PROBLEM SOLVING APPLIED TO LANGUAGE GENERATION

Douglas E. Appelt
Stanford University, Stanford, California
SRI International, Menlo Park, California

This research was supported at SRI International by the Defense Advanced Research Projects Agency under contract N00039-79-C-0118 with the Naval Electronic Systems Command. The views and conclusions contained in this document are those of the author and should not be interpreted as representative of the official policies, either expressed or implied, of the Defense Advanced Research Projects Agency, or the U. S. Government. The author is grateful to Barbara Grosz, Gary Hendrix and Terry Winograd for comments on an earlier draft of this paper.

I. Introduction

Previous approaches to designing language understanding systems have considered language generation to be the activity of a highly specialized linguistic facility that is largely independent of other cognitive capabilities. All the requisite knowledge for generation is embodied in a "generation module" which, with appropriate modifications to the lexicon, is transportable between different domains and applications. Application programs construct "messages" in some internal representation, such as first order predicate calculus or semantic networks, and hand them to the generation module to be *translated* into natural language. The application program decides *what to say*; the generation module decides *how to say it.*

In contrast with this previous work, this paper proposes an approach to designing a language generation system that builds on the view of language as action which has evolved from speech act theory (see Austin [2] and Searle [11]). According to this view, linguistic actions are actions planned to satisfy particular goals of the speaker, similar to other actions like moving and looking. Language production is integrated with a speaker's problem solving processes. This approach is founded on the hypothesis that planning and performing linguistic actions is an activity that is not substantially different from planning and performing other kinds of physical actions. The process of producing an utterance involves planning actions to satisfy a number of different kinds of goals, and then efficiently coordinating the actions that satisfy these goals. In the resulting framework, there is no distinction between deciding what to say and deciding how to say it.

This research has proceeded through a simultaneous, integrated effort in two areas. The first area of research is the theoretical problem of identifying the goals and actions that occur in human communication and then characterizing them in planning terms. The second is the more applied task of developing machine-based planning methods that are adequate to form plans based on the characterization developed as part of the work in the first area. The eventual goal is to merge the results of the two areas of effort into a planning system that is capable of producing English sentences.

Rather than relying on a specialized generation module, language generation is performed by a general problem-solving system that has a great deal of knowledge about language. A planning system, named KAMP (Knowledge and Modalities Planner), is currently under development that can take a high-level goal and plan to achieve it through both linguistic and non-linguistic actions. Means for satisfying multple goals can be integrated into a single utterance.

This paper examines the goals that arise in a dialog, and what actions satisfy those goals. It then discusses an example of a sentence which satisfies several goals simultaneously, and how KAMP will be able to produce this and similar utterances. This system represents an extension to Cohen's work on planning speech acts [3]. However, unlike Cohen's system which plans actions on the level of informing and requesting, but does not actually generate natural language sentences, KAMP applies general problem-solving techniques to the entire language generation process, including the construction of the utterance.

II. Goals and Actions used in Task Oriented Dialogues

The participants in a dialogue have four different major types of goals which may be satisfied, either directly or indirectly, through utterances. *Physical goals*, involve the physical state of the world. The physical state can only be altered by actions that have physical effects, and so speech acts do not serve directly to achieve these goals. But since physical goals give rise to other types of goals as subgoals, which may in turn be satisfied by speech acts, they are important to a language planning system. Goals that bear directly on the utterances themselves are *knowledge state goals*, *discourse goals*, and *social goals*.

Any goal of a speaker can fit into one of these four categories. However, each category has many sub-categories, with the goals in each sub-category being satisfied by actions related to but different from those satisfying the goals of other sub-categories. Delineating the primary categorizations of goals and actions is one objective of this research.

Knowledge state goals involve changes in the beliefs and wants held by the speaker or the hearer. They may be satisfied by several different kinds of actions. Physical actions affect knowledge, since as a minimum the agent knows he has performed the action. There are also actions that affect only knowledge and do not change the state of the world — for example, reading, looking and speech acts. Speech acts are a special case of knowledge-producing actions because they do not produce knowledge directly, like looking at a clock. Instead, the effects of speech acts manifest themselves through the recognition of *intention*. The effect of a speech act, according to Searle, is that the hearer recognizes the speaker's intention to perform the act. The hearer then knows which speech act has been performed, and because of rules governing the communication processes, such as the Gricean maxims [4], the hearer makes inferences about the speaker's beliefs. These inferences all affect the hearer's own beliefs.

Discourse goals are goals that involve maintaining or changing the state of the discourse. For example, a goal of focusing on a different concept is a type of discourse goal [5, 9, 12]. The utterance *Take John, for instance* serves to move the participants' focusing from a general subject to a specific example. Utterances of this nature seem to be explainable only in terms of the effects they have, and not in terms of a formal specification of their propositional content.

Concept activation goals are a particular category of discourse goals. These are goals of bringing a concept of some object, state, or event into the hearer's immediate conciousness so that he understands its role in the utterance. Concept activation is a general goal that subsumes different kinds of speaker reference. It is a low-level goal that is not considered until the later stages of the planning process, but it is interesting because of the large number of interactions between it and higher-level goals and the large number of options available by which concept activations can be performed.

Social goals also play an important part in the planning of utterances. These goals are fundamentally different from other goals in that frequently they are not effects to be achieved so much as constraints on the possible behavior that is acceptable in a given situation. Social goals relate to politeness, and are reflected in the surface form and content of the utterance. However, there is no simple "formula" that one can follow to construct polite utterances. *Do you know what time it is?* may be a polite way to ask the time, but *Do you know your phone number?* is not very polite in most situations, but *Could you tell me your phone number?* is.

What is important in this example is the exact propositional content of the utterance. People are expected to know phone numbers, but not necessarily what time it is. Using an indirect speech act is not a sufficient condition for politeness. This example illustrates how a social goal can influence *what* is said, as well as *how* it is expressed.

Quite often the knowledge state goals have been assigned a special priviliged status among all these goals. Conveying a propsition was viewed as the primary reason for planning an utterance, and the task of a language generator was to somehow construct an utterance that would be appropriate in the current context. In contrast, this research attempts to take Halliday's claim [7] seriously in the design of a computer system:

> "We do not, in fact, first decide what we want to say independently of the setting and then dress it up in a garb that is appropriate to it in the context. . . . The 'content' is part of the total planning that takes place. There is no clear line between the 'what' and the 'how'. . ."

The complexity that arises from the interactions of these different types of goals leads to situations where the content of an utterance is dictated by the requirement that it fit into the current context. For example, a speaker may plan to inform a hearer of a particular fact. The context of the discourse may make it impossible for the speaker to make an abrupt transition from the current topic to the topic that includes that proposition. To make this transition according to the communicative rules may require planning another utterance. Planning this utterance will in turn generate other goals of informing, concept activation and focusing. The actions used to satisfy these goals may affect the planning of the utterance that gave rise to the subgoal. In this situation, there is no clear dividing line between "what to say" and "how to say it".

III. An Integrated Approach to Planning Speech Acts

A probem-solving system that plans utterances must have the ability to describe actions at different levels of abstraction, the ability to specify a partial ordering among sequences of actions, and the ability to consider a plan globally to discover interactions and constraints among the actions already planned. It must have an intelligent method for maintaining alternatives, and evaluating them comparatively. Since reasoning about belief is very important in planning utterances, the planning system must have a knowledge representation that is adequate for representing facts about belief, and a deduction system that is capable of using that representation efficiently.

KAMP is a planning system, which is currently being implemented, that builds on the NOAH planning system of Sacerdoti [10]. It uses a possible-worlds semantics approach to reasoning about belief and the effects that various actions have on belief [8] and represents actions in a data structure called a *procedural network*. The procedural network consists of nodes representing actions at some level of abstraction, along with *split* nodes, which specify several partially ordered sequences of actions that can be performed in any order, or perhaps even in parallel, and *choice* nodes which specify alternate actions, any one of which would achieve the goal.

Figure 1 is an example of a simple procedural network that represents the following plan: The top-level goal is to achieve P. The downward link from that node in the net points to an expansion of actions and subgoals, which when performed or achieved, will make P true in the resulting world. The plan consists of a choice between two alternatives. In the first the agent A does actions A1 and A2, and no commitment has been made to the ordering of these two parts of the plan. After both of those parts have been completely planned and executed, then action A3 is performed in the resulting world. The other alternative is for agent B to perform action A4.

It is an important feature of KAMP that it can represent actions at several levels of abstraction. An INFORM action can be considered as a high level action, which is expanded at a lower level of abstraction into concept activation and focusing actions. After each expansion to a lower level of abstraction, KAMP invokes a set of procedures called *critics* that examine the plan globally, considering the interactions between its parts, resolving conflicts, making the best choice among available alternatives, and noticing redundant actions or actions that could be subsumed by minor alterations in another part of the plan. The control structure could be described as a loop that makes a plan, expands it, criticizes the result, and expands it again, until the entire plan consists of executable actions.

The following is an example of the type of problem that KAMP has been tested on: A robot named Rob and a man named John are in a room that is adjacent to a hallway containing a clock. Both Rob and John are capable of moving, reading clocks, and talking to each other, and they each know that the other is capable of performing these actions. They both know that they are in the room, and they both know where the hallway is. Neither Rob nor John knows what time it is. Suppose that Rob knows that the clock is in the hall, but John does not. Suppose further that John wants to know what time it is, and Rob knows he does. Furthermore, Rob is helpful, and wants to do what he can to insure that John achieves his goal. Rob's planning system must come up with a plan, perhaps involving actions by both Rob and John, that will result in John knowing what time it is.

Rob can devise a plan using KAMP that consists of a choice between two alternatives. First, if John could find out where the clock is, he could go to the clock and read it, and in the resulting state would know the time. So, Rob can tell John where the clock is, reasoning that this information is sufficient for John to form and execute a plan that would achieve his goal.

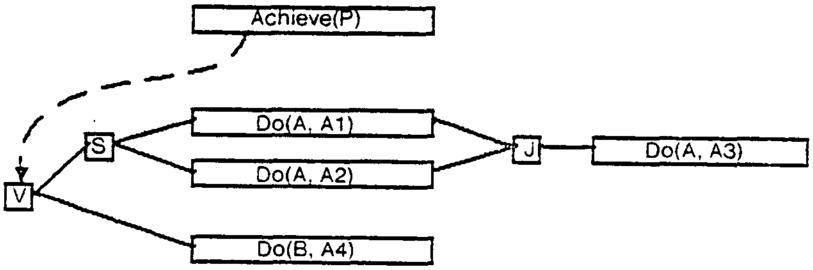

Figure 1
A Simple Procedural Network

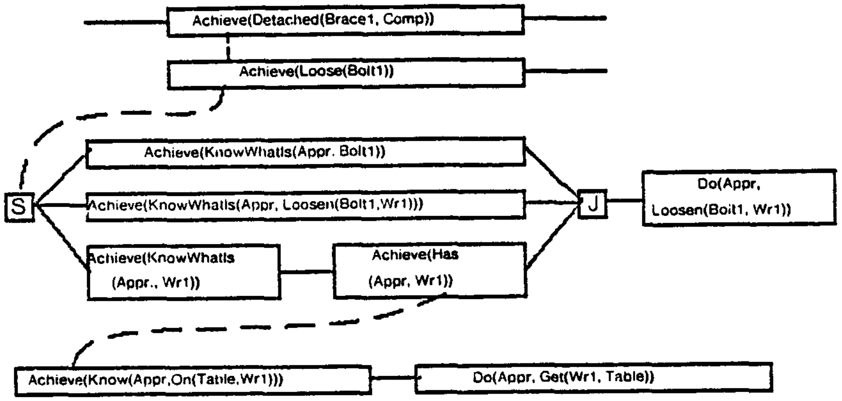

Figure 2
A Plan to Remove a Bolt

The second alternative is for Rob to move into the hall and read the clock himself, move back into the room, and tell John the time.

As of the time of this writing, KAMP has been implemented and tested on problems involving the planning of high level speech act descriptions, and performs tasks comparable to the planner implemented by Cohen. A more complete description of this planner, and the motivation for its design can be found in [1]. The following example is intended to give the reader a feeling for how the planner will proceed in a typical situation involving linguistic planning, but is not a description of a currently working system.

An expert and an apprentice are cooperating in the task of repairing an air compressor. The expert is assumed to be a computer system that has complete knowledge of all aspects of the task, but has no means of manipulating the world except by requesting the apprentice to do things, and furnishing him or her with the knowledge needed to complete the task.

Figure 2 shows a partially completed procedural network. The node at the highest level indicates the planner's top-level goal, which in this case is removing a particular object (BRACE1) from an air compressor. It knows that this goal can be achieved by the apprentice executing a particular unfastening operation involving a specific wrench and a specific bolt. The expert knows that the apprentice can do the action if he knows what the objects involved in the task are, and knows what the action is (i.e. that he knows how to do the action). This is reflected in the second goal in the split path in the procedural network. Since the plan also requires obtaining a wrench and using it, a goal is also established that the apprentice knows where the wrench is: hence the goal ACHIEVE(Know(Apprentice, On(Table, Wr1))).

Assume that the apprentice knows that the part is to be removed, and wants to do the removal, but does not know of a procedure for doing it. This situation would hold if the goal marked with an asterisk in figure 2 were unsatisfied. The expert must plan an action to inform the apprentice of what the desired action is. This goal expands into an INFORM action. The expert also believes that the apprentice does not know where the wrench is, and plans another INFORM action to tell him where it is located.

The planner tests the ACHIEVE goals to see if it believes that any of them are already true in the current state of the world. In the case we are considering, KAMP's model of the hearer should indicate that he knows what the bolt is, and what the wrench is, but doesn't know what the action is, i.e. that he should use that particular wrench to loosen that bolt, and he doesn't know the location of the wrench. If informing actions are planned to satisfy those goals that are not already satisfied, then that part of the plan looks like Figure 3.

Each of the INFORM actions is a high-level action that can be expanded. The planner has a set of standard expansions for actions of this type. In NOAH, these actions were written in SOUP code. In this planner, they are represented in situation-action rules. The conditional of the rule involves tests on the type of action to be performed, the hearer's knowledge, and social goals. The action is to select a particular strategy for expanding the action. In this case, a rule such as *If you are expanding an inform of what an action involving the hearer as agent is, then use an IMPERATIVE syntactic construct to describe the action.* The planner then inserts the expansion shown in Figure 4 into the plan.

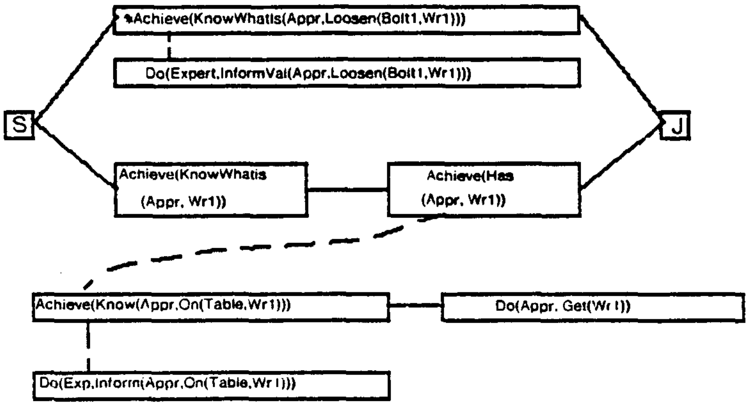

Figure 3
Planning to Inform

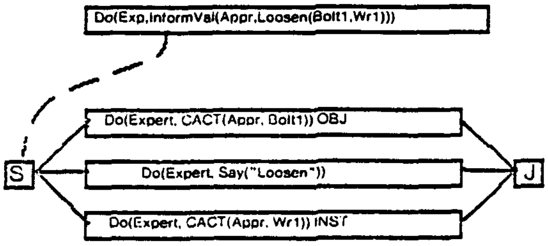

Figure 4
Expanding the INFORM Act

This sub-plan is marked by a tag indicating that it is to be realized by an imperative. The split specifies which lower level actions are performed by the utterance of the imperative. At some point, a critic will choose an ordering for the actions. Without further information the sentence could be realized in any of the following ways, some of which sound strange when spoken in islolation:

Loosen Bolt1 *with* Wr1.

With Wr1 *loosen* Bolt1.

Bolt1 *loosen with* Wr1.

The first sentence above sounds natural in isolation. The other two might be chosen if a critic notices a need to realize a focusing action that has been planned. For example, the second sentence shifts the focus to the wrench instead of the bolt, and would be useful in organizing a series of instructions around what tools to use. The third would be used in a discourse organized around what object to manipulate next.

Up to this point, the planning process has been quite straightforward, since none of the critics have come into play. However, since there are two INFORM actions on two branches of the same split, the COMBINE-CONCEPT-ACTIVATION critic is invoked. This critic is invoked whenever a plan contains a concept activation on one branch of the split, and an inform of some property of the activated object on the other branch. Sometimes the planner can combine the two informing actions into one by including the property description of one of the informing acts into the description that is being used for the concept activation.

In this particular example, the critic would attach to the Do(Expert, CACT(Appr., Wr1)) action the constraint that one of the realizing descriptors must be ON(Wr1, Table), and the goal that the apprentice knows the wrench is on the table is marked as already satisfied.

Another critic, the REDUNDANT-PATH critic, notices when portions of two brances of a split contain identical actions, and collapses the two branches into one. This critic, when applied to utterance plans will often result in a sentence with an *and* conjunction. The critic is not restricted to apply only to linguistic actions, and may apply to other types of actions as well.

Other critics know about action subsumption, and what kinds of focusing actions can be realized in terms of which linguistic choices. One of these action subsumption critics can make a decision about the ordering of the concept activations, and can mark discourse goals as phantoms. In this example, there are no specific discourse goals, so it is possible to chose the default *verb-object-instrument* ordering.

On the next next expansion cycle, the concept activations must be expanded into utterances. This means planning descriptors for the objects. Planning the right description requires reasoning about what the hearer believes about the object, describing it as economically as possible, and then adding the additional descriptors recommended by the action subsumption critic. The final step is realizing the descriptors in natural language. Some descriptors have straightforward realizations as lexical items. Others may require planning a prepositional phrase or a relative clause.

IV. Formally defining linguistic actions

If actions are to be planned by a planning system, they must be defined formally so they can be used by the system. This means explicitly stating the preconditions and effects of each action. Physical actions have received attention in the literature on planning, but one aspect of physical actions that has been ignored are their effects on knowledge. Moore [8] suggests an approach to formalizing the knowledge effects of physical actions, so I will not pursue that further at this time.

A fairly large amount of work has been done on the formal specification of speech acts on the level of informing and requesting, etc. Most of this work has been done by Searle [11], and has been incorporated into a planning system by Cohen [3].

Not much has been done to formally specify the actions of focusing and concept activation. Sidner [12] has developed a set of formal rules for detecting focus movement in a discourse, and has suggested that these rules could be translated into an appropriate set of actions that a generation system could use. Since there are a number of well defined strategies that speakers use to focus on different topics, I suggest that the preconditions and effects of these strategies could be defined precisely and they can be incorporated as operators in a planning system. Reichmann [9] describes a number of focusing strategies and the situations in which they are applicable. The focusing mechanism is driven by the speaker's goal that the hearer know what is currently being focused on. This particular type of knowledge state goal is satisfied by a variety of different actions. These actions have preconditions which depend on what the current state of the discourse is, and what type of shift is taking place.

Consider the problem of moving the focus back to the previous topic of discussion after a brief digression onto a different but related topic. Reichmann points out that several actions are available. One such action is the utterance of "anyway" which signals a more or less expected focus shift. She claims that the utterance of "but" can achieve a similar effect, but is used where the speaker believes that the hearer believes that a discussion on the current topic will continue, and that presupposition needs to be countered. Each of these two actions will be defined in the planning system as operators. The "but" operator will have as an additional precondition that the hearer believes that the speaker's next utterance will be part of the current context. Both operators will have the effect that the hearer believes that the speaker is focusing on the previous topic of discussion.

Other operators that are available include explicity labeled shifts. This operator expands into planning an INFORM of a focus shift. The previous example of *Take John, for instance.* is an example of such an action.

The precise logical axiomitization of focusing and the precise definitions of each of these actions is a topic of current research. The point being made here is that these focusing actions can be specified formally. One goal of this research is to formally describe linguistic actions and other knowledge producing actions adequately enough to demonstrate the feasibility of a language planning system.

V. Current Status

The KAMP planner described in this paper is in the early stages of implementation. It can solve interesting problems in finding multiple agent plans, and plans involving acquiring and using knowlege. It has not been applied directly to language yet, but this is the next step in research.

Focusing actions need to be described formally, and critics have to be defined precisely and implemented. This work is currently in progress.

Although still in its early stages, this approach shows a great deal of promise for developing a computer system that is capable of producing utterances that approach the richness that is apparent in even the simplest human communication.

REFERENCES

[1] Appelt, Douglas, *A Planner for Reasoning about Knowledge and Belief,* Proceedings of the First Conference of the American Association for Artificial Intelligence, 1980.

[2] Austin, J., *How to Do Things with Words,* J. O. Urmson (ed.), Oxford University Press, 1962

[3] Cohen, Philip, *On Knowing What to Say: Planning Spech Acts,* Technical Report #118, University of Toronto, 1978

[4] Grice, H. P., *Logic and Coversation,* in Davidson, ed., The Logic of Grammar, Dickenson Publishing Co., Encino, California, 1975.

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

[6] Halliday, M. A. K., *Language Structure and Language Function,* in Lyons, ed., New Horizons in Linguistics.

[7] Halliday, M. A. K., Language as Social Semiotic, University Park Press, Baltimore, Md., 1978.

[8] Moore, Robert C., *Reasoning about Knowledge and Action,* Ph.D. thesis, Massachusetts Institute of Technology, 1979

[9] Reichman, Rachel, *Conversational Coherency,* Center for Research in Computing Technology Technical Report TR-17-78. Harvard University, 1978.

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

[11] Searle, John, Speech Acts, Cambridge Universiy Press, 1969

[12] Sidner, Candace L. *Toward a Computational Theory of Definite Anaphora Comprehension in English Discourse,* Massichusetts Institute of Technology Aritificial Intelligence Laboratory technical note TR-537, 1979.