

PLANNING NATURAL-LANGUAGE UTTERANCES

Douglas E. Appelt

SRI International
Menlo Park, California

ABSTRACT

This paper describes recent research on a natural-language-generation system that is based on planning. A system named KAMP is described that is capable of producing English sentences as part of a plan to enlist the cooperation of another agent in achieving a goal involving a change in the physical state of the world. The planner uses knowledge about the different subgoals to be achieved and linguistic rules about English to produce sentences that satisfy multiple goals through the realization multiple illocutionary acts.

I. INTRODUCTION

This paper describes a view of natural-language production similar to that of Allen [1], Cohen and Perrault [6], namely that speakers produce utterances with the intention of satisfying particular goals, and that a hearer's understanding of an utterance depends on how he interprets the utterance as a component of what he believes is the speaker's plan.

A system named KAMP (for **K**nowledge **A**nd **M**odalities **P**lanner) has been developed that plans natural-language utterances, starting with a high-level description of the speaker's goals. The system can be viewed as an extension of speech-act planning research by Cohen [6], but while Cohen was only concerned with producing abstract descriptions of speech acts, KAMP allows the planning to be extended to the level of the production of English sentences, integrating both physical and linguistic actions that satisfy discourse, knowledge-state and referring goals into utterances that achieve multiple goals simultaneously.

This research has addressed the following three major problems:

- Developing a domain-independent multiple-agent planning system called KAMP.
- Extending the possible-worlds semantics representation of propositional attitudes developed by Moore [9] to handle mutual knowledge and wanting.

- Capturing linguistic knowledge in the axioms, critics, and procedures used by KAMP to facilitate the planning of *linguistic* actions.

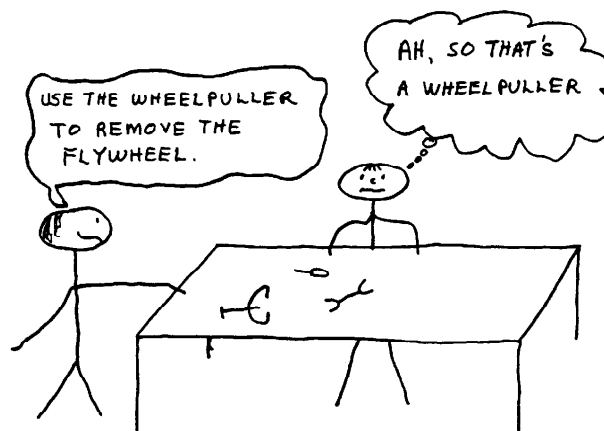


Figure 1

Satisfying Multiple Goals with a Request

II. WHY PLAN UTTERANCES?

Figure 1 illustrates a typical situation arising when two people cooperate on a common task. The speaker plans an utterance that has multiple effects on the hearer. The speaker points to one of the tools on the table, saying "Use the wheelpuller to remove the flywheel." The hearer, who is observing the speaker while he makes the request and knows that the speaker is pointing to the particular tool, thinks to himself, "Ah, so that's a wheelpuller. I was wondering how I was going to get the flywheel off."

The speaker's utterance in Figure 1 is syntactically very simple, but a surprising amount of complex reasoning is required for a speaker to produce such an utterance and know that it will have the intended effect on the hearer. The speaker obviously wants to request that the hearer do something. However, before he can make the request, he has to determine whether the hearer has enough knowledge to carry it out. If not, then the speaker has to know that the hearer can form a plan for acquiring the knowledge, or he must furnish the knowledge himself. In this example, the speaker *informs* the hearer that he should use the wheelpuller as part of the same utterance that he uses to

* 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.

request the action he wants performed.

The speaker uses the noun phrase "the wheelpuller" to refer to a particular tool. In Figure 1, it is evident from the hearer's reaction that he did not know what a wheelpuller was. It is also clear that the speaker knew what the hearer did not know, because he performed a pointing action to clarify his intention to refer. Although the speaker knew that the hearer did not know what a wheelpuller was, he knew that the hearer would know after understanding the sentence. The utterance of Figure 1 also serves to inform the hearer that the object to which the speaker is pointing is a wheelpuller. For the speaker to make that inference, he had to know that the hearer would know that he did not intend the object to which he was pointing be the referent of the noun phrase "the flywheel." He knows that because he knows the hearer understands that the flywheel is not a tool, and, therefore, cannot fill the instrument role of "remove." Under different circumstances, the speaker could point to the flywheel, utter a sentence identical to the one in Figure 1, and reason that the hearer would learn that the object to which he was pointing was the flywheel.

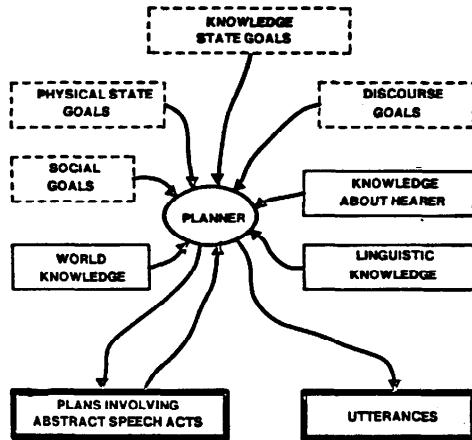


Figure 2

The Organization of a Language Planning System

III. THE KAMP LANGUAGE PLANNING SYSTEM

It is clear from the above example that a model of language production that simply transforms a logical form to a surface utterance does not account for the way that people use utterances to satisfy multiple goals, because the model does not include the reasoning that the speaker and hearer perform to infer each other's plans. Furthermore, utterances do more than alter the participants' knowledge and wants. They influence the participants' emotional attitudes and affect the state of the ongoing discourse. Utterances can be planned with the intention of achieving goals along these dimensions as well. Therefore, instead of a simple transducer from logical form to surface English sentences, KAMP is organized like the planner in Figure 2.

Appelt [2][3] discussed the organization of KAMP as a hierarchical planner similar in overall organization to Sacerdoti's NOAH. KAMP has two descriptions of actions at each level in the action hierarchy: (1) A full axiomatization in terms of possible worlds, and (2) a shorter, more intuitive description called an *action summary*. KAMP uses the action summaries as a heuristic device to propose plans that it then verifies using the possible worlds axiomatization. The heuristic plan generation process is implemented by the NOAH-like hierarchical planner, and the verification process is implemented by a first-order-logic theorem prover.

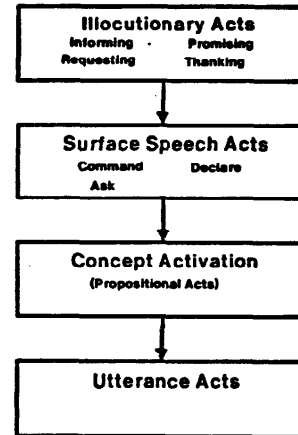


Figure 3

A Hierarchy of Actions Related to Language

Figure 3 illustrates the hierarchy of actions that is used by KAMP to plan linguistic actions. The central problem of building a language-planning system around KAMP is formulating the correct axioms, and incorporating the correct action summaries and critic procedures into KAMP that describe the actions of the hierarchy in Figure 3.

IV. AXIOMATIZING KNOWLEDGE ABOUT INTENSIONAL CONCEPTS

Axiomatizing the actions of Figure 3 requires the ability to specify the way in which performing actions affects the knowledge and wants of agents. Moore's possible-worlds-semantics approach [9] solves this problem for knowledge and its relation to action. Describing Moore's approach in detail is not possible here, but the central idea is to axiomatize the possible-worlds semantics of a modal object language in a first-order meta language. Thus, the semantics of a statement like $\text{Know}(A, P)$ is represented as " P is true in every possible world consistent with what A knows."

It is necessary to reason about mutual knowledge (i.e., knowledge that A knows that B knows that A knows ... ad infinitum) to plan referring expressions (see Clark & Marshall, [5]). KAMP reasons about the mutual knowledge shared by two agents by reasoning about what is true in the union of the sets of possible worlds consistent with

two agents' knowledge. An "agent" called the *kernel* of *A* and *B* is defined, for whom the worlds consistent with his knowledge are precisely that union. This approach is a generalization of the "any fool" approach advocated by McCarthy et al. [7].

Wanting is represented in KAMP by a relation between an agent and a set of possible worlds called a *situation*. The situation is a set of possible alternatives to the current world which an agent wants. The situation that an agent wants can be characterized by different propositions according to what he knows. An agent is said to want *P* if there is some situation he wants such that *P* is true in every possible world that is a member of the situation, with the terms of *P* evaluated with respect to the agent's knowledge. This representation allows a connection between knowledge and wanting, which, while ignoring many of the subtle problems associated with wanting and intention, is adequate for solving many planning problems that arise in the task-oriented domains under consideration. This approach allows such reasoning as, for example, if John wants to meet the president of the United States, and if John knows that the president of the United States is Ronald Reagan, then John wants to meet Ronald Reagan.

V. AXIOMATIZING LINGUISTIC ACTIONS

As illustrated in Figure 3, the most abstract linguistic actions in KAMP's hierarchy are *illocutionary acts*. These are actions such as informing, requesting, promising, thanking, etc. that can be characterized as communicative acts independent of any particular linguistic realization.

Speakers do not perform illocutionary acts directly, but rather perform them *by means of surface speech acts*. When a speaker plans a surface speech act, he selects the propositional content of the sentence he is going to utter (which may be different from the propositional content of the illocutionary act in the case of indirect speech acts), and selects a particular syntactic structure that is used for the realizing the illocutionary act. A fundamental choice made at this level is whether to use an imperative, interrogative or declarative sentence. Each surface speech act has a syntactic structure tree associated with it that evolves as the plan is expanded to include more constituents that are expanded to progressively lower levels.

The relationship between illocutionary acts and surface speech acts is similar to the relationship between walking across the room and a sequence of muscle movements. One action is performed by performing the others. What distinguishes this relationship from that of walking and muscle movements is that the particular illocutionary act that is being performed depends on the hearer's recognition of what the speaker is trying to do. A particular surface speech act, for example, "Can you reach the tool on the top shelf?" can in one case be a request to retrieve a tool, and in another case a request to inform the speaker of the ability to retrieve a tool.

As the example illustrates, it is not even the case that there is a one-to-one correspondence between illocutionary

acts and surface speech acts. In this case several informing and requesting actions are being performed as part of a single surface speech act.

KAMP has an axiomatization of each illocutionary act and surface speech act it knows about in terms of the possible worlds approach outlined above. This paper will not describe the axioms in detail. The interested reader is referred to Appelt [2] for more information. The general approach to axiomatizing illocutionary acts is to describe only what Austin [4] refers to as illocutionary effects, not perlocutionary effects. In other words, the effect of informing a hearer that *P* is not that the hearer then believes *P*, but that the hearer knows that the speaker wants him to know that the speaker believes *P*. However, the speaker wanting the hearer to know that the speaker believes *P* is a reasonable *precondition* for the sincere performance of an informing action. Therefore, the effects of an illocutionary act can be said to produce the mutual knowledge between the speaker and the hearer that the act has been performed. All deductions about the change in the knowledge of the participants follow from knowing that the action has been performed and from their mutual knowledge of the conditions on the action's performance.

Surface speech acts include *concept activation actions* as part of their realization on the next lower level of abstraction. Concept activation actions perform referring at a high enough level of abstraction so that they are not constrained to be purely linguistic actions. When a concept activation action is expanded to a lower level of abstraction, it can result in planning 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, and may be realized by a plan that includes either physical or linguistic actions.

Although concept activation actions can be realized through physical actions, the planner must reason about their interaction with the linguistic actions being planned. Therefore, concept activation actions are expanded into two components, an *intention-communication component* communicating the speaker's intention to refer, and a *linguistic-realization component* realizing the action in a sentence in accordance with the grammatical rules.

The lowest level actions of Figure 3 are the utterance acts. Utterance acts consist of the utterance of particular sequences of words. The component of KAMP that produces utterance acts from a plan of hierarchical linguistic actions and the constituent-structure trees associated with surface speech acts is quite simple because at this point, no modifications to the plan are made because decisions are completely determined by grammatical rules, and the speaker's intentions do not influence the process. The final stage of planning consists primarily of making obligatory modifications required by the grammar, such as subject-verb agreement, proper auxiliary affixes, insertion of reflexive pronouns, and the like.

VI. CONCLUSION

The development of KAMP has been the first step toward a theory of planning natural-language utterances that allows the satisfaction of multiple goals in a single surface utterance, that plans utterances tailored to the specific knowledge of an intended hearer as well as the context of the discourse, and that provides for the integration of physical and linguistic actions.

There are a number of areas in which the concepts developed in KAMP can be profitably applied and extended. One major area is the planning of extended discourse. Currently, KAMP plans only very simple dialogues. It may plan more than one utterance if it wants to perform several illocutionary acts and it cannot figure out a way in which one can subsume the others. The resulting dialogues will be coherent because the illocutionary acts are naturally tied together by being part of the same plan. However, to move beyond simple dialogues consisting of alternating one or two sentence turns, more complex, abstract discourse-level actions must be defined. McKeown [8] incorporates such strategies in a language generation system, and such actions need to be formalized in a planning framework for use by a system like KAMP.

KAMP currently keeps track of discourse focus primarily so it can generate appropriate referring expressions. When planning an extended discourse, the planner would also be concerned about the speaker's need to inform the hearer of topic shifts. Topic shifting actions, similar to those described by Reichman [10], must be formalized so they can be planned when appropriate.

The primary focus of research on KAMP has been on planning natural-language utterances. However, KAMP is a general tool that can serve as the basis of multiple-agent planning systems in a variety of domains. There are many problems concerning planning to acquire knowledge, cooperation among several agents with limited resources, for which KAMP seems useful.

VII. ACKNOWLEDGEMENTS

The author is grateful to Barbara Grosz and Jane Robinson for comments on an earlier draft of this paper.

REFERENCES

- [1] Allen, James and C. R. Perrault, *Analyzing Intention in Utterances*, Artificial Intelligence, vol. 15, no. 3, December 1980.
- [2] Appelt, Douglas E., *A Planner for Reasoning about Knowledge and Action*, Proceedings of the First Annual National Conference on Artificial Intelligence, 1980.
- [3] Appelt, Douglas E., *Planning Natural Language Utterances to Satisfy Multiple Goals*, SRI International Artificial Intelligence Center Technical Note No. 259, 1982.
- [4] Austin, J., *How to do things with Words*, J. O. Urmson (ed.), Oxford University Press, 1962.
- [5] Clark, Herbert, and C. Marshall, *Definite Reference and Mutual Knowledge*, in Joshi et. al. (eds.), **Elements of Discourse Understanding**, Cambridge University Press, Cambridge, 1981.
- [6] Cohen, Philip and C. R. Perrault, *Elements of a Plan Based Theory of Speech Acts*, Cognitive Science, vol. 3, pp. 177-212, 1979.
- [7] McCarthy, John, M. Sato, T. Hayashi, and S. Igarashi, *On the Model Theory of Knowledge*, Stanford Artificial Intelligence Laboratory Memo AIM-312, 1978.
- [8] McKeown, Kathleen R., *Generating Relevant Explanations: Natural Language Responses to Questions about Database Structure*, Proceedings of the First Annual National Conference on Artificial Intelligence, 1980.
- [9] Moore, Robert C., *Reasoning about Knowledge and Action*, SRI International Technical Note No. 191, 1980.
- [10] Reichman, Rachel, *Conversational Coherency*, Cognitive Science, Vol. 2, No. 4, 1978.
- [11] Sacerdoti, Earl, **A Structure for Plans and Behavior**, Elsevier North-Holland, Inc., Amsterdam, 1977.