PLANNING TEXT FOR ADVISORY DIALOGUES

Johanna D. Moore
UCLA Department of Computer Science
and
USC/Information Sciences Institute
4676 Admiralty Way
Marina del Rey, CA 90292-6695, USA

Cécile L. Paris
USC/Information Sciences Institute
4676 Admiralty Way
Marina del Rey, CA 90292-6695, USA

ABSTRACT
Explanation is an interactive process requiring a dialogue between advice-giver and advice-seeker. In this paper, we argue that in order to participate in a dialogue with its users, a generation system must be capable of reasoning about its own utterances and therefore must maintain a rich representation of the responses it produces. We present a text planner that constructs a detailed text plan, containing the intentional, attentional, and rhetorical structures of the text it generates.

INTRODUCTION
Providing explanations in an advisory situation is a highly interactive process, requiring a dialogue between advice-giver and advice-seeker (Pollack et al., 1982). Participating in a dialogue requires the ability to reason about previous responses, e.g., to interpret the user's follow-up questions in the context of the ongoing conversation and to determine how to clarify a response when necessary. To provide these capabilities, an explanation facility must understand what it was trying to convey and how that information was conveyed, i.e., the intentional structure behind the explanation, including the goal of the explanation as a whole, the subgoal(s) of individual parts of the explanation, and the rhetorical means used to achieve them.

Researchers in natural language understanding have recognized the need for such information. In their work on discourse analysis, Grosz and Sidner (1986) argue that it is necessary to represent the intentional structure, the attentional structure (knowledge about which aspects of a dialogue are in focus at each point), and the linguistic means of the discourse. In contrast, most text generation systems (with the notable exception of KAMP (Appelt, 1985)) have used only rhetorical and attentional information to produce coherent text (McKeown, 1985; McCoy, 1985; Paris, 1988b), omitting intentional information, or conflating intentional and rhetorical information (Hovy, 1988b). No text generation system records or reasons about the rhetorical, the attentional, as well as the intentional structures of the texts it produces.

In this paper, we argue that to successfully participate in an explanation dialogue, a generation system must maintain the kinds of information outlined by Grosz and Sidner as well as an explicit representation of the rhetorical structure of the texts it generates. We present a text planner that builds a detailed text plan, containing the intentional, attentional, and rhetorical structures of the responses it produces. The main focus of this paper is the plan language and the plan structure built by our system. Examples of how this structure is used in answering follow-up questions appear in (Moore and Swartout, 1989).

WHY A DETAILED TEXT PLAN?
In order to handle follow-up questions that may arise if the user does not fully understand a response given by the system, a generation facility must be able to determine what portion of the text failed to achieve its purpose. If the generation system only knows the top-level discourse goal that was being achieved by the text (e.g., persuade the hearer to perform an action), and not what effect the individual parts of the text were intended to have on the hearer and how they fit together to achieve this top-level goal, its only recourse is to use a different strategy to achieve the top-level goal. It is not able to re-explain or clarify any part of the explanation. There is thus a need for a text plan to contain a specification of the intended effect of individual parts of the text.
on the hearer and how the parts relate to one
another. We have developed a text planner
that records the following information about
the responses it produces:

- the information that Grosz and Sidner
  (1986) have presented as the basics of a
discourse structure:

  - **intentional structure:** a representa-
tion of the effect each part of
the text is intended to have on the
hearer and how the complete text
achieves the overall discourse pur-
pose (e.g., describe entity, persuade
hearer to perform an action).

  - **attentional structure:** information
about which objects, properties and
events are salient at each point
in the discourse. User's follow-
up questions are often ambiguous.
Information about the attentional
state of the discourse can be used
to disambiguate them (cf. (Moore
and Swartout, 1989)).

- in addition, for generation we require the
following:

  - **rhetorical structure:** an agent must
understand how each part of the
text relates rhetorically to the oth-
ers. This is necessary for linguis-
tic reasons (e.g., to generate the
appropriate clausal connectives in
multi-sentential responses) and for
responding to requests for elabora-
tion/clarification.

  - **assumption information:** advice-
giving systems must take knowl-
dge about their users into account.
However, since we cannot rely on
having complete user models, these
systems may have to make assump-
tions about the hearer in order to
use a particular explanation stra-
 tegy. Whenever such assumptions
are made, they must be recorded.

The next sections describe this new text plan-
er and show how it records the information
needed to engage in a dialogue. Finally, a brief
comparison with other approaches to text gen-
eration is presented.

**TEXT PLANNER**

The text planner has been developed as part
of an explanation facility for an expert sys-
tem built using the Explainable Expert Sys-
tems (EES) framework (Swartout and Smo-
lia, 1987). The text planner has been used
in two applications. In this paper, we draw
our examples from one of them, the Program
Enhancement Advisor (PEA) (Neches et al.,
1985). PEA is an advice-giving system in-
tended to aid users in improving their Com-
mon Lisp programs by recommending trans-
f ormations that enhance the user's code. The
user supplies PEA with a program and in-
dicates which characteristic of the program
should be enhanced (any combination of read-
ability, maintainability, and efficiency). PEA
then recommends transformations. After each
recommendation is made, the user is free to
ask questions about the recommendation.

We have implemented a top-down hier-
archical expansion planner (à la Sacerdoti
(1975)) that plans utterances to achieve dis-
course goals, building (and recording) the inten-
tional, attentional, and rhetorical struc-
ture of the generated text. In addition, since
the expert system explanation facility is in-
tended to be used by many different users,
the text planner takes knowledge about the
user into account. In our system, the user
model contains the user's domain goals and
the knowledge he is assumed to have about the
domain.

**THE PLAN LANGUAGE**

In our plan language, intentional goals are
represented in terms of the effects the speaker intents his utterance to have on the hearer.
Following Hovy (1988a), we use the terminol-
ogy for expressing beliefs developed by Cohen
and Levesque (1985) in their theory of rational
interaction, but have found the need to
extend the terminology to represent the types
of intentional goals necessary for the kinds of
responses desired in an advisory setting.
Although Cohen and Levesque have subse-
 quently retracted some aspects of their theory
of rational interaction (Cohen and Levesque,
1987), the utility of their notation for our pur-
poses remains unaffected, as argued in (Hovy,
1989). 2

---

1 PEA recommends transformations that improve
the 'style' of the user's code. It does not attempt to
understand the content of the user's program.

2 Space limitations prohibit an exposition of their
terminology in this paper. We provide English para-
phrases where necessary for clarity. (ENS 8 R x)
should be read as 'the speaker believes the speaker
and hearer mutually believe x.'
Rhetorical structure is represented in terms of the rhetorical relations defined in Rhetorical Structure Theory (RST) (Mann and Thompson, 1987), a descriptive theory characterizing text structure in terms of the relations that hold between parts of a text (e.g., CONTRAST, MOTIVATION). The definition of each RST relation includes constraints on the two entities being related as well as constraints on their combination, and a specification of the effect which the speaker is attempting to achieve on the hearer's beliefs. Although other researchers have categorized typical intersentential relations (e.g., (Grimes, 1975, Hobbs, 1978)), the set of relations proposed by RST is the most complete and the theory sufficiently detailed to be easily adapted for use in generation.

In our plan language, each plan operator consists of:

- **an effect**: a characterization of what goal(s) this operator can be used to achieve. An effect may be an intentional goal, such as persuade the hearer to do an action or a rhetorical relation, such as provide motivation for an action.

- **a constraint list**: a list of conditions that must be true before the operator can be applied. Constraints may refer to facts in the system's knowledge base or in the user model.

- **a nucleus**: the main topic to be expressed. The nucleus is either a primitive operator (i.e., speech acts such as inform, recommend and ask) or a goal (intentional or rhetorical) which must be further expanded. All operators must contain a nucleus.

- **satellites**: subgoal(s) that express additional information which may be needed to achieve the effect of the operator. When present, satellites may be specified as required or optional.

Examples of our plan operators are shown in Figures 1 and 2. The operator shown in Figure 1 can be used if the speaker (S) intends to persuade the hearer (H) to intend to do some act. This plan operator states that if an act is a step in achieving some domain goal(s) that the hearer shares, one way to persuade the hearer to do the act is to motivate the act in terms of those domain goals. Note that this plan operator takes into account not only the system's knowledge of itself, but also the system's knowledge about the user's goals, as embodied in a user model. If any domain goals that satisfy the constraints are found, this operator will cause the planner to post one or more MOTIVATION subgoals. This plan operator thus indicates that one way to achieve the intentional goal of persuading the hearer to perform an action is by using the rhetorical means MOTIVATION.
Plans that achieve intentional goals and those that achieve rhetorical relations are distinguished for two reasons: (1) so that the completed plan structure contains both the intentional goals of the speaker and the rhetorical means used to achieve them; (2) because there are many different rhetorical strategies for achieving any given intentional goal. For example, the system has several plan operators for achieving the intentional goal of describing a concept. It may describe a concept by stating its class membership and describing its attributes and its parts, by drawing an analogy to a similar concept, or by giving examples of the concept. There may also be many different plan operators for achieving a particular rhetorical strategy. (The planner employs selection heuristics for choosing among applicable operators in a given situation (Moore and Swartout, 1989).)

Our plan language allows both general and specific plans to be represented. For example, Figure 2 shows a plan operator for achieving the rhetorical relation MOTIVATION. This is a very specific operator that can be used only when the act to be motivated is a replacement (e.g., replace setq with setf). In this case, one strategy for motivating the act is to compare the object being replaced and the object that replaces it with respect to the domain goal being achieved. On the other hand, the operator shown in Figure 3 is general and can be used to achieve mutual belief of any assertion by first informing the hearer of the assertion and then, optionally, by persuading him of that fact. Because we allow very general operators as well as very specific ones, we can include both domain-independent and domain-dependent strategies.

A DETAILED EXAMPLE

Consider the sample dialogue with our system shown in Figure 4, in which the user indicates that he wishes to enhance the maintainability of his program. While enhancing maintainability, the system recommends that the user perform the act replace-I, namely 'replace setq with setf', and thus posts the intentional goal (BMB S H (GOAL H Eventually(DONE H replace-I))). This discourse goal says that the speaker would like to achieve the state where the speaker believes that the hearer and speaker mutually believe that the replacement eventually be done by the hearer.

The planner then identifies all the operators whose effect field matches the discourse goal to be achieved. For each operator found, the planner checks to see if all of its constraints are satisfied. In doing so, the text planner attempts to find variable bindings in the expert system's knowledge base or the user model that satisfy all the constraints in
**Figure 5:** High-level Plan Operator for Recommending an Act

```
EFFECT: (BMB S H (GOAL H Eventually(DONE H ?act)))
CONSTRAINTS: none
NUCLEUS: (RECOMMEND S H ?act)
SATELLITES: (((BMB S H (COMPETENT H (DONE H ?act))) *optional*)
  ((PERSUADE S H (GOAL H Eventually(DONE H ?act))) *optional*))
```

**Figure 6:** System goals leading to replace `setq` with `setf`

```
apply-SETQ-to-SETF-transformation
apply-local-transformations-whose-rhs-use-is-more-general-than-lhs-use
apply-local-transformations-that-enhance-maintainability
enhance-maintainability
enhance-program
```

---

the constraint list. Those operators whose constraints are satisfied become candidates for achieving the goal, and the planner chooses one based on: the user model, the dialogue history, the specificity of the plan operator, and whether or not assumptions about the user’s beliefs must be made in order to satisfy the operator’s constraints.

Continuing the example, the current discourse goal is to achieve the state where it is mutually believed by the speaker and hearer that the hearer has the goal of eventually executing the replacement. This discourse goal can be achieved by the plan operator in Figure 5. This operator has no constraints. Assume it is chosen in this case. The nucleus is expanded first, causing (RECOMMEND S H replace-1) to be posted as a subgoal. RECOMMEND is a primitive operator, and so expansion of this branch of the plan is complete.

Next, the planner must expand the satellites. Since both satellites are optional in this case, the planner must decide which, if any, are to be posted as subgoals. In this example, the first satellite will not be expanded because the user model indicates that the user is capable of performing replacement acts. The second satellite is expanded, posting the intentional subgoal to persuade the user to perform the replacement. A plan operator for achieving this goal using the rhetorical relation MOTIVATION was shown in Figure 1.

When attempting to satisfy the constraints of the operator in Figure 1, the system first checks the constraints (GOAL S ?domain-goal) and (STEP replace-1 ?domain-goal). These constraints state that, in order to use this operator, the system must find an expert system goal, ?domain-goal, that replace-1 is a step in achieving.

This results in several possible bindings for the variable ?domain-goal. In this case, the applicable system goals, listed in order from most specific to the top-level goal of the system, are shown in Figure 6.

The last constraint of this plan operator, (EMB S H (GOAL H ?domain-goal)), is a constraint on the user model stating that the speaker and hearer should mutually believe that ?domain-goal is a goal of the hearer. Not all of the bindings found so far will satisfy this constraint. Those which do not will not be rejected immediately, however, as we do not assume that the user model is complete. Instead, they will be noted as possible bindings, and each will be marked to indicate that, if this binding is used, an assumption is being made, namely that the binding of...
?domain-goal is assumed to be a goal of the user.

In this example, since the user is using the system to enhance a program and has indicated that he wishes to enhance the maintainability of the program, the system infers the user shares the top-level goal of the system (enhance-program), as well as the more specific goal enhance-maintainability. Therefore, these are the two goals that satisfy the constraints of the operator shown in Figure 1.

The text planner prefers choosing binding environments that require no assumptions to be made. In addition, in order to avoid explaining parts of the reasoning chain that the user is familiar with, the most specific goal is chosen. The plan operator is thus instantiated with enhance-maintainability as the binding for the variable ?domain-goal. The selected plan operator is recorded as such, and all other candidate operators are recorded as untried alternatives.

The nucleus of the chosen plan operator is now posted, resulting in the subgoal (MOTIVATION replace-1 enhance-maintainability). The plan operator chosen for achieving this goal is the one that was shown in Figure 2. This operator motivates the replacement by describing differences between the object being replaced and the object replacing it. Although there are many differences betweensetq and setf, only the differences relevant to the domain goal at hand (enhance-maintainability) should be expressed. The relevant differences are determined in the following way. From the expert system's problem-solving knowledge, the planner determines what roles setq and setf play in achieving the goal enhance-maintainability. In this case, the system is enhancing maintainability by applying transformations that replace a specific construct with one that has a more general usage. Setq has a more specific usage than setf, and thus the comparison between setq and setf should be based on the generality of their usage.

Finally, since the term generalized-variable has been introduced, and the user model indicates that the user does not know this term, an intentional goal to define it is posted: (BMB S H (KNOW H generalized-variable)). This goal is achieved with a plan operator that describes concepts by stating their class membership.
and describing their attributes. Once completed, the text plan is recorded in the dialogue history. The completed text plan for response (3) of the sample dialogue is shown in Figure 7.

ADVANTAGES

As illustrated in Figure 7, a text plan produced by our planner provides a detailed representation of the text generated by the system, indicating which purposes different parts of the text serve, the rhetorical means used to achieve them, and how parts of the plan are related to each other. The text plan also contains the assumptions that were made during planning. This text plan thus contains both the intentional structure and the rhetorical structure of the generated text. From this tree, the dominance and satisfaction-precedence relationships as defined by Grosz and Sidner can be inferred. Intentional goals higher up in the tree dominate those lower down and a left to right traversal of the tree provides satisfaction-precedence ordering. The attentional structure of the generated text can also be derived from the text plan. The text plan records the order in which topics appear in the explanation. The global variable $\text{local-context}$ always points to the plan node that is currently in focus, and previously focused topics can be derived by an upward traversal of the plan tree.

The information contained in the text plan is necessary for a generation system to be able to answer follow-up questions in context. Follow-up questions are likely to refer to the previously generated text, and, in addition, they often refer to part of the generated text, as opposed to the whole text. Without an explicit representation of the intentional structure of the text, a system cannot recognize that a follow-up question refers to a portion of the text already generated. Even if the system realizes that the follow-up question refers back to the original text, it cannot plan a text to clarify a part of the text, as it no longer knows what were the intentions behind various pieces of the text.

Consider again the dialogue in Figure 4. When the user asks 'What is a generalized variable?' (utterance (4) in Figure 4), the query analyzer interprets this question and posts the goal: (BMB S H (KNOW H generalized-variable)). At this point, the explainer must recognize that this discourse goal was attempted and not achieved by the last sentence of the previous explanation. Failure to do so would lead to simply repeating the description of a generalized variable that the user did not understand. By examining the text plan of the previous explanation recorded in the dialogue history, the explainer is able to determine whether the current goal (resulting from the follow-up question) is a goal that was attempted and failed, as it is in this case. This time, when attempting to achieve the goal, the planner must select an alternative strategy. Moore (1989b) has devised recovery heuristics for selecting an alternative strategy when responding to such follow-up questions. Providing an alternative explanation would not be possible without the explicit representation of the intentional structure of the generated text. Note that it is important to record the rhetorical structure as well, so that the text planner can choose an alternative rhetorical strategy for achieving the goal. In the example under consideration, the recovery heuristics indicate that the rhetorical strategy of giving examples should be chosen.

RELATED WORK

Schemata (McKeown, 1985) encode standard patterns of discourse structure, but do not include knowledge of how the various parts of a schema relate to one another or what their intended effect on the hearer is. A schema can be viewed as a compiled version of one of our text plans in which all of the non-terminal nodes have been pruned out and only the leaves (the speech acts) remain. While schemata can produce the same initial behavior as one of our text plans, all of the rationale for that behavior has been compiled out. Thus schemata cannot be used to participate in dialogues. If the user indicates that he has not understood the explanation, the system cannot know which part of the schema failed to achieve its effect on the hearer or which rhetorical strategy failed to achieve this effect. Planning a text using our approach is essentially planning a schema from more finely-grained plan operators. From a library of such plan operators, many varied schemata can result, improving the flexibility of the system.

In an approach taken by Cohen and Appelt (1979) and Appelt (1985), text is planned by reasoning about the beliefs of the hearer and speaker and the effects of surface speech

---

4 We are also currently implementing another interface which allows users to use a mouse to point at the noun phrases or clauses in the text that were not understood (Moore, 1989b).
acts on these beliefs (i.e., the intentional effect). This approach does not include rhetorical knowledge about how clausal units may be combined into larger bodies of coherent text to achieve a speaker's goals. It assumes that appropriate axioms could be added to generate large (more than one- or two-sentence) bodies of text and that the text produced will be coherent as a by-product of the planning process. However, this has not been demonstrated.

Recently, Hovy (1988b) built a text structure which produces a coherent text when given a set of inputs to express. Hovy uses an opportunistic planning approach that orders the inputs according to the constraints on the rhetorical relations defined in Rhetorical Structure Theory. His approach provides a description of what can be said when, but does not include information about why this information can or should be included at a particular point. Hovy's approach conflates intentional and rhetorical structure and, therefore, a system using his approach could not later reason about which rhetorical strategies were used to achieve intentional goals.

STATUS AND FUTURE WORK

The text planner presented is implemented in Common Lisp and can produce the text plans necessary to participate in the sample dialogue described in this paper and several others (see (Moore, 1989a, Paris, 1988a)). We currently have over 60 plan operators and the system can answer the following types of (follow-up) questions:

- Why?
- Why conclusion?
- Why are you trying to achieve goal?
- Why are you using method to achieve goal?
- Why are you doing act?
- How do you achieve goal?
- How did you achieve goal (in this case)?
- What is a concept?
- What is the difference between concept1 and concept2?
- Huh?

The text planning system described in this paper is being incorporated into two expert systems currently under development. These systems will be installed and used in the field. This will give us an opportunity to evaluate the techniques proposed here.

We are currently studying how the attentional structure inherent in our text plans can be used to guide the realization process, for example in the planning of referring expressions and the use of cue phrases and pronouns. We are also investigating criteria for the expansion and ordering of optional satellites in our plan operators. Currently we use information from the user model to dictate whether or not optional satellites are expanded, and their ordering is specified in each plan operator. We wish to extend our criteria for satellite expansion to include other factors such as pragmatic and stylistic goals (Hovy, 1988a) (e.g., brevity) and the conversation that has occurred so far. We are also investigating the use of attentional information to control the ordering of these satellites (McKeown, 1985). We also believe that the detailed text plan constructed by our planner will allow a system to modify its strategies based on experience (feedback from the user). In (Paris, 1988a), we outline our preliminary ideas on this issue. We have also begun to study how our planner can be used to handle incremental generation of texts. In (Moore, 1988), we argue that the detailed representation provided by our text plans is necessary for execution monitoring and to indicate points in the planning process where feedback from the user may be helpful in incremental text planning.

CONCLUSIONS

In this paper, we have presented a text planner that builds a detailed text plan, containing the intentional, attentional, and rhetorical structures of the responses it produces. We argued that, in order to participate in a dialogue with its users, a generation system must be capable of reasoning about its past utterances. The text plans built by our text planner provide a generator with the information needed to reason about its responses. We illustrated these points with a sample dialogue.

REFERENCES


guistics, pages 49–60, University of Chicago, Chicago, Illinois, July 8-12.


