THE GENERATION OF HIGH-LEVEL STRUCTURE FOR EXTENDED EXPLANATIONS

David J. Mooney (mooney@dewey.udel.edu)
Sandra Carberry (carberry@dewey.udel.edu)
Kathleen F. McCoy (mccoy@dewey.udel.edu)

Department of Computer and Information Sciences
University of Delaware
Newark, Delaware 19716

ABSTRACT

This paper analyzes the structural features of naturally-occurring extended explanations and argues that current generation methodologies are inadequate for determining high-level structure. It presents a computational model based on the hypothesis that high-level structure - composed of a *unifying framework* and its associated *basic blocks* - can be determined by bottom-up processes that attempt to satisfy speaker, listener, and *compositional* goals, after which top-down strategies can be used to organize the material about the selected framework.

1 INTRODUCTION

In this paper, we describe the structural characteristics of extended, planned[1] explanations involving complex physical devices and present a computational model for generating such explanations. Our investigation suggests that the organizational strategies currently employed for structuring short explanations are inadequate for generating the high-level structure characteristic of that found in naturally-occurring extended explanations, which typically require several pages of text. Our computational model is based on the hypothesis that text structure is not completely recursive as others have claimed ([GS86], [Rei78], [Pol86], [MT88]), but rather that the high-level structure of extended explanations is determined by processes separate from those which organize text at lower levels.

Section 2.1 provides a brief overview of current models for structuring text, followed by a description of the *basic block*, the unit of discourse on which our model is based, in Section 2.2. Section 2.3 describes the characteristics of high-level structure of extended explanations, followed by a description of our strategy for generating this structure in Section 3; a complete description is contained in [MCM89].

[1]Emphasis is placed on *planned* to distinguish these explanations from discourse in which the material is developed mutually by the participants as the discourse progresses.

2 THE BASIC BLOCK MODEL OF EXTENDED DISCOURSE

In order to generate an extended explanation, a natural language system must determine the basic content to be conveyed; the next step is to cohesively organize this material. As anyone who has had to organize large amounts of information into a coherent text can attest, there are many possible combinations of that material, some more cohesive than others. Frequently, deciding *how* to organize a large body of material is more difficult than determining *what* to include. Our research is concerned with the identification of a coherent unifying framework about which an extended explanation can be organized and the criteria for selecting from among several frameworks when more than one viable alternative exists.

2.1 OTHER APPROACHES TO TEXT STRUCTURE

2.1.1 COMPUTATIONAL APPROACHES

A number of researchers (e.g., [GS86], [MT88], [Rei78], [Pol86]) have argued that discourse is composed of hierarchically structured segments and that this structure is completely recursive in nature. Two general methodologies have been applied to the structuring of explanations: schemas ([McK85], [McC85], [Par87]) and rhetorical structure theory (RST) ([HM89], [Hov88], [MP88], [MS88]).

A schema is a discourse strategy that captures a typical pattern of discourse associated with a particular discourse purpose, (e.g., providing an analogy or evidence). Schemas can be thought of as templates composed of an ordered sequence of *rhetorical predicates*, which

> "characterize the predicating acts a speaker may use and delineate the structural relation between propositions in a text."[2]

[2]From [McK85] page 9

These predicates are intended to capture the structural relations that hold between clauses in a text. The predicates are used recursively, capturing the structure of text at any level.

RST, developed by Mann and Thompson ([MT88]), was originally a tool for the analysis of text. RST claims that, except for a small number of highly-stylized forms, all coherent texts have an RST decomposition. RST posits a small number of relations, comparable to McKeown's rhetorical predicates, that exist between segments of text. Because each relation has associated with it well-defined intended effects and conditions necessary for it to hold, RST lends itself well to a generation methodology based on a top-down, hierarchical planning formalism ([Sac77]). Thus, like McKeown's rhetorical predicates, RST claims to account for the structure of text at any level of the discourse hierarchy.

While these methods have proven to be effective for organizing *short* pieces of text, we maintain that they are inadequate for generating the characteristic structure of *extended* explanations at the level of the *primary segments,* which occupy the first level of the discourse hierarchy. We contend that the characteristics exhibited by the primary segments of extended explanations, to be described in the next section, cannot be captured by recursive processes. Rather, we maintain that high-level structure must be generated by a separate, bottom-up process, after which recursive organizational strategies can be applied at lower levels.

2.1.2 RHETORIC

Rhetoric, the formal study of the art of good writing, provides general strategies for organizing text at a high level that are absent from the computational models. Analysis - "the method of explanation whereby a subject is divided into its separate component parts"[3] - is possibly the most instrumental of these strategies. There are no hard-and-fast rules for determining what constitutes an appropriate analysis of a subject. As [WA60] observes, a subject may be classified in as many ways as it has characteristics/parts/stages/etc. However, there are three criteria which experts ([WA60], [Are75], [Tho57], [Dan67], [KP66]) mutually consider essential for a satisfactory organizational strategy:

1. The scheme should be logical; a single, consistent criterion should be used for the analysis (e.g. time, steps in a process).

2. The scheme should exhaust all of the possibilities; everything to be conveyed should be encompassed by the scheme.

3. The resultant categories should be mutually exclusive; nothing should belong to more than one.

[3] [Are75] page 107

While the type of explanation with which this paper is concerned exhibits a high-level organization reflective of these criteria, the criteria by themselves do not provide the specificity necessary for computational generation. These guidelines include no suggestions for dealing with situations in which no logical, all-inclusive framework can be identified, nor do they offer suggestions for selecting among several organizational schemes which meet the prescribed criteria equally well. Furthermore, the guidelines are not sufficient in-and-of themselves to account for all of the observed phenomena discussed in the following sections.

2.2 BASIC BLOCKS

Our model is based on a discourse unit which we have termed a *basic block.* A basic block consists of two elements:

1. an organizational focus, such as a person or location, and

2. a set of concepts related to that focus.

The focus is what makes a cohesive unit of the material in the block; it is the thread common to all of this material, whether directly or indirectly.

A basic block will be realized as a primary segment of text which occupies the first level of the discourse hierarchy. In a coherent discourse, the foci on which the basic blocks are based are themselves related, each representing a different aspect of some unifying framework. These points are demonstrated by the testimony from which the basic block in Figure 1 was extracted[4].

This block references a particular time frame: zero to thirty seconds of the accident at Three Mile Island[5]. The remaining blocks of that testimony are similarly constructed around time frames, e.g., one to six minutes, six minutes to one hour, etc. Observed frameworks demonstrate a gamut of types: properties of the concepts (location, time), planning strategies in which events are involved (medical diagnosis), and characteristics that are not only inherent in the material but also due in part to the speaker's perception of them (significant factors). There appears to be no limit to what can constitute an acceptable framework, *only that it is derived from the material itself* and not from an independent device solely concerned with text structure. What may be a potential framework for one set of material may be totally inadequate for another. Note that these features are reflective of the guidelines suggested by analysis.

In addition to forming a cohesive unit, basic block structure is explicitly distinguished in the following two ways. First, it is often explicitly marked. In

[4] Space limitations prevent inclusion of the complete text.

[5] Three Mile Island is a nuclear power plant located in the state of Pennsylvania in the United States. It suffered a near-meltdown in 1979.

Now, what happened at Three Mile Island was that a feedwater transient was experienced, and by that I mean, simply, we lost feedwater to the plant momentarily.
Now, with loss of feedwater to the steam generator, the plant will experience a momentary pressurization above its normal pressure. This pressurization is released by a relief valve located at the top of the pressurizer. It simply opens and lets out a little bit of steam to take care of the excess pressure that is -interruption-
Then at 15 seconds into the event-keeping in mind that the valves opened maybe 5 seconds into the event-at 15 seconds the pressure started coming down because the valve had opened and cut off the pressure.
The valve should have reclosed when it got back down to about 2,250 pounds; it did not reclose. The pressure proceeded to come on down. At about 30 seconds into the event, this water here started disappearing, of course, because now you are continuing to remove a very large amount of heat here, which is then coming off as secondary side steam generation, and this water will proceed to disappear if you do not start replacing it.
And the auxiliary feedwater which normally comes on to make sure this does not go dry came on at about 30 seconds into the event. And at least the pumps were running. So this picture here is just the first 30 seconds and-or thereabouts.

And this figure here is indicative of the situation from 1 to 6 minutes into the event.

Figure 1: A Typical Primary Segment

Figure 1, the speaker closes the block with explicit reference to its focus: "So this picture here is just the first 30 seconds and-or thereabouts." The subsequent block in this testimony is also well marked, this time by its initial sentence (which is the last sentence of the figure): "And this figure here is indicative of the situation from 1 to 6 minutes." The basic block structure is further distinguished by certain forms of repetition, whose use is closely tied to the basic block structure. While a brief allusion to these forms is made below, a detailed discussion is contained in [MCM89].

2.3 CHARACTERISTICS OF COHERENT, HIGH-LEVEL STRUCTURE

Given the existence of a high-level structure as evidenced by explicit markings and repetition, we must consider how such a framework is chosen. If the only consideration were the identification of a set of related foci which can partition the material to be conveyed, then *any* set of related concepts about which the material can be cohesively organized would suffice. Consider the motivation behind the block of Figure 1. On the surface, the events in this block all occurred within the first thirty seconds of the accident. However, it is doubtful whether the driving force behind the construction of this block was to communicate the time frame in which these events occurred; rather, what is of importance is their relative sequence in the total series of events, their cause-effect relations, and their impact on the resulting accident. One could argue that the individual events in this block represent a cause-effect chain, and hence their mutual grouping; but this chain is continued in the next block of the testimony. Thus, this argument alone cannot account for this segmentation. Apparently, other factors beyond the ability to cohesively juxtapose clauses contribute to the segmentation and the high-level framework about which it is constructed.

In all of the dialogues examined, the blocks are of approximately the same size. Apparently, speakers choose organizations that tend to be well-balanced. However, balance does not seem to be the only criterion taken into consideration; if it were, we could expect to see perfectly balanced organizations in which each of the concepts to be conveyed is realized by its own basic block in addition to those well-balanced organizations found in our text analysis whose blocks are comprised of many concepts. The size of the component basic blocks seems to be a further consideration in determining what constitutes an acceptable framework. No lengthy discourse organized about a single primary segment has been observed, nor have any in which primary segments are composed of single concepts. The "ideal" size of a basic block seems to be three or four paragraphs in length for an explanation of three pages.

We contend that the high-level structure of extended explanations reflects the characteristics of an "ideal" framework in which:

- The basic block foci about which the material is organized reflect various aspects of the unifying framework.
- Basic blocks are the same size.
- The length of each basic block approximates the ideal size.

Our basic block model of discourse posits that the attainment of each of these features, termed *compositional goals*, is instrumental in the selection of a high-level framework.

However, not all observed explanations conform to the basic block model as presented thus far. For example, in explanations that are concluded by a final summary, one would expect the summary to emphasize the major points stressed by the speaker in the body of the text. In fact, there is a strong correlation between the segments comprising such a summary and the foci about which the basic blocks are constructed (see [MCM89]). However, some explanations have been examined in which material that appears in a summary has *not* been included in the text. Another apparent exception has been observed in explanations that exhibit a well-defined high-level structure that accounts for all of the basic blocks except for the final one, which is at best loosely related to the others.

We postulate that a speaker attempts to identify a framework that is capable of coherently expressing all of the material he wishes to convey while satisfying the compositional goals equally well. However, a speaker is rarely blessed with such an ideal situation. The problem appears to be one of finding a satisfactory, rather than an optimal, unifying framework. We hypothesize that the satisfaction of some goals will be sacrificed in favor of others so that a framework that provides the best *overall* solution can be achieved. This relaxation of goal constraints explains the above anomalies and the variability of size and balance observed in basic block structure.

3 A MODEL FOR GENERATION

We have developed a computational model of discourse generation that captures the structural characteristics observed in naturally-occurring explanations. Our model is based on the hypothesis that the high-level structure of a discourse can be determined by bottom-up processes that attempt to satisfy speaker, listener, and compositional goals. Once this organization has been established, top-down processes are used to organize the information into basic blocks and to supplement that information based on the choice of framework. Only after the basic block structure has been established will detailed organization within a block and realization into text proceed. The remainder of this section describes our strategy for identifying an organizational framework and the resulting basic blocks given an initial set of concepts to be conveyed. We conclude by establishing our model within the context of a complete text generating system.

3.1 IDENTIFICATION OF CANDIDATE FRAMEWORKS

We suggest that a speaker, when organizing an extended explanation, will be faced with one of the following general situations:

- He already has a well-defined organizational structure in which the material to be conveyed has already been organized.
- He has a set of goals about which he wants to structure the discourse. In this situation, the goals serve as basic block foci about which he must now attempt to partition his material.
- The speaker is starting "from scratch"; he needs to find a unifying framework about which he can structure what he wants to say. It is this situation which our strategy intends to capture.

Our strategy assumes that the generation process starts with some initial state of affairs and a communicative goal to be achieved. Based on this, a rough set of information to be conveyed will be identified. The selection of this material is not our concern; rather, we are concerned with determining a reasonable high-level structure for this information. We assume that the material to be conveyed has been categorized into at least two levels of importance: that which must be included at all costs (the *kernel*), and that which would be nice to include, but due to time, style, or coherence may be left out. The process of identifying candidate frameworks begins with the kernel.

The implication of the discussion in Section 2 is that a unifying framework will be some feature to which all of the kernel concepts can be related. In our model, we assume that the domain knowledge is maintained as a hierarchical network. Such a representation suggests that a node to which all of these concepts converge could serve as the unifying framework. The basic block foci, which represent various aspects of this framework, would then correspond to children of this node.

Given a set of propositions to be conveyed, one is randomly selected from the kernel and a trace is performed upwards from each of its arguments through the hierarchy. The traversal of the hierarchy is performed using generalization links, e.g., ISA, IS-PART, and SUBGOAL, incrementing counters associated with each node that is traversed. Additionally, CAUSE/EFFECT links can be followed when events are encountered. Property links, such as color and size, are used only if they were included as part of the kernel. Once the traversal has been completed for each element of the kernel, the nodes are ranked according to their counters; this will give some idea of the number of concepts that converge on each.

At this point, the general characteristics of basic blocks can be used to eliminate some obvious non-candidate frameworks. Since no extended explanations have been observed to be composed of a single, massive basic block, a node that has the same value as its parent can be eliminated from consideration as a candidate framework. By the same token, since no organization has been observed in which a basic block is constructed for each concept, any node that has a number of children on the same order as the number of concepts in the kernel can also be eliminated from consideration. Furthermore, nodes at which very few concepts converge can be disregarded since an acceptable framework must account for most (ideally all) of the kernel.

The remaining nodes represent the potential candidates for the unifying framework. In all likelihood, none of these candidates will account for all of the kernel concepts. However, it may be possible to connect such concepts to a candidate by finding a link between these concepts and ones which are already associated with a candidate, possibly via a property they have in common. Additionally, it may be possible to include such material in a final summary or in a "catchall" block. That speakers do introduce such material in this way is evidenced by our transcript analysis. We claim that the inclusion of such material in a final summary is warranted if the speaker initially intended to include a final summary and if the amount of material is of the same order as that comprising the individual segments of the summary; for the creation of a catchall block to be viable, there must be sufficient material to construct an adequate basic block and the material must form a cohesive unit.[6]

3.2 SELECTION OF THE UNIFYING FRAMEWORK

Once potential frameworks have been identified, each must be evaluated according to how well it meets the criteria described in Section 2.3:

- How thoroughly does a candidate account for the selected material?
- How uniformly does a candidate distribute the concepts among the resulting basic blocks?
- How closely do the generated blocks conform to the ideal size?

Additionally, a candidate may be evaluated as to how well it meets the needs of a user model.

The actual blocks are constructed around a candidate's immediate descendents, not the candidate itself. So, for example, if a candidate were the node *workstations*, its children - actual workstations at which things occurred - will become the basic block foci. The ideal case is one in which each child of the candidate accounts for approximately the same number of concepts. The balance each candidate achieves can be determined by comparing the counters of its children. At the same time, the candidate will be rated according to how closely its blocks conform to the ideal size. Lastly, we can rate the candidates on how well they meet the demands, if any, imposed by a user model. For example, if we know the user is familiar with the location in which the events occurred, basic blocks based on location should be given higher ratings; organizing an explanation around a framework with which the listener is familiar will facilitate his assimilation of the information.

[6] See [MCM89] for a more complete discussion.

3.3 GENERATION OF TEXT

At this point, the high-level structure of the explanation has been determined. In the remainder of this paper, we will attempt to place our strategy in the context of a complete system by giving a brief overview of the remaining phases of our model.

Once a framework has been selected, the concepts to be conveyed can be partitioned about each of the resulting foci. At this point, there may be some concepts for which the framework does not account. The system must determine whether to include them as a separate, trailing basic block, as part of a final summary, or to eliminate them altogether (refer to Section 3.1). In addition, depending on several factors (e.g., verbosity constraints, choice of framework, etc.), those concepts of moderate import may be integrated into the framework by performing a traversal of the networks or by finding links to concepts already incorporated by the framework (as described in Sections 3.1 and 3.2). These steps complete the process of partitioning the material about the unifying framework.

The next step is to determine the order of presentation for the blocks. The order in which the basic blocks are presented may be imposed by the material itself or by some external factor. If no overriding considerations exist, the material may present obvious "natural" choices, such as sequential ordering by time or by cause-effect. On the other hand, there may be a predetermined preference for a particular order, e.g., by their importance as determined by some external process.

Once the blocks have been ordered, text generation can proceed. We intend to use an established, low-level text structuring strategy (e.g., RST or schemas) for this phase. We anticipate that additional information may be included in order to satisfy low-level discourse goals (e.g., supplying information requisite for the user's comprehension of the material to be presented); thus, we do not claim that the entire contents of a text have been determined before structuring is performed. Rather than organizing and realizing text for the blocks *en mass*, we have adopted a strategy in which these steps are performed for each block individually. Text for one block is presented to the user before proceeding to the organization and generation of text for the next block. This is to preclude wasted effort which may result from reorganization necessitated by a user's questions.

4 CONCLUSIONS

This paper has presented a computational model for generating the high-level structure for extended explanations. It is being implemented as part of DIALS (Delaware Intelligent Advisory Language System), using the NIKL knowledge representation system. No claims are made that the methodology represents a psychological model of composition. However, it is maintained that the text produced by this model is typical of that encountered in many naturally-occurring extended explanations. The model does not presume to generate all acceptable organizations; indeed, there are many well-structured explanations for which it does not account. We have presented the simplest case in which a high-level organization is constructed from a parent and its children in a taxonomic hierarchy. However, many other possibilities exist, several of which are currently being examined within the semantic network paradigm. In addition, we are investigating the generation of appropriate recovery strategies in response to interruptions and their impact on the planning strategy. We are also incorporating several forms of repetition that are widely used in extended explanations.

The unit of discourse which is the major motivation for this model is the basic block, a primary segment of text which occupies the first level of the discourse hierarchy. It consists of an organizational focus and text constructed about that focus. While methods such as RST and schemas are adequate for local organization of text, they are insufficient for the determination of high-level structure, providing neither criteria which describe what constitutes an acceptable framework for a discourse, nor any clues as to how this framework should be constructed. This paper provides both. We believe that our model provides strategies essential for bridging the gap between the generation of short explanations and the successful generation of extended explanations.

References

[Are75] L. A. Arena. *Linguistics and Composition.* Georgetown University Press, 1975.

[Dan67] R. W. Daniel. *A Contemporary Rhetoric.* Little, Brown and Company, 1967.

[GS86] B. J. Grosz and C. L. Sidner. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12, 1986.

[HM89] E. Hovy and K. McCoy. Focusing your rst: a step toward generating coherent multisentential text. In *Proceedings of the Eleventh Annual Meeting of the Cognitive Science Society*, Cognitive Science Society, 1989.

[Hov88] E. Hovy. Approaches to the planning of coherent text. Fourth International Workshop on Natural Language Generation, 1988.

[KP66] T. S. Kane and L. J. Peters. *A Practical Rhetoric of Expository Prose.* Oxford University Press, 1966.

[McC85] K. F. McCoy. *Correcting Object-Related Misconceptions.* PhD thesis, University of Pennsylvania, 1985.

[McK85] K. McKeown. *Text Generation.* Cambridge University Press, 1985.

[MCM89] D. Mooney, M. S. Carberry, and K. McCoy. *The Identification of a Unifying Framework for the Organization of Extended, Interactive Explanations.* Technical Report 90-1, University of Delaware, 1989.

[MP88] J. Moore and C. Paris. Constructing coherent text using rhetorical relations. In *Proceedings of the Tenth Annual Meeting of the Cognitive Science Society*, Cognitive Science Society, 1988.

[MS88] J. D. Moore and W. R. Swartout. A reactive approach to explanation. Fourth International Workshop on Natural Language Generation, 1988.

[MT88] W. C. Mann and S. A. Thompson. Rhetorical structure theory: toward a functional theory of text organization. *Text*, 8, 1988.

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

[Pol86] L. Polanyi. *The Linguistic Discourse Model: Towards a Formal Theory of Discourse Structure.* Technical Report 6409, BBN, 1986.

[Rei78] R. Reichman. Conversational coherency. *Cognitive Science*, 2, 1978.

[Sac77] E. Sacerdoti. *A Structure for Plans and Behavior.* Elsevier North-Holland, 1977.

[Tho57] W. N. Thompson. *Fundamentals of Communication.* McGraw Hill, 1957.

[WA60] C. V. Wicker and W. P. Albrecht. *The American Technical Writer.* American Book Company, 1960.