Title
Constructing Coherent Text using Rhetorical Relations

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 10(0)

Authors
Moore, Johanna D.
Paris, Cecile L.

Publication Date
1988

Peer reviewed

Constructing Coherent Text using Rhetorical Relations*

Johanna D. Moore
UCLA Department of Computer Science
and
USC/Information Sciences Institute

Cecile L. Paris
USC/Information Sciences Institute

Introduction

Language is planned behavior: speakers produce complex combinations of utterances to cause certain effects on their hearers, such as persuading them, informing them or requesting something from them [Aus75], [Gri75], [CP79], [AP80], [App81], [CL85]. Researchers in the field of text generation have concentrated their efforts on producing a computational model of the process of generating coherent text to achieve a speaker's conversational goals. But producing text is only a small part of a much broader cognitive faculty of language which includes such tasks as responding to follow-up questions, elaborating or clarifying texts that were not completely understood by the hearer, and learning rhetorical strategies to achieve conversational goals. To perform these other tasks, it is not sufficient to simply produce coherent text. An agent must understand the text it produces in terms of how each part of the text relates to the others and how the complete text achieves its goal.

Computational models of text generation put forth to date fall into two extremes: *plan*-based approaches and *schema*-based approaches. Using a plan-based approach [CP79] [App81], a text plan is produced by reasoning about the beliefs of the hearer and speaker and the effects of various speech acts (e.g., REQUEST, INFORM) on these beliefs. This approach thus captures the intended effects of an utterance on the hearer. However, it does not include rhetorical knowledge which is knowledge about how to combine individual speech acts into larger pieces of coherent text. As a result, systems using the plan-based approach cannot produce texts to describe objects, compare and contrast two entities, explain how a device works, or justify a result.

To produce larger bodies of text, researchers have turned to schema-based approaches (e.g., [McK82], [McC85], [Par87]) which employ script-like structures (schemata) to generate coherent multi-sentential texts achieving a given discourse goal. Schemata encode standard patterns of discourse using rhetorical predicates, but do not include knowledge of how the various parts of the schema relate to one another or what their intended effect on the hearer is. As a result, if the hearer does not understand the text, it is very hard to recover, i.e., to know which part of the schema failed to achieve its effect on the hearer and produce text to correct the misunderstanding.

Recently, Hovy has begun to incorporate rhetorical relations into the plan-based approach to produce multi-sentential texts, but has concentrated solely on ordering a set of propositions [Hov88]. His system assumes that the information to be said is chosen in advance and given to the system as input. We believe that the tasks of choosing what to say and how to organize it cannot be divided in this way: they are intertwined and can influence one another. The choice of what to say next is in fact largely dependent on what has already been said.

In this work, we are concerned with generating coherent multi-sentential texts, deciding both what to say and how to organize it, in such a way that a system can recover from failure. We want our generation system to explicitly record how the different parts of the text are related and how they affect the hearer's beliefs. To this end, we propose a plan structure which relates discourse goals, expressed in terms of their intended effects on the hearer, to the rhetorical means used to achieve them. Rhetorical means might be low

*The research described in this paper was supported under DARPA Grant #MDA 903-81-C-0335. The authors would like to thank Eduard Hovy and William Swartout for their comments and suggestions on this paper.

level speech acts, such as INFORM, or rhetorical relations, as defined in Rhetorical Structure Theory (RST) [MT87], such as MOTIVATION.

This paper presents the structure of the text planning operators in our model and the planning mechanism that uses these operators to construct text. To illustrate how a text plan is produced, we show an example. We discuss the advantages of this approach and show how it can be seen as a computational model that unifies both the schema-based and the plan-based approaches presented to date.

The Plan Language and the Planning Mechanism

The planning mechanism we use is a conventional top-down hierarchical expansion planner [Sac75]. The planner begins with a high-level discourse goal of the speaker and successively refines it into other discourse subgoals or rhetorical means for achieving them. This process continues until the entire plan is refined into primitive operators, i.e., speech acts. Discourse goals are represented in terms of the effects that the speaker intends his or her utterance to have on the hearer's knowledge. Following Hovy in [Hov88] we have adopted the terminology for expressing beliefs developed by Cohen and Levesque in their theory of rational interaction [CL85].[1]

There are two types of plan operators in our system: operators whose effects are characterizations of the hearer's beliefs, and operators that achieve a particular rhetorical relation. Figure 1 shows a plan operator whose effect is the state where the speaker believes that the hearer and speaker mutually believe that it is a goal of the hearer that a given act eventually be done by the hearer. Figure 2 shows a text plan capable of achieving the rhetorical relation MOTIVATION which relates an act to the goal it achieves. When instantiated, the plan will inform the hearer that the act is a step towards achieving the goal and may elaborate as to why this is the case.

Each plan operator consists of:

- **an effect:** a characterization of what goal(s) this operator can be used to achieve.
- **a constraint list:** a list of conditions that must be true before the operator can be applied. Constraints may refer to facts or relations in the domain of discourse or states of the hearer's knowledge. Some of the constraints correspond to the constraints on a nucleus and satellite as specified in RST. Additional constraints had to be introduced in order to actually *construct* text, as RST was developed as a tool for *analyzing* text.
- **a nucleus and a satellite:** each is a partially ordered sequence of operators that, when executed, may achieve the effects of this operator. The nucleus is the focus of the text produced by this operator and must be present. It could be a speech act or a desired state, which can later be expanded. Satellites are additional parts of the plan. They may be required or optional.

When a goal is posted, the planner finds all the plan operators whose effects match the goal and whose constraints are satisfied. It selects one such operator and posts the steps listed in the nucleus and satellite as subgoals to be satisfied. The completed text plan is given as input to a text generation system, Penman in our case [MM83].

An Example

One of the prototype systems we are building is an explanation component for an expert system that aids users in improving their LISP code by recommending transformations that enhance the user's code along the dimensions of readability, maintainability, or efficiency.[2] The most natural way for a dialogue to begin is with a recommendation by the system. The user is then free to ask questions about this recommendation.

Suppose for example that the user wanted to enhance the readability of the program under consideration. The expert system might wish to recommend that the user replace calls to the function CAR with calls to the function FIRST. To do so, the expert system would post the following discourse goal to the text planner:

(BMB S H (GOAL H Eventually(DONE H replace-1)))

[1]Space limitations prohibit an exposition of their terminology in this paper. For clarity, we will paraphrase the terminology in English in our examples.

[2]Details regarding the expert system can be found in [NSM85].

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

Figure 1: High-level Text Plan for Recommending an Act

EFFECT: (MOTIVATION ?act ?goal)
CONSTRAINTS: ((GOAL S ?goal) and
(BMB S H (GOAL H ?goal)) and
(STEP ?act ?goal))
NUCLEUS: (INFORM S H (STEP ?act ?goal))
SATELLITE: (BMB S H (STEP ?act ?goal)) *optional

Figure 2: Text Plan for Motivation

where
replace-1 = (replace (ACTOR user) (REPLACEE car-function) (REPLACER first-function))

This goal says that the speaker would like to achieve the state where the speaker believes that the hearer and speaker mutually believe that it is a goal of the hearer that the replacement eventually be done by the hearer. One of the plans for achieving this goal is shown in Figure 1. This plan operator has no constraints on its application. The nucleus becomes a subgoal to achieve a state in which the hearer and speaker mutually believe that it is a goal of the speaker that the hearer do this act. This subgoal will be achieved by the plan operator shown in Figure 3. RECOMMEND is a primitive speech act, so this node will be a leaf in the completed text plan.

Satellites of a plan operator capture the kind of discourse patterns that were captured in the schema-based approach and indicate other information that could be said to give a more elaborated text. In this plan operator, the first satellite corresponds to making the hearer competent to perform the act by telling him or her anything that he or she needs to know to do so. The second satellite calls for persuading the hearer to perform the act.

The satellites of the high-level plan in Figure 1 are both marked "optional," indicating that it would be sufficient to simply state the recommendation. The planner could choose not to expand any of the satellites and await feedback from the hearer. In this case, the system would simply recommend that the hearer perform the action. If the hearer is not satisfied with this text and indicates this by asking a question, such as "why?", the system can examine the recorded plan to find optional parts of the plan that were not expanded and determine if any of these could be used in answering the question. For example, the goal of persuading the hearer would be posted in response to the hearer's "why?".

Alternatively, the planner could choose to post any of the optional satellites as subgoals. In the context of this expert system, we assume that the user is competent to perform the recommended transformation. Therefore, the first satellite will not cause any text to be generated. Suppose the second optional satellite is

EFFECT: (BMB S H (GOAL S (GOAL H Ev(DONE H ?act)))
CONSTRAINTS: none
NUCLEUS: (RECOMMEND S H ?act)
SATELLITES: none

Figure 3: Text Plan for Recommending an Act

```
EFFECT: (PERSUADE H (GOAL H Ev(DONE H ?act)))
CONSTRAINTS: G = the set of g s.t. ((BMB S H (GOAL H g)) and (STEP ?act g))
NUCLEUS: loop for g in G
             (MOTIVATION ?act g)
SATELLITES: none
```

Figure 4: Text Plan for Persuading by Motivating an Act

posted as a subgoal. The planner must then find rhetorical means for achieving the discourse goal "persuade." One of the plans for persuading a hearer to do an act is shown in Figure 4. This plan has a constraint which requires finding a non-empty set of goals which are mutually believed to be goals of the hearer and which the recommended act is a step towards achieving. This information may be retrieved from the expert system's problem-solving knowledge and the user model.

Constraints referring to the hearer's beliefs are checked against a user model which represents the state of the hearer's knowledge. We do not, however, wish to depend on this user model being either complete or correct. Furthermore, since we are assuming a highly interactive model of conversation in which we rely on the feedback from the hearer to indicate lack of understanding, the planner may assume that a constraint on the hearer's knowledge is true even if it is not explicitly indicated in the user model, keeping track of any such assumptions made. If the hearer indicates dissatisfaction with a text, the system can examine the recorded text plan to determine what assumptions could have been erroneous. It can then re-plan the text using a plan operator that does not require these assumptions or by providing the hearer with the additional knowledge necessary to make these constraints true. If more verbose text is desired (e.g., because the system has information that the hearer is a novice), the planner may immediately plan text to make the constraint true.

In this example, there is one goal which satisfies the constraint, namely ENHANCE-1, the goal to enhance the readability of the program. The planner thus posts a single subgoal:

(MOTIVATION replace-1 enhance-1)

Motivation is a rhetorical relation defined in RST. Note that this plan operator combined with the operator which satisfies the MOTIVATION subgoal shown in Figure 2, operationalize this rhetorical relation. We have added the necessary constraints to indicate when this relation can be used in constructing a text, indicating how and what to look for in the knowledge base in order to use this rhetorical strategy to achieve this discourse goal. For the sake of brevity, we will not go through the rest of this example in detail here. Figure 5 presents the final text plan for achieving the original high-level discourse goal. Once the text plan is completed, it is transformed into a representation suitable as input for the Penman text generation system. The text that will be generated is:[3]

> You should replace (CAR x) with (FIRST x) because that will enhance the readability of the program. To enhance the readability of the program, the system applies readability enhancing transformations. CAR-to-FIRST is a readability enhancing transformation because its left-hand-side is a function whose function name is a technical word and its right-hand-side is a function whose function name is an English word.

Note in Figure 5 that the rhetorical relations provide a context for choosing appropriate cue words to link the different parts of the text when realizing the text in English [Hov88]. For example, the two pieces of text linked with the rhetorical relation MOTIVATION, are connected with the cue word "because".

Advantages of this Approach

The Need for a Detailed Text Plan

As we have seen in the example, a text plan produced using this method provides a detailed representation of the text produced by the system, indicating how parts of the plan are related and which purposes different

[3]The implementation of the text planner is not yet complete. However, we expect to complete it by the time of the conference.

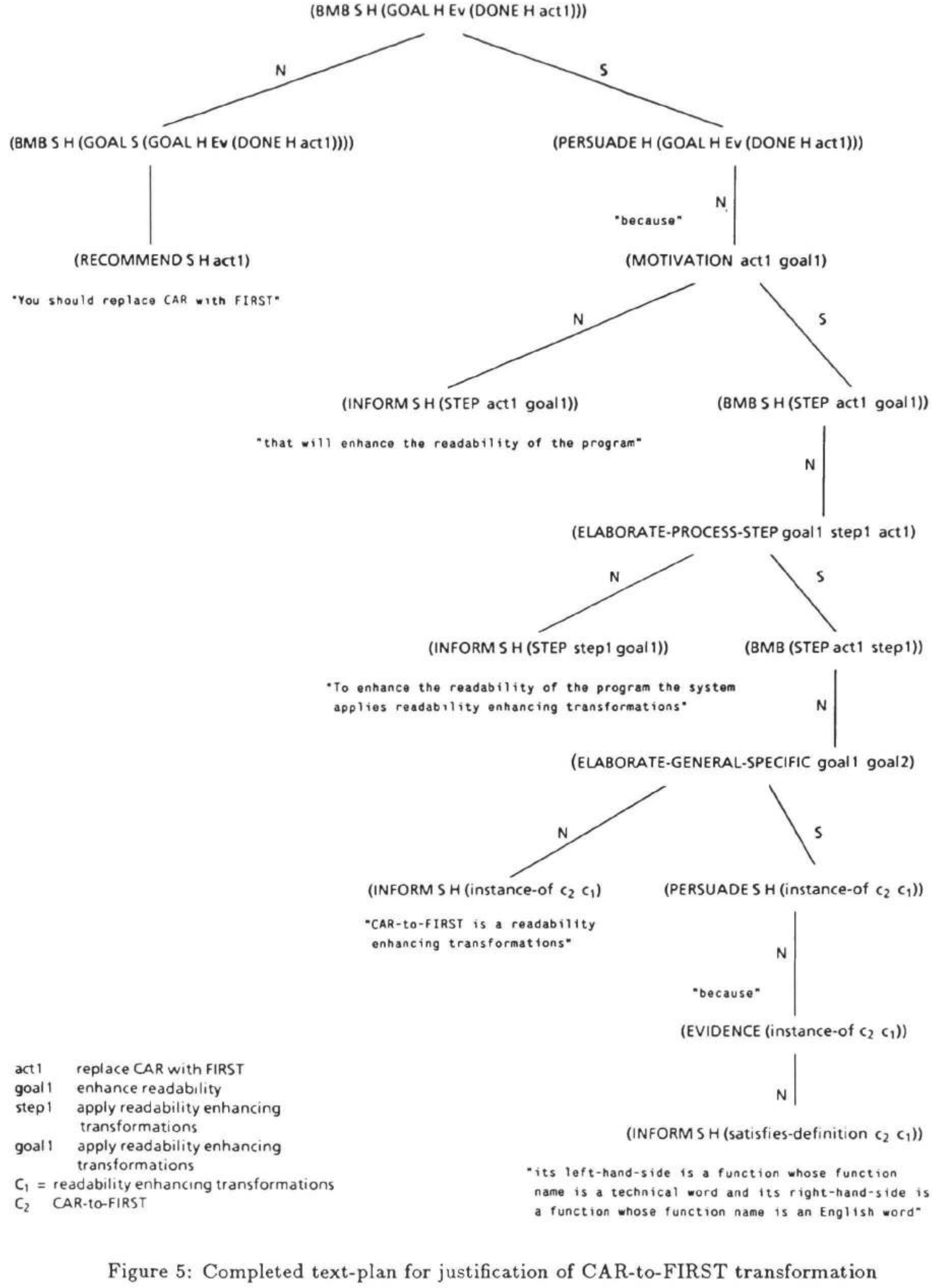

Figure 5: Completed text-plan for justification of CAR-to-FIRST transformation

parts of the generated text serve. This structure can be useful in aiding several of the cognitive tasks that have thus far been difficult for natural language generation systems, such as disambiguating follow-up questions, selecting perspective when describing or comparing objects, and providing further explanations.[4] In addition, the completed text plan can be used to update a user model upon receiving follow-up questions, as it is possible to examine the completed plan to determine whether information taken from the user model might have been inaccurate.

Unifying the Schema-based and Plan-based Approaches

Both the schema-based approach and the plan-based approach are useful in a text generation system: the plan-based approach is necessary to reason about the hearer's beliefs to produce an utterance the hearer will understand; and the schema-based approach is useful in situations where standard patterns of discourse can be identified. A generation system should thus have access to both of these kinds of text plans. In fact, instead of being two very different approaches, these two approaches should be two extremes of a continuum, where a schema can be viewed as a compiled version of a text-plan, with all the non-terminal nodes pruned out and only the leaves remaining. They can produce the same behavior, but all of the rationale for the behavior has been compiled out.[5] Because of this "compilation," a schema is computationally efficient. It may be necessary however to recover the information which has been compiled out in order to handle some discourse situations.

We believe that the approach proposed in this paper provides a computational model that unifies the schema-based and the plan-based approaches. The plan structure proposed can express the whole spectrum of plans: from primitive plans for speech acts to plans that achieve very high-level discourse goals and resemble schemata. Once a plan for achieving a high-level discourse goal has been constructed, it will be possible to use its "compiled" form (just the leaves) for the sake of computational efficiency. By keeping the detailed plan that resulted in this schema, the system will still be able to recover from failure. Most plans, however, like the ones presented in our examples, can be considered as a middle ground between the schemata and the plans that were used in previous plan-based approaches.

Future Work

Many issues remain to be addressed. In particular, we must develop a set of heuristics for plan selection as well as for deciding when an optional satellite should be expanded. We must also devise a control mechanism for these heurictics.

We have mentioned that the rhetorical relations used to form a plan can provide information as to which cue word is appropriate. Rhetorical relations may relate paragraphs as well as simple propositions. The planner must recognize this fact and act upon it in the appropriate way to ensure that the resulting text is easy to read: for example, the current focus might have to be reintroduced before a new proposition is presented. We will investigate the use of *critics* that will examine a plan for global interactions and make decisions about how to realize a large text plan in understandable English.

We are also investigating the possibility of remembering "good" plans, namely plans that were successful in presenting the information to the user without triggering a flock of follow-up questions. In a similar vein, we are looking into "learning" plans for achieving high-level discourse goals, based on the information presented to the user and follow-up questions asked. We would like the system to be able to learn schemata that are used often.

Conclusions

In this paper, we proposed a computational model for text planning which relates effects on the hearer's beliefs to the rhetorical means that can be used to achieve these effects. As a result, we are operationalizing Rhetorical Structure Theory for constructing text. We have argued that a detailed text plan which explicitly records the rhetorical means used to construct the text is necessary in order to respond to follow-up questions. The text planner proposed can be used to generate either verbose texts or answers to follow-up questions, in a more interactive situation.

[4] See [MS88] for more details about a system that uses such a plan to do these tasks.

[5] Note that this is similar to Swartout's argument in the XPLAIN system [Swa83], where Swartout points out that having implicit knowledge is usually not enough: one might need to recover the knowledge omitted in a compiled form.

References

[AP80] James F. Allen and C. Raymond Perrault. Analyzing intention in utterances. *Artificial Intelligence*, 15:143–178, 1980.

[App81] Douglas E. Appelt. *Planning Natural Language Utterances to Satisfy Multiple Goals*. PhD thesis, Stanford University, 1981.

[Aus75] J. L. Austin. *How to Do Things With Words*. Harvard University Press, Cambridge, MA., 1975. Second Edition.

[CL85] Philip R. Cohen and Hector J. Levesque. Speech acts and rationality. In *Proceedings of the Twenty-Third Annual Meeting of the Association for Computational Linguistics*, pages 49–60, University of Chicago, Chicago, Illinois, July 8-12 1985.

[CP79] Philip R. Cohen and C. Raymond Perrault. Elements of a plan-based theory of speech acts. *Cognitive Science*, 3:177–212, 1979.

[Gri75] Joseph E. Grimes. *The Thread of Discourse*. Mouton, The Hague, Paris, 1975.

[Hov88] Eduard H. Hovy. Planning coherent multisentential text. In *Proceedings of the Twenty-Sixth Annual Meeting of the Association for Computational Linguistics*, State University of New York, Buffalo, New York, 7-10 June 1988.

[McC85] Kathleen F. McCoy. *Correcting Object-Related Misconceptions*. PhD thesis, University of Pennsylvania, December 1985. Published by University of Pennsylvania as Technical Report MS-CIS-85-57.

[McK82] Kathleen R. McKeown. *Generating Natural Language Text in Response to Questions About Database Structure*. PhD thesis, University of Pennsylvania, 1982. Published by University of Pennsylvania as Technical Report MS-CIS-82-5.

[MM83] William C. Mann and Christian Matthiessen. *Nigel: A Systemic Grammar for Text Generation*. Technical Report RR-83-105, USC/Information Sciences Institute, February 1983.

[MS88] Johanna D. Moore and William R. Swartout. A reactive approach to explanation. 18-21 July 1988. To be presented at the Fourth International Workshop on Natural Language Generation.

[MT87] William C. Mann and Sandra A. Thompson. Rhetorical structure theory: a theory of text organization. In Livia Polanyi, editor, *The Structure of Discourse*, Ablex Publishing Corporation, Norwood, N.J., 1987. Also available as USC/Information Sciences Institute Technical Report Number RS-87-190.

[NSM85] Robert Neches, William R. Swartout, and Johanna D. Moore. Enhanced maintenance and explanation of expert systems through explicit models of their development. *IEEE Transactions on Software Engineering*, SE-11(11), November 1985.

[Par87] Cecile L. Paris. *The Use of Explicit User Models in Text Generation: Tailoring to a User's Level of Expertise*. PhD thesis, Columbia University, October 1987.

[Sac75] Earl D. Sacerdoti. *A structure for plans and behavior*. Technical Report TN-109, SRI, 1975.

[Swa83] William R. Swartout. XPLAIN: a system for creating and explaining expert consulting systems. *Artificial Intelligence*, 21(3):285–325, September 1983. Also available as ISI/RS-83-4.