PLANNING NATURAL LANGUAGE
REFERRING EXPRESSIONS

Douglas E. Appelt

SRI International
Menlo Park, California

ABSTRACT

This paper describes how a language-planning system can produce natural-language referring expressions that satisfy multiple goals. It describes a formal representation for reasoning about several agents' mutual knowledge using possible-worlds semantics and the general organization of a system that uses the formalism to reason about plans combining physical and linguistic actions at different levels of abstraction. It discusses the planning of concept activation actions that are realized by definite referring expressions in the planned utterances, and shows how it is possible to integrate physical actions for communicating intentions with linguistic actions, resulting in plans that include pointing as one of the communicative actions available to the speaker.

I. INTRODUCTION

One of the most important constituent processes of natural-language generation is the production of referring expressions, which occur in almost every utterance. Referring expressions often carry the burden of informing the hearer of propositions as well as referring to objects. Therefore, many phenomena that are observed in dialogues can-

Figure 1
Satisfying Multiple Goals with a Referring Expression

not be explained by the simple view that referring expressions are descriptions of the intended referent sufficient to distinguish the referent from other objects in the domain or in focus.

Consider the situation (depicted in Figure 1) in which two agents, an apprentice and an expert, are cooperating on a common task, such as disassembling an air compressor. Several tools are lying on the workbench, and although the apprentice knows that the objects are there, he may not necessarily know where they are. The expert might say:

Use the wheelpuller to remove the flywheel. (1)

while pointing at the wheelpuller. The apprentice may think to himself at this point, "Ah, ha, so *that's* a wheelpuller," and then proceed to remove the flywheel.

What the expert is accomplishing through the utterance of (1) by using the noun phrase "the wheelpuller" cannot be fully explained by treating definite referring expressions simply as descriptions that are uniquely true of some object, even taking focusing [7][11] into account. The expert uses "the wheelpuller" to refer to an object that in fact uniquely fits the description predicated of it, so this simple analysis is incapable of accounting for the effects the expert intends his utterance to have.

If one takes the knowledge and intentions of the speaker and hearer into account, a more accurate account of the speaker's use of the referring expression can be developed. The apprentice does not know what the object is that fits the description "the wheelpuller". The expert knows that the apprentice doesn't know this, and performs the pointing action to guarantee that his intentions will be recognized correctly.

The apprentice must recognize what the expert is trying to communicate by pointing — he must realize that pointing is not just a random gesture, but is intended by the speaker to be recognized as a communicative act by the hearer in much the same way as his utterances are recognized as communicative acts. Furthermore, the apprentice must recognize how the pointing act is correlated with the utterance the expert is producing. Although there is no specific deictic reference in the expert's utterance, it is clear that he does not mean the flywheel, since we will assume that the apprentice can determine that the object

he is pointing to is a tool. The apprentice realizes that the object the expert is pointing to is the intended referent of "the wheelpuller," but in the process, he also acquires the information that the expert believes the object he is pointing to is a wheelpuller, and that the expert has also *informed* him of that fact.

A language-planning system called KAMP (for **Know**ledge **A**nd **M**odalities **P**lanner) has been developed that can plan utterances similar to example (1) above, coordinate the linguistic actions with physical actions, and know that the utterance it plans will have the intended multiple effects on the hearer. KAMP builds on Cohen and Perrault's idea of planning speech acts [4], but extends the planning activity down to the level of constructing surface English sentences. A detailed description of the entire KAMP system can be found in [2]. The system has been implemented and tested on examples in a cooperative equipment assembly domain, such as the one in example (1). This paper develops and extends some of the ideas of an early prototype system described in [1].

The reference problems that KAMP addresses are a subset of a more general problem, which, following Cohen [5] will be called 'identification.' Whenever a speaker makes a definite reference, he intends the hearer to identify some object in the world as the referent. Identifying a referent [*] requires that the agent perform some cognitive activity, such as the simple case of matching the description with what he knows, or in some cases plan to perform perceptual actions that lead to the identification. KAMP simplifies the problem by not considering perceptual actions, and assumes that there is some 'perceptual field' common to the participants in a dialogue, and that the objects that lie within that field are mutually known to the participants, along with the observable properties and relations that hold among them.

For example, the speaker and hearer in (1) are assumed to mutually know the size, shape and location of all objects on the workbench. The agents may not know unobservable properties of the objects, such as the fact that a particular tool is a wheelpuller. Similarly, the participants are assumed to be mutually aware of physical actions that take place within their perceptual field, without explicitly performing any perceptual actions. When the expert points at the wheelpuller, the apprentice is simply assumed to know that he is doing it.

II. KNOWLEDGE REPRESENTATION

KAMP uses an intensional logic to describe facts about the world, including the knowledge of agents. The possible-worlds semantics of this intensional logic is axiomatized in first-order logic as described by Moore [8]. The axiomatization enables KAMP to reason about how the knowledge of both the speaker and the hearer changes as they perform actions.

Moore's central idea is to axiomatize operators such as **Know** as relations between possible worlds. For example, if W_0 denotes the real world, then **Know**(John, P) means P is true in every possible world that is consistent with what John knows. This is stated formally in the axiom schema:

$$\forall w_1\, T(w_1, \textbf{Know}(A, P)) \equiv \\ \forall w_2\, K(A, w_1, w_2) \supset T(w_2, P). \tag{1}$$

The predicate $T(w, P)$ means that P is true in possible world w. The predicate $K(A, w_1, w_2)$ means that w_2 is consistent with what A knows in w_1.

Actions are described by treating possible worlds as state variables, and axiomatizing actions as relations between possible worlds. Thus, $R(E, w_1, w_2)$ means that world w_2 is the result of event E happening in world w_2.

It is important that a language planning system reason about mutual knowledge while planning referring expressions [3][5]. Failure to consider the mutual knowledge of the speaker and hearer can lead to the failure of the reference. KAMP uses an axiomatization of mutual knowledge in terms of relations on possible worlds. An agent's knowledge is described as everything that is true in all possible worlds compatible with his 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. [*] To state this fact formally, an individual called the *kernel* of A and B is defined such that the set of possible worlds compatible with the kernel's knowledge is the set of all worlds compatible with either A's knowledge or B's knowledge. This leads to the following definition of mutual knowledge:

$$\forall w_1\, T(w_1, \textbf{MutuallyKnow}(A, B, P)) \equiv \\ \forall w_2\, K(\text{Kernel}(A, B), w_1, w_2) \supset T(w_2, P). \tag{2}$$

In (2), $T(w, P)$ means that the object language proposition P is true in possible world w, and $K(a, w_1, w_2)$ is a predicate that describes the relation between possible worlds that means that w_2 is a possible alternative to w_1 according to a's knowledge. The second axiom needed is:

$$\forall x, w_1, w_2\, K(x, w_1, w_2) \supset \forall y\, K(\text{Kernel}(x, y), w_1, w_2) \tag{3}$$

Axiom (3) 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.

III. THE KAMP PLANNING SYSTEM

KAMP is a multiple-agent planning system designed around a NOAH-like hierarchical planner [10]. KAMP uses two descriptions of each action available to the planning agent: a complete axiomatization of the action using the possible-worlds approach outlined above, and an *action*

---

[*] What it means to *identify* an object is somewhat problematic. KAMP assumes that identification means that the referring description conjoined with focusing knowledge picks out the same individual in all possible worlds consistent with what the agent knows.

[*] Notice that the "intersection" of the propositions believed by two agents is represented by the *union* of possible worlds compatible with their knowledge.

summary consisting of a simplified description of the action that serves as a heuristic to aid in proposing plans that are likely to succeed. KAMP forms a plan using the simplified action summaries first, and then verifies the plan using the full axiomatization. Since the possible-worlds axioms lend themselves more efficiently to proving a plan correct than in generating a plan in the first place, such an approach results in a system that is considerably more efficient than one relying on the possible-worlds axioms alone.

Because action summaries represent actions in a simplified form, the planner can ignore details of the effects of communicative acts to produce a plan that is likely to work in most circumstances. For example, if a simplified description of the effects of informing states that the hearer knows the proposition, then the planner can reason that a plan to achieve the goal of the hearer knowing P is likely to include the action of informing him that P is true. In the relatively unlikely event that this description is inadequate, this fact will be detected during the verification phase where the more complete description is invoked.

The flow of control during KAMP's heuristic plan-generation phase is similar to that of NOAH's. If a goal needs to be satisfied, KAMP searches for actions that can achieve the goal and inserts them into the plan, along with the preconditions, which become new goals to be satisfied. When the entire plan has been expanded to one level of abstraction, then if there is a lower level, all high-level actions that have low-level expansions are expanded.

Between each stage of expansion, *critics* are invoked that examine the plan for global interactions between actions, and make changes in the structure of the plan to avoid the bad effects of the interactions and take advantage of the beneficial ones. Critics play an important role in the planning of referring expressions, and their functions are described more fully in Section IV.
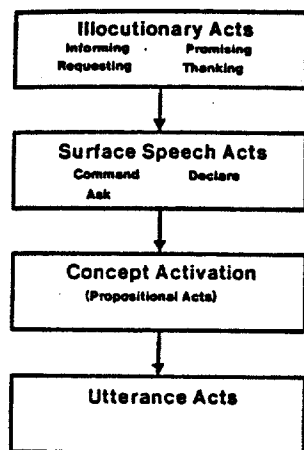

Figure 2
A Hierarchy of Actions Related to Language

KAMP's hierarchy of linguistic actions is illustrated in Figure 2. The hierarchy consists of *illocutionary acts, surface speech-acts, concept-activation actions,* and *utterance*

acts. Illocutionary acts are speech acts such as informing and requesting, which are planned at the highest level without regard for any specific linguistic realization. The next level consists of surface speech-acts, which are abstractions of the actions of uttering particular sentences with particular syntactic structures. At this level the planner starts making commitments to particular choices in syntactic structure, and linguistic knowledge enters the planning process. One surface speech-act can realize one or more illocutionary acts. The next level consists of concept-activation actions, which entail the planning of descriptions that are mutually believed by the speaker and hearer to refer to objects in the world. This is the level of abstraction at which noun phrases for definite reference are planned. Finally, at the lowest level of abstraction are utterance acts, consisting of the utterance of specific words.

IV. PLANNING CONCEPT-ACTIVATION ACTIONS

Concept-activation actions describe referring at a high enough level of abstraction so that they are not constrained to have purely linguistic realizations. When a concept-activation action is expanded to a lower level of abstraction, it can result in the planning of a noun phrase within the surface speech-act of which the concept activation is a part, *and* physical actions such as pointing that also communicate the speaker's intention to refer.

KAMP can plan referential definite noun phrases that realize concept-activation actions. (The planning of attributive and indefinite referring expressions has not yet been addressed.) KAMP recognizes the need to plan a concept activation when it is expanding a surface speech-act. The surface speech-act is planned with a particular proposition that the hearer has to come to believe the speaker wants him to know or want. It is necessary to include whatever information the hearer needs to recognize what the proposition is, and this leads to the necessity of referring to the particular objects mentioned in the proposition. The planner often reasons that some objects do not need to be referred to at all. For example, in requesting a hearer to remove the pump from the platform in an air-compressor assembly task, if the hearer knows that the pump is attached to the platform and nothing else, it is not necessary to mention the platform, since it is sufficient to say "Remove the pump," for the hearer to recognize the following proposition:[*]

$$\text{Want}(S, \text{Do}(H, \text{Remove}(\text{pump1}, \text{platform1}))).$$

The planning of a concept-activation action is similar to the planning of an illocutionary act in that the speaker is trying to get the hearer to recognize his intention to perform the act. This means that all that is necessary from a high-level planning point of view is that the speaker perform some action that signals to the hearer that the

speaker wants to refer to the object. This is often done by incorporating a mutually believed description of the object into the utterance, but there is no requirement that the means by which the speaker communicates this intention be linguistic. For example, the speaker could point at an object (almost always a communicative act), or perhaps throw it at the hearer (not so clearly communicative but definitely attention-getting. The hearer has to reason whether there are any communicative intentions behind the act.)

Since concept-activation actions are planned during the expansion of surface speech-acts, the actions that realize them must somehow become part of the utterance being planned. Therefore, all concept-activation actions are expanded with two components: an *intention-communication* component and a *surface-linguistic* component. The intention-communication component is an abstraction of the speaker's plan to communicate his intention to refer, and may be realized by a plan that includes physical and linguistic actions. The surface-linguistic component consists of the realization (in some linguistic expression) of the intention-communication component *as part of the surface speech-act being planned*, which means that the realization must be grammatically consistent with the sentence.

The following two axiom schemata describe concept activation in KAMP's possible worlds representation:

$$\forall w_1, w_2 \, R(Do(A, \mathrm{Cact}(B, C)), w_1, w_2) \supset \\ T(w_1, \mathbf{Want}(A, \mathrm{Active}(A, B, C))) \land \quad (4) \\ T(w_2, \mathrm{Active}(A, B, C))$$

$$\forall w_1, w_2 \, R(Do(A, \mathrm{Cact}(B, C)), w_1, w_2) \supset \\ \forall w_3 \, K(\mathrm{Kernel}(A, B), w_2, w_3) \supset \\ \exists w_4 \, R(Do(A, \mathrm{Cact}(B, C)), w_4, w_3) \land \quad (5) \\ K(\mathrm{Kernel}(A, B), w_1, w_4)$$

Axiom schema (4) says that when an agent A performs a concept activation for an agent B, he must first want the object C to be active, and as a result of performing it, C becomes active with respect to A and B. Axiom schema (5) says that after agent A performs the action, the two agents A and B mutually know that the action has been performed.

The consequence for the planner of axiomatizing concept activation as in (4) and (5) is that the problem of activating a concept now becomes one of getting the hearer to know that the speaker wants a particular concept to be active. This is the role of the intention-communication component in the expansion of the concept activation.

KAMP knows about two types of actions that produce knowledge about what concepts a speaker wants to be active. One is an action called *describe*, which is ultimately expanded into a linguistic description corresponding to the concept the speaker intends to activate, and the other is called *point*, which is a generalized pointing action. The point action is assumed to directly communicate the intention to activate a concept, thereby avoiding the problem of observing a gesture and deciding whether it is a pointing,

or an attempt to scratch an itch.

The following schema defines the *describe* action:

$$\forall w_1 w_2 \, R(Do(A, \mathrm{Describe}(B, D)), w_1, w_2) \supset \\ T(w_1, \exists x \, D^*(x) \land (\forall y \, D^*(y) \supset x = y)) \equiv \quad (6) \\ T(w_1, \mathbf{Want}(A, \mathrm{Active}(A, B, x)))$$

Axiom (6) says that the precondition for an agent to perform an action of describing using a particular description D is that the speaker wants an object to be active if and only if it uniquely fits the description predicated of it. In (6), the symbol D denotes a *description* consisting of object language predicates that can be applied to the object being described. It could be defined as

$$D \equiv \lambda x. (D_1(x) \land \ldots \land D_n(x))$$

where the $D_i(x)$ are the individual *descriptors* that comprise the description. The symbol D^* denotes a similar expression, which includes all the descriptors of D conjoined with a set of predicates that describe the focus of the discourse.* An axiom similar to (5) is also needed to assert that the speaker and hearer will mutually know, after the action is performed, that it has taken place. Therefore, if the speaker and hearer mutually know of an object that satisfies D in focus, then they mutually know that the speaker wants it to be active.

The pointing action is much simpler because it does not require either the speaker or the hearer to know anything at all about the object.

$$\forall w_1, w_2 \, R(Do(A, \mathrm{Point}(B, X)), w_1, w_2) \supset \\ T(w_1, \mathbf{Want}(A, \mathrm{Active}(A, B, X))). \quad (7)$$

According to the above axiom, if an agent points at an object, that implies that he wants the object to be active. As usual, an axiom similar to (5) is required to assert that the agents mutually know the action has been performed.

Axioms (4) and (5) work together with (6) and (7) to produce the desired effects. When a speaker utters a description, or points, he communicates his intention to refer. When he performs the concept-activation action by incorporating the surface-linguistic component of his action into a surface speech-act, his intentions are carried out. Because the equivalence of axiom (6) can be used in both directions, if the speaker wants an object to be active, then one can reason that he knows the description predicated of it is true.

A major problem facing the planner is deciding when the necessary conditions obtain to be able to take advantage of the interactions between (6) and (7). Since this task involves examining several actions in the plan, it is performed by a *critic* called the *action-subsumption critic*. This critic notices when the speaker is informing the hearer

---

* A complete discussion of focusing in KAMP is beyond the scope of this paper. KAMP uses an axiomatization of Sidner's focusing rules [11] to keep track of focus shifts.

of a predication that could be included in the description associated with a concept activation. When such an interaction is noticed, the critic proposes a modification to the plan. If the surface-linguistic component does not insist that the modification is impossible given the grammar, then the action subsumption is carried out.

In example (1), for instance, the expert has a high-level plan that includes the performance of two illocutionary acts: requesting that the apprentice remove the pump using a particular tool (call it *tool1*), and informing the apprentice that *tool1* is a wheelpuller. The action subsumption critic notices that in the request the expert is referring to *tool1* and also wants to inform the hearer of a property of *tool1*. Therefore, it proposes combining the property of being a wheelpuller into the description used for referring to *tool1* while making the request.

V. CONCLUSION

This paper has described a formalism for describing the action of referring in a manner that is useful for a generation system based on planning, like KAMP. The central idea is to divide referring into two tasks: an intention-communication task and a surface-linguistic task. By so doing, it is possible to axiomatize different actions that communicate a speaker's intention to refer. Thus, the planner is able to produce plans that produce natural-language referring expressions, but take the larger context of the speaker's nonlinguistic actions into account as well.

KAMP currently plans only simple definite reference. One promising extension of this approach for future research is to extend the *active* predicate to apply to intensional concepts in addition to the extensional ones now required for definite reference. We hope this will allow for the planning of attributive and indefinite reference as well. KAMP currently does not plan quantified noun phrases, nor can it refer generically, nor can it refer to collections of entities. Much basic research needs to be done to extend KAMP to handle these other cases, but we hope that the formalism outlined here will provide a good base from which to investigate these extensions.

VI. ACKNOWLEDGEMENTS

The author is grateful to Barbara Grosz, Bob Moore and Nils Nilsson for comments on earlier drafts of this paper.

VII. REFERENCES

[1]   Appelt, Douglas E., *Problem Solving Applied to Language Generation,* Proceedings of the 18th Annual Meeting of the ACL, 1980.

[2]   Appelt, Douglas E., *Planning Natural Language Utterances To Satisfy Multiple Goals,* SRI International Technical Note No. 259, 1982.

[3]   Clark, Herbert, and C. Marshall, *Definite Reference and Mutual Knowledge,* in Joshi et. al. (eds.), **Elements of Discourse Understanding**, Cambridge University Press, Cambridge, 1981.

[4]   Cohen, Philip and C. R. Perrault, *Elements of a Plan-Based Theory of Speech Acts,* Cognitive Science, vol. 3, pp. 177–212, 1979.

[5]   Cohen, Philip, and H. Levesque, *Speech Acts and the Recognition of Shared Plans,*, Proceedings of the Canadian Society for Computational Studies in Intelligence, 1980.

[6]   Cohen, Philip, *The Need for Referent Identification as a Planned Action,* Proceedings of IJCAI-7, 1981.

[7]   Grosz, Barbara J., *Focusing and Description in Natural Language Dialogs,* in Joshi et al. (eds.), **Elements of Discourse Understanding: Proceedings of a Workshop on Computational Aspects of Linguistic Structure and Discourse Setting**, Cambridge University Press, Cambridge, 1980.

[8]   Moore, Robert C., *Reasoning about Knowledge and Action,* SRI International Technical Note No. 191, 1980.

[9]   Olson, D., *From Utterance to Text: The Bias of Language in Speech and Writing,* Harvard Educational Review, Vol. 47, No. 3, August, 1977.

[10] Sacerdoti, Earl, **A Structure for Plans and Behavior,** Elsevier North-Holland, Inc., Amsterdam, 1977.

[11] Sidner, Candacl L., *Toward a Computational Theory of Definite Anaphora Comprehension in English,* MIT Technical Report AI-TR-537, 1979.