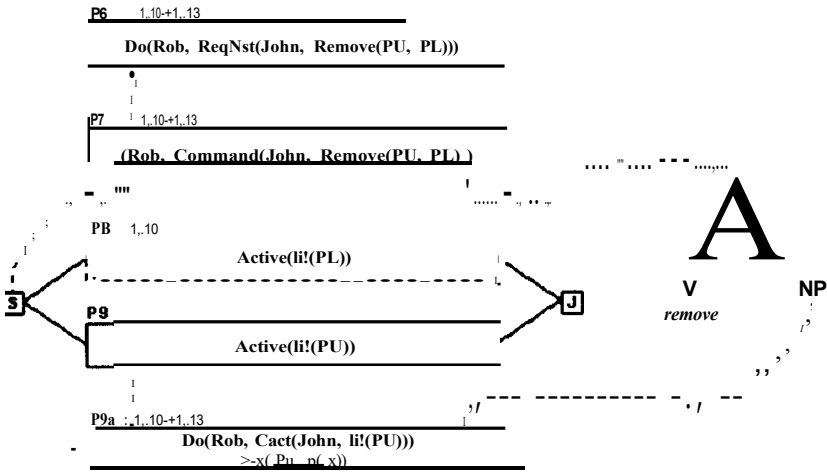
In summary, what KAMP has done so far is formulate a plan in which John adopts Rob's goal of removing the compressor pump from its platform, and through additional planning, has discovered some information that John needs to know to carry out the plan, and has planned to provide him with this information by means of an additional informing action.

Next, the planner must expand illocutionary acts in the plan to surface speech acts. This step may require some complex reasoning about when a surface speech act **will** be recognized as a particular kind of illocutionary act, for example, when a question will be understood as a request to perform a physical action as opposed to a request for information. For reasons explained in Section 3.5, surface speech acts are assumed to be interpreted directly, so the utterance of an imperative sentence is planned to realize the request in **P6**.

KAMP reasons that enough information must be included in the utterance so the hearer will recognize the intended proposition Remove(PU, PL). This entails conveying the predicate 'Remove' and activating concepts@(PU) and @(PL). As described in Section 3.6, concepts can be activated by being inferentially related to active ones as well as being activated directly. In this problem, axioms are included that state that Rob and John both know that the platform is the only object to which the pump is attached. Therefore, any action that removes the pump must remove it from the platform, and it is only necessary to say "Remove the pump" to have the hearer recognize the entire proposition.

Fig. 9 illustrates the interaction between the two components of surface speech-act planning-intention communication and linguistic realization. The intention communication component of this surface speech act consists of concept activation goals (nodes **P8** and **P9**) for each of the concepts mentioned in the intended proposition. Because KAMP has reasoned that it does not need to mention the platform, node **P8** is marked as a phantom. The phantom action will most likely not be reflected in the final utterance, but can be noticed by critics and later reactivated if the critic decides that it could satisfy another goal by referring to the platform with an appropriate description.

The linguistic realization component consists of choosing a basic syntactic structure for the sentence and relating it to the actions of the intention communication component. According to its grammar, the planner knows that an imperative sentence has the structure "V NP (PP)*" and associates such a structure with the surface speech-act node **P7**.



F10. 9. Expanding a surface speech act.

KAMP now turns attention to expanding the goal node **P9**, activating the concept@(PU). Intention communication in this case is very simple, because (according to the initial axiomatization of the domain) there is only one object that is mutually believed by the speaker and hearer to be a pump. Therefore, the concept activation action **P9a** is planned, and its subordinate describe action, choosing

.Ax(Pump(x))

as a description, as described in Section 3.6.

KAMP has now reached the point at which the criticism portion of the expansion-criticism cycle begins. As explained in Section 3.4, each critic has a simple test that it applies to the plan to see if it is applicable. The action-subsumption critic's test works by examining pairs of illocutionary acts, such as the informing action **P16** and the request **P6** (see Fig. 8) to see if they are connected in a way that permits action subsumption. Actions **P16** and **P6** are connected by the fact that the wrench referred to in **P16** is an instrument of the action requested by **P6**. Because the verb chosen for **P6**, remove, can take an instrument case, the critic realizes that the informing action **P16** can be subsumed by the request **P6**, provided that reference to the instrument is made explicitly in the utterance. Since the simple grammar doesn't constrain the number of adverbial PPs that appear in the sentence, the addition of a prepositional phrase is adopted as a subsumption strategy.

The action-subsumption critic must also determine whether all the pre-conditions for the subsumption candidate are also satisfied in the world when the subsuming action is going to be 'performed'. All the conditions, namely

that Rob is in the same location as John, and Rob knows that $Tool(B1) = WRI$, are satisfied in this situation. The critic adds concept activation **P16a** of Fig. 10 to the plan for intention communication, and adds the prepositional phrase with preposition *with* to the syntax tree. Once the addition is performed, then the planner reasons that the description $Ax(Wrench(x))$ is adequate to activate @ (**WRI**).

The subsumption of the informing action means that the hearer's knowledge will have changed by the time he executes the action of removing the pump. Because the exact effect of this additional knowledge on the plan is difficult to determine, the entire expansion of node **P4** is discarded and replanned. Fig. 10 shows the procedural net after criticism by the action-subsumption critic.

KAMP now turns its attention to goal **P12**, that John has wrench WRI in his possession. For John to have the wrench, he must know where it is, and he must go there and get it. According to our model, John does not have this knowledge, but Rob does; accordingly KAMP plans for Rob to perform an additional informing action (**P24** in Fig. 11) to tell John the wrench's location.

In the next criticism cycle, the action-subsumption critic finds a situation analogous to the one with informing action **P16**. Action **P17** is a candidate for subsumption by the request because it informs the hearer of a property of one of the case arguments of the main verb being planned for the request. As in the previous case, the informing action is relocated so that it follows the request, and the part of the plan that may be affected by the hearer's new knowledge is discarded and replanned, as before. The description used in the concept-activation action **P16a** is then augmented with the new descriptor to yield

$$Ax[(Wrench(x) \wedge (Loc(x) = Loc(TB1)))]$$

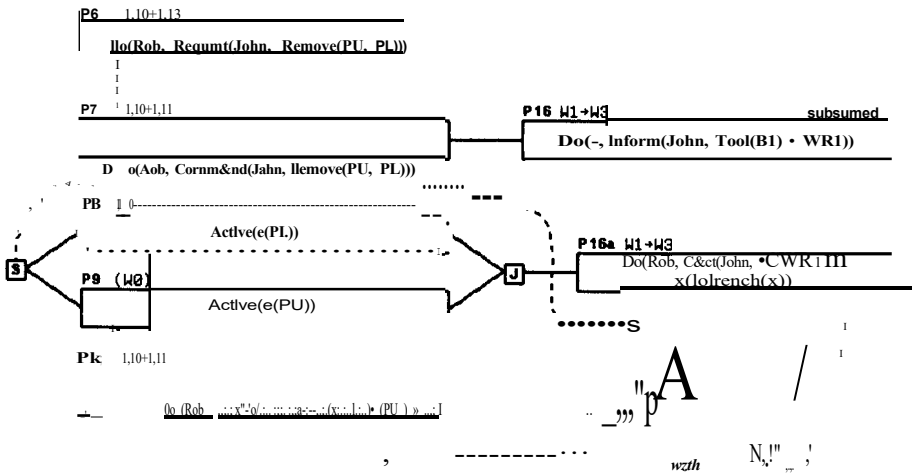


FIG. 10. Subsuming the informing action.

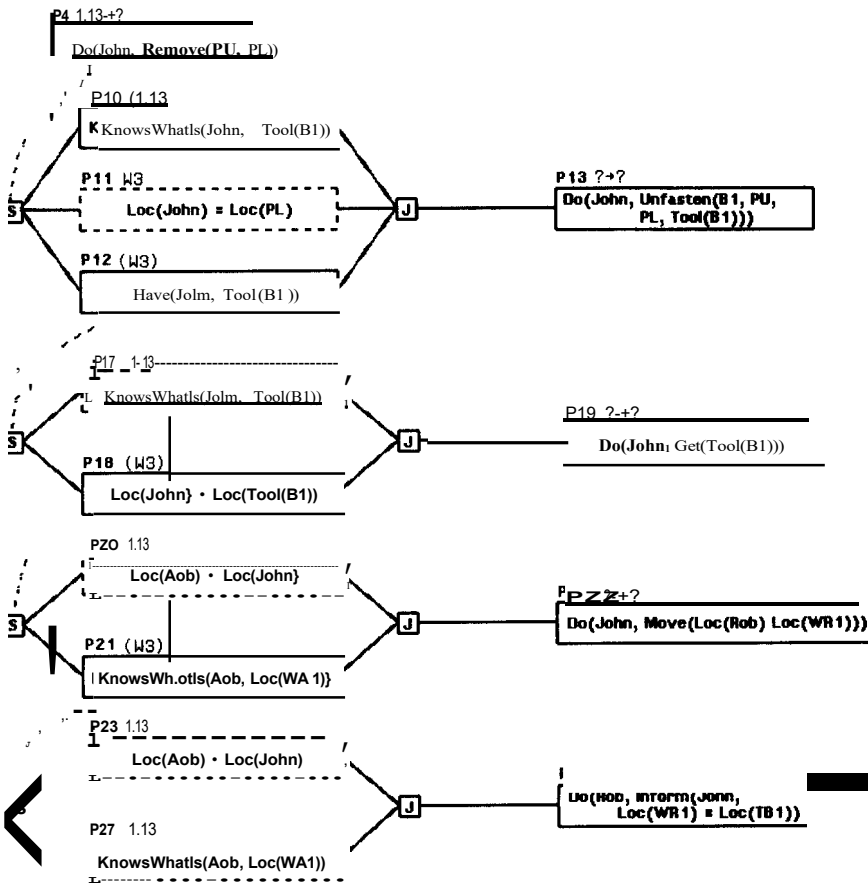


FIG. 11. Rob plans to inform John of the wrench's location.

and modifying the linguistic realization, shown in Fig. 12 to attach a prepositional phrase to the noun phrase referring to the wrench.

The plan is completed when the planner plans a concept activation of TB 1, using the description $h(\text{Tool-box}(x))$. In the completed plan, Rob says to John "Remove the pump with the wrench in the toolbox." Before the plan is actually adopted and executed, KAMP verifies that it will actually work, since the action summaries employ simplifications that may in some cases lead to the formulation of an incorrect plan. In this case the verification is possible because according to axiom (8), the hearer acquires the knowledge through the describe actions associated with concept activation **P16a** that would have been provided by inform actions **P16** and **P24** had they not been subsumed.

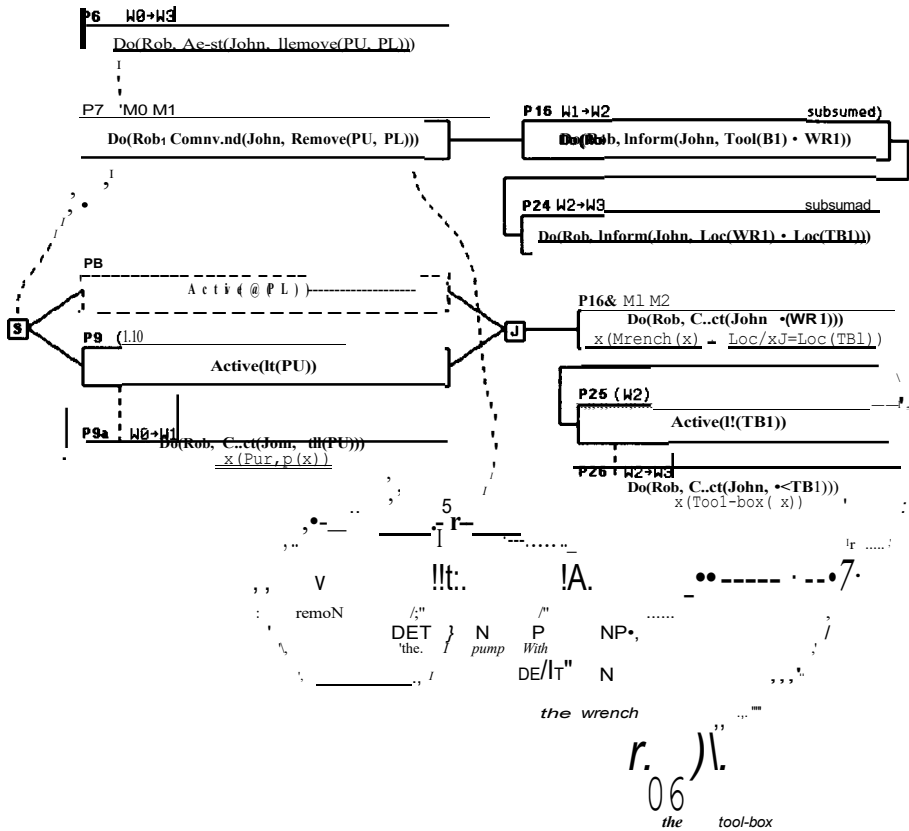


FIG. 12. Incorporating the wrench's location into the referring expression.

5. Conclusion

This research has focused on how speakers plan referring expressions that can be coordinated with physical actions and that may satisfy multiple goals. Producing such utterances given only a description of a speaker's goals is not a simple process; it requires a powerful system that is capable of general reasoning about agents' beliefs and intentions. It is difficult to envision any alternative to utterance planning that will account for the wide range of behavior observed in human communication.

The KAMP system is a useful vehicle for the investigation of a theory of language generation based on planning. Adapting KAMP from a general-purpose hierarchical planner to a language planner involved axiomatizing the various linguistic actions (illocutionary acts, surface speech acts, describing, pointing and concept activation) in terms of the possible-worlds formalism, incorporat-

ing these axioms and action summaries into KAMP and designing plan critics that focus on interactions typical of linguistic actions. The result of incorporating these capabilities into KAMP is a system capable of producing English sentences with complex referring expressions as part of an agent's plan.

Several important research issues in planning referring expressions have been raised by the work done on KAMP, but have received only cursory examination to date. The linguistic coverage of the grammar needs to be extended considerably, and the work on TELEGRAM has been an important step in that direction. As has been cited earlier, there are a number of different purposes for which speakers use noun phrases. Not only are the concept activation actions examined in depth in this article realized by noun phrases, but also identification requests, attributive uses of definite descriptions, and a variety of other phenomena. It will be an important test of the theory to see if these other actions can be axiomatized and easily accommodated within this general framework.

KAMP has proven to be a useful tool for the investigation of planning referring expressions and utterances in general, and promises to be useful in developing a speech-act theory to account for many aspects of natural language use.

ACKNOWLEDGMENT

This research was supported in part by the Office of Naval Research under contract N0014-80-C-0296 and in part by the National Science Foundation under grant MCS-8115105. The author is grateful to the AI Journal referees for pointing out the shortcomings of the first draft. The author is particularly indebted to Barbara Grosz for providing thorough comments on several drafts of this article.

REFERENCES

1. Sacerdoti, E. *A Structure for Plans and Behavior* (North-Holland, Amsterdam, 1977).
2. Mann, W.C. and Matthiessen, C., *Nigel: A systemic grammar for text generation*, University of Southern California Information Sciences Institute Tech. Rept. ISI/RR-83-105, 1983.
3. McKeown, K., *Generating natural language text in response to questions about data base structure*, Ph.D. Dissertation, University of Pennsylvania, Philadelphia, PA, 1982.
4. Bruce, B.C., *Belief systems and language understanding*, BBN Technical Rept. No. 2973, Cambridge, MA, 1975.
5. Cohen, P. and Perrault, C.R., *Elements of a plan based theory of speech acts*, *Cognitive Sci.* 3 (1979) 177-212.
6. Appelt, D., *Planning English Sentences* (Cambridge University Press, Cambridge, U.K., 1985).
7. Evans, D. *Situations and speech acts: Toward a formal semantics of discourse*, Ph.D. Dissertation, Department of Linguistics, Stanford University, Stanford, CA, 1981.
8. Barwise, J. and John, P., *Situations and Attitudes* (MIT Press, Cambridge, MA, 1983).
9. Appelt, D.E., *A planner for reasoning about knowledge and action*, in: *Proceedings National Conference on Artificial Intelligence* (1980) 131-133.
10. Appelt, D.E., *Planning natural-language referring expressions*, in: *Proceedings 20th Annual Meeting of the Association for Computational Linguistics*, (1982) 108--112.
11. Prince, E.F., *Toward a taxonomy of given-new information*, in: P. Cole (Ed.), *Radical Pragmatics* (Academic Press, New York, 1981).

12. Moore, R.C., Reasoning about knowledge and action, SRI International AI Center Tech. Note No. 191, Menlo Park, CA, 1980.
13. Searle, J., *Speech Acts: An Essay in the Philosophy of Language* (Cambridge University Press, Cambridge, U.K., 1969).
14. McCarthy, J. and Hayes, P., Some philosophical problems from the standpoint of artificial intelligence, in B. Meltzer and D. Michie (Eds.), *Machine Intelligence 4* (Edinburgh University Press, Edinburgh, 1969).
15. Clark, H. and Marshall, C., Definite reference and mutual knowledge, in: A.K. Joshi, I.A. Sag and B.L. Webber (Eds.), *Elements of Discourse Understanding* (Cambridge University Press, Cambridge University Press, Cambridge, U.K., 1981).
16. Cohen, P. and Perrault, C.R., It's for your own good: a note on inaccurate reference, in: A.K. Joshi, I.A. Sag and B.L. Webber (Eds.), *Elements of Discourse Understanding* (Cambridge University Press, Cambridge, U.K., 1981).
17. Nadathur, G. and Joshi, A.K., Mutual beliefs in conversational systems: their role in referring expressions, in: *Proceedings Eighth International Joint Conference on Artificial Intelligence*, Karlsruhe, West Germany (1983) 603--605.
18. Stickel, M.E., Theory resolution: building in nonequational theories, in: *Proceedings National Conference on Artificial Intelligence*, Washington, DC (1983) 391-397.
19. Bruce, B.C. and Newman, D., Interacting plans, *Cognitive Sci.* **2** (1978) 195-233.
20. Fikes, R.E., and Nilsson, N., STRIPS: a new approach to the application of theorem proving to problem solving, *Artificial Intelligence* **2** (1971).
21. Grice, H.P., Logic and conversation, in: Davidson (Ed.), *The Logic of Grammar* (Dickenson, Encino, CA, 1975).
22. Allen, F. and Perrault, C.R., Analyzing intention in utterances, *Artificial Intelligence* **15** (1980) 14>-178.
23. Appelt, D.E., TELEGRAM: A grammar formalism for language planning, in: *Proceedings Eighth International Conference on Artificial Intelligence*, Karlsruhe, West Germany (1983) 595-599.
24. Grosz, B.J., Joshi, A.K. and Weinstein S., Providing a unified account of definite noun phrases in discourse, in: *Proceedings 21st Annual Meeting of the Association of Computational Linguistics* (1983) 44-50.
25. Grosz, B.J., Focusing and description in natural language dialogs, in: A.K. Joshi, I.A. Sag and B.L. Webber (Eds.), *Elements of Discourse Understanding: Proceedings of a Workshop on Computational Aspects of Linguistic Structure and Discourse Setting* (Cambridge University Press, Cambridge, U.K., 1980) 84-105.
26. Rosch, E., Mervis, C., Gray, W., Johnson, D. and Boyes-Braem, P., Basic objects in natural categories, *Cognitive Psycho.* **8** (1976) 382-439.
27. Kay, M., Functional unification grammar: A formalism for machine translation, in: *Proceedings Tenth International Conference on Computational Linguistics* (1984) 75-78.
28. Cohen, P., The need for identification as a planned action, in: *Proceedings Seventh International Joint Conference on Artificial Intelligence*, Vancouver, BC (1981) 31-36.
29. Deutsch (Grosz), B., Typescripts of task-oriented dialogs, SRI International AI Center Tech. Note No. 146, Menlo Park, CA, 1975.

Received October 1982; revised version received October 1983, August 1984