Textual Constraints, Rhetorical Relations and Communicative Intentions in PLANDoc Narratives

Kathleen McKeown
Dept. of Computer Science
Columbia University
New York, NY 10027

Karen Kukich
Software Methods Research Group
Bellcore
Morristown, NJ 07960-6438

1. Introduction

While communicative intention certainly plays a role in language generation, we have found that there is a need for textual and rhetorical relations as well in our document generation system, PLANDoc. PLAN-Doc will produce multi-page narratives describing the activities of telephone network planning engineers. These narratives will be read by managers and auditors as well as other planning engineers.

Since the reports contain summaries, sometimes determining what information to include and how to organize it is based solely on **textual constraints.** Summaries must be concise and information can be packed in using constructions such as ellipsis, conjunction and modification. Thus, textual constraints that rely on syntactic structure, lexical information, and parallels in the semantic types of the underlying information must play a role in organizing a summary.

Since the intended audience of the narratives is broad, we often need to determine what information to include in the narrative based on general presentational factors that aim to improve the clarity of the text. **Rhetorical relations** are useful for modelling this sort of constraint.

There are also cases where it is necessary to communicate the planning engineer's **intent** in taking a specific action. However, we cannot model this computationally since we have no access to information about intention. While we cannot automatically generate information based on intent, we do allow the engineer to enter statements that describe her reasons for an action. These user-authored statements can be interspersed with machine-generated text.

In the following sections, we first give some background on the PLANDoc system and the narratives that we are aiming to generate. We follow this with examples that illustrate why different types of constraints are needed in determining the content and organization of a PLANDoc narrative. All examples are taken from a set of narratives written by an experienced planning engineer for actual runs of the underlying planning system that PLANDoc is based on.

2. PLANDoc Background

Currently, planning engineers use a powerful software tool, the PLAN system, that helps them derive 20-year relief plans optimizing the timing, placement and cost of new facilities in a route in the telephone network. Documentation of relief plans is needed to justify their implied expenditures to managers, auditors, and regulators as well as to provide a record of the engineer's reasoning for future network studies. PLANDoc will generate documentation using a trace of an engineer's interaction with PLAN.

Given a model of the current configuration of a route along with forecasts of future service demands, the planning engineer first calls upon PLAN to generate a Base relief plan for the route. Although the Base plan is the most economical long-term solution, it generally needs to be refined to account for practical and intangible factors known to the engineer but unknown to the computer. For example, the engineer may need to postpone the placement of equipment in a certain CSA (Carrier Serving Area) due to a fixed cap on near-term expenditures, or she may decide to activate DLC (Digital Loop Carrier) equipment in another CSA because small manholes in that area make the placement of additional copper cables impractical. For this reason PLAN includes a powerful Interactive Refinement Module that allows the engineer to do 'what-if' modeling to explore the effects of various refinements to the Base plan. The engineer draws upon her expertise to determine what refinements to try, often including some that she fully expects

to be sub-optimal, if only for the purpose of proving the point to managers, auditors and regulators. In the end, some subset of refinements is selected as the final Proposed Relief Plan.

3. PLANDoc Narratives

We employed an experienced planning engineer to write sample PLANDoc narratives based on actual PLAN runs. These manually-written narratives serve as models for our system development. The opening and closing paragraphs of PLANDoc narratives are summaries of the Base Plan and the final Proposed Relief Plan. The main body of PLANDoc narratives is a sequence of one-paragraph descriptions of each refinement scenario the engineer chooses to include. In addition to the machine-generated refinement paragraphs, the engineer is given the opportunity to enter a manually-written Engineer's Note for each refinement. The purpose the Engineer's Note is to provide motivation or to describe the reasoning behind the refinement. Thus, the main body of the narrative consists of alternating refinement paragraphs and engineering notes.

The Engineer's Notes are clearly meant to communicate intentions. In contrast, we have observed that some of our model refinement paragraphs (to be generated automatically): 1) appear to be organized by textual relations that govern their discourse structure; 2) include sentences that seem to be motivated by rhetorical relations; and 3) include sentences that require access to communicative goals in order to be generated. We provide examples of each.

3.1. Evidence of Textual Constraints in PLANDoc Paragraphs

In the following refinement paragraph, 8 messages have been grouped together based on textual constraints. Because of the similarity of each individual message and resulting sentence, the messages can be combined using conjunction to produce a very concise sentence.

1) This refinement demanded the following changes: denied DLC activations to CSA 2119, 2653, 2755, and 2757; used a cutover strategy of "ALL" for CSA 2605 and 2713; changed the "ALL" DLC system to "idlc96" for CSA 2605; and demanded DLC activation for CSA 2651 in 1994Q1.

In a second summary example shown below, the sentence in italics appears to have been placed in this position based on its relation to the previous sentences. In the body of the narrative, the date is usually included with each separate activation. To generate a summary like this, PLANDoc would have to note that it could group a series of activations initially and delay mentioning the date since it can refer back to them using a phrase such as "All these new activations." Such a decision can only be opportunistically based on similarities between data that can be grouped together and the available of an appropriate means for referring back to them.

2) The BASE plan for this route called for activating new DLC sites with fiber T-1 support in CSA 2117, CSA 2651, CSA 2105, CSA 2113, CSA 2115, CSA 2703, CSA 2755 and CSA 2757. DLC activations with copper conditioned pairs for T-1 support included CSA 2119, CSA 2653, CSA 2713, CSA 2523 and CSA 2451. Fiber was activated to CSA 2605. *All these new activations are in the first 10 years of the study.*

The use of constraints such as these for organizing content, which are opportunistically based on the surrounding text, seem only remotely related to rhetorical relations and communicative intentions, at least relative to the following examples. Such constraints are more textual in nature.

3.2. Evidence of Rhetorical Structure Relations in PLANDoc Paragraphs

Unlike the previous examples, some PLANDoc narrative paragraphs include sentences that appear to be motivated by rhetorical structure relations. Examples include the following:

3) This refinement demanded that fiber be activated to CSA 2907 in 1996. *This CSA was activated for DLC in 1994Q4 in RUNID 'dlc_2907'.*

4) This refinement denied activation of CSA 2119 and CSA 2120 *(the 2 DLC sites activated in the BASE plan).*

5) This refinement demands DLC activation of CSA 2551 in 1996 with 40 pairs cutover from copper to DLC at that time. *CSA 2551 was activated in 1997 in the BASE plan with no cutover.*

6) This refinement denied the *new* DLC activation *in the BASE plan* of CSA 4704.

These paragraphs require intelligent content planning because the italicized sentences refer not to actions in the current refinement scenario but to actions in some previous scenarios, in the Base plan or input data. Their underlying propositions are not among those passed to the text generator for the current refinement paragraph. Hence, a content planner capable of exploiting factors such as rhetorical relations is needed in order to seek out those propositions for inclusion. We take the rhetorical structure relation underlying all of the above examples to be that of CONTRAST since in each case the italicized sentence explicitly contrasts old information with new information.

One might surmise that the author had the communicative goal of calling the reader's attention to the contrast between the actions in the current refinement and previous information. However, for implementation purposes it seems sufficient for the content planning module to operate based on the inferred rhetorical relation of CONTRAST.

3.3. Evidence of Communicative Intentions in PLANDoc Paragraphs
Although intentional relations are evident in the previous examples, it can be characterized as a general intention that might be attributed to any author. There is nothing specific to the individual planning situation that causes the information to be added as in every report the difference between new refinements and the base plan must be clear. Thus, existing rhetorical relations are sufficient to make computational implementation possible. In contrast, the following examples embody some intentional relations that are beyond our capacity to implement.

7) This refinement activates CSA 4111 for DLC in 1994Q4. *CSA 4111 was activated in 1997 in the BASE plan* and **24 gauge cable was placed in section 4109 in 1994Q.**

ENGR NOTE: **This is an attempt to defer/eliminate the coarse gauge cable. It doesn't.**

8) This refinement demands the activation of fiber to CSA 2168b in 1996 and to CSA 2115 in 1994Q4. *These fiber activations eliminate manhole rebuilds in sections 2111, 2112 and 2115. Ducts and coarse gauge cable in sections 2165 and 2117 are also eliminated or deferred.* **Plus, they get fiber turned up in business areas at a lower cost than serving the DLC on copper.**

The first example includes both the refinement paragraph and the its accompanying engineering note. In both examples, neither the italicized nor the bold sentences refer to actions in the current refinement scenario, so all would require the effort of an intelligent content planner in order to be included.

Note that the italicized sentences in these examples are similar to the ones in the previous examples in that an intelligent content planner could exploit the rhetorical relation CONTRAST as a basis for their inclusion. The bold sentences are far more problematic in that their only basis for inclusion appears to be a communicative intention that can only be revealed in an engineering note. Note also that the communicative intention is related to a domain-specific telephone engineering principle. This principle is specific to the current planning problem and not to all planning problems. Unless we can discover and explicitly represent a finite set of such principles, it will not be possible to include such sentences in our machine-generated refinement paragraphs.

4. Conclusion
In written reports such as the ones PLANDoc will generate, communicative intent is not sufficient for determining the content and organization of the report. Since the audience is broad and planning scenarios are numerous, we cannot reason about the individual intentions and knowledge of writer and reader in determining what should go into the report. Often knowledge about the general kind of information that is needed is sufficient for determining what to include. This can be implemented using rhetorical relations. The opportunistic use of constraints based on the surrounding text, possible syntactic structures that can be

used, and available phrasings to determine content is distinct from current techniques for organizing content. However, this capability seems crucial when summaries are required.

References

Elhadad, M., FUF: The Universal Unifier - User Manual, Version 5.0, CUCS-038-91, Columbia University, 1991.

Kukich, K., K. McKeown, J. Lim, J. Phillips and N. Morgan, "User needs analysis and design methodology for an automated documentation generator", Bellcore Technical Memo submitted to the 2nd Annual Bellcore User-Centered Design Symposium, Bellcore, Piscataway NJ, 1993.

Mann, W., and S. Thompson, "Rhetorical Structure Theory: A theory of text organization", Technical Report Number ISI/RS-87-190, Information Sciences Institute, 1987.

McKeown, K. R., Text Generation: Using Discourse Strategies and Focus Constraints to Generate Natural Language Text, Cambridge University Press, Cambridge, England, 1985.