Observations and Directions in Text Structure

John A. Hughes and Kathleen F. McCoy,
University of Delaware
Newark, DE 19716
hughes@cis.udel.edu, mccoy@cis.udel.edu

1 Introduction

Early work in text structuring, such as [McK85] and [MT88] showed that texts of all types and genres seem to be composed of a small number of simple, intuitive units, variously referred to as rhetorical relations and rhetorical predicates. These two bodies of work differed in whether these units were best viewed as the bricks or the mortar of text structure, but in either case a small set of primitives seemed to suffice for all texts.

McKeown showed how a text generation system can make use of these sorts of primitives to produce coherent, informative texts. However, not long after that, it became obvious that McKeown's schematized block-stacking approach to generation compiled out too much information about a speaker's intentional goals in choosing the blocks s/he did, and that more of this information should be recorded in the process of text structuring to allow for such niceties as flexibility in answering follow up questions or requests for elaboration. Following this reasoning, [MS91] instead utilized the more "mortar-centered" approach of RST for text generation.

2 Text Structure Primitives: A Tool or Side-Effect?

However, there is even more to the information-hiding nature of McKeown's schemas than just intentional structure, which primitive-based text structuring approaches seem to suffer from in general. They all seem to hide certain "psychological" dynamics which undoubtedly underlie almost all human linguistic interaction. For example, it is possible that McKeown's schema for defining an object might be derivable from more general principles of information processing: Definitions tend to begin with an identification to give the hearer a default set of attributes in one fell swoop, then follow that with a shorter list of specific attributes to override those inherited. But even this approach might not be best schematized, as the following text seems to serve quite well as a definition:

Q: "What's a kumquat?"
A: "It's like a teeny little orange, but really sour."

This represents yet another technique, not relying upon inheritance but on the hearer's world knowledge, to convey a great amount of attributive information with very few resources. As McKeown's schemas were only arrived at through a process of text analysis, we know nothing about exactly what might make a particular schematized presentation of information effective at any given time, or how to choose between schematic texts if more than one is possible. No doubt certain hearers might prefer many examples, or perhaps a particular concept lends itself very well to analogy, but at present we have no deep understanding as to what would indicate that examples or analogies are useful.

Further, a reasonable analysis of an existing text can conflate, confuse or obscure the actual processes that went into composing it with other factors only secondary to the cohesiveness of the text. As Moore & Pollack point out in [MP92], there can be non-isomorphic relations between text spans on different dimensions; we add that it is therefore unclear which of these relations contribute most to the coherence of the text. It is possible that some relations are really side-effects of others. In Moore & Pollack's example,

"Come home by 5:00. Then we can go to the hardware store before it closes. That way, we can finish the bookshelves tonight."

there is a right-branching tree of Motivation relations, as well as a left- branching tree of Condition relations both structuring this text. However, the Motivation relations seem to be sufficient to structure this text, for probably the most likely way to motivate someone to do an act is to show how some desirable state of affairs is conditional on that act (and in fact, this is exactly one type of plan operator Moore & Swartout's generation system would use). In this case, the "informational" relations seem directly dependent on the intuitively more central "presentational" ones. This seems to suggest that the mere fact that relations appear in text after text in the analysis stage does not imply that those relations are the most useful tools to plan with.

In fact, this last point seems to be an essential one to explore. Both McKeown and Mann & Thompson found regularities in the texts they analyzed, and used these regularities to characterize coherent texts. But his characterization can only serve as a first approximation. Perhaps they've shown that coherent texts have RST trees or fit certain schemas, but it's clearly not the case that schematized texts or ones with RST trees are necessarily well-formed, even if they obey some additional constraints (e.g., focusing).

3 Choosing Appropriate Instantiations

In the same way that no theory we know of can guide a search among possible rhetorical strategies (Analogy, Identification, etc) as to their effects on the hearer and appropriateness, the content of individual relations and predicates are underspecified to such an extent that given a rich enough knowledge base, a single well-formed rhetorical skeleton could be incarnated with any amount of suboptimal or inappropriate information making for an ineffective or incoherent text.

To illustrate, imagine that one's task is to generate an explanation of the important social factors leading up to and resulting from the American Civil War. In describing the circumstances of the start of the war, there are many attitudes, events, and characteristics of the mid-1800's we could choose from to achieve setting up the circumstances. Among them, we know about the tensions between the industrialized North and the agricultural South, institutionalized racism, perhaps something about the attitudes of the people at the time towards their country and their lifestyles, and many others. But a rich knowledge base also contains a lot of information less useful given our task; the opening up of the Orient to trade with the West, institutionalized discrimination against women, and perhaps something about style of dress or what a typical day was like on a South Carolina plantation. All of these serve as providing circumstantial information, and given a perfect text from a history book which makes use of some circumstantial information along the lines of the first set above, it's quite likely one could find information equally circumstantial along the lines of the second set, focussing on the very same notions (economics, prejudicial attitudes, or some aspect of human interest), but whose use would have impaired the logical flow of the text. Similarly, in giving Background information about Abraham Lincoln, some information such as his attitudes towards state's rights and secession, and perhaps his legendary morality and honesty, must be sifted out of a rich knowledge base at the expense of other information about his mole, his beard, his childhood, or his clinical depressions, though all are construed as Attributive relations which focus on Lincoln and in some sense give the reader Background information about Lincoln.

Such a problem seems to involve calculating relevance, which has not received much attention as it relates to formal text structure devices like schemas or RST tree construction. An interesting question, however, is how clearly relevance selection can be divorced from formal structuring. There is no inherent irrelevance in any information contained in a knowledge base. Given different gross organizations and aims of the text, like a tale of the Civil War as gothic romance, hoop skirts and oppression of women might be mentionable within the same text as descriptions of Southern political

rage, and in writing a humanistic psychological treatise one can expound at length on Lincoln's melancholy and similar attributes where very different attributes were mentioned in the sociological text. It seems that all these different types of texts share the property that the skeleton of rhetorical relations are fleshed out with information which is in some way linked to a basic backbone or kernel of information which organizes the text – the "gist" of the text. If the gist of the text is to give a sequential historical account of a War, this backbone might consist of sequential Causes and Effects, with events and players introduced as they figure in events mentioned, and elaborated with information which is related not simply in one way to the immediate context, forming a tree, but perhaps to several parts of the backbone, forming a graph. Thus in a text which aims at a specifically sociological account of the war, hoop skirts and Lincoln's mole, though associated in the knowledge base to aspects of the Civil War, do not relate in any interesting way to other pieces of information in the gist of the tale and don't warrant mention, while his attitudes as a politician not only characterize him as personal attributes but also might serve as pointers to why he behaved as he did or why events unfolded as they did.

4 Where To Go From Here

The rhetorical relations and predicates proposed in the literature, then, are probably best viewed as approximations to be analyzed at a deeper level. It seems that the recurring patterns of relations and predicates that text analysis reveals are best viewed as common linguistic, goal-satisfaction techniques – techniques for enhancing understanding. Promising directions for future research in that field would then be along several lines: Discerning the underlying principles that make rhetorical techniques so useful, discerning the dynamic, situational phenomena which license the use of these techniques, and discerning what sorts of information best carries out the work of the techniques in any given text.

For the first tack, we would need to consider the effectiveness and efficiency of the various techniques. While the systems of Moore & Swartout [Moo89, MS91] and Maybury [May92] try to incorporate rhetorical techniques into an intentional, goal-oriented framework, they have not shown how one technique should be chosen above the other when both are possible. Maybury does not address the issue at all. Moore & Swartout list selectional heuristics which prefer operators that make the least demands on the hearer, but this very transparent use of the user model is not sufficient to make the most effective choices. For example, in describing a missing person to a police officer, the literature suggests such varied options as compare/contrast, attributive descriptions, analogical descriptions, description by parts and subparts, and so on. The most effective options in that situation are probably the attributive ("He's 5'11", 190lbs, brown hair, gray eyes, pale") and compare/contrast descriptions ("He's about your build but a little shorter"), but to decide this requires some understanding of the principles (perhaps psychological or cognitive) which underlie the identity of each of the techniques. Efficiency considerations also come into play. In the current scenario, given a choice between the two most effective techniques, we might prefer a compare/contrast description technique if a good candidate for comparison is available. The rationale might be that comparison to an entity transmits almost all of that entity's attributes with a single invocation, and the contrast allows quick overriding of a few salient attributes.

The second tack, discerning the dynamics of communication which license the use of these predicates and relations, is intimately related to the first, for the communication situation provides the impetus for searching for a technique in the first place. The "missing person" scenario above creates several interacting needs and desires on the parts of the participants. The relation of rhetorical techniques to these goals needs to be more fully focussed on. Similarly, the situation establishes a perspective on what is to be said; since the police will want to look for the missing person, his physical attributes are highlighted by the situation, giving the impulse and facility for

creating an attributive description or a comparative description highlighting physical features. In an another example, from Mann & Thompson,

> "What if you're having to clean floppy drive heads too often? Ask for Syncom diskettes, with burnished Ectype coating and dust-absorbing jacket liners..."

the overall relation seems to be one of Solutionhood. However, this relation in and of itself does not reflect the persuasive nature of the text. It does however reflect more general principles of what a reader desires from a product (a solution), how to attract the interest of a reader (pose a question s/he will want an answer to), and so on. A cursory glance through magazine advertisements shows an unusually large number of ads structured this way, and with very good reason.

And in following the third tack, determining appropriate context, it's useful to consider a wider range of texts than seems to have been considered before. We chose here to think about historical narratives to illustrate how a very rich knowledge base might confound selection of relevant information, but at the same time suggest a richer model of coherence based on more than just recursive, nested constituents. Typical domains in which text structure is studied are task-oriented in nature or are based on knowledge bases rich in isa-links but little else. To get at a more fundamental understanding of what factors influence text structure, one should look across many different text types and seek points of similarity. For example, is there a connection between the way a well-drawn comparison provides an efficient classification of an item in the hearer's complex knowledge base and the way a well-chosen set of background information provides support and explanation for the complex information to come? .

In the end, the inventory of techniques a generation system draws upon will need to be indexable by these interpersonal dynamics. They will also need to be characterizable in terms which facilitate constructing effective, efficient text. These considerations seem to demand an examination of rhetorical structures in terms of the mental states of the conversation partners, their perspectives and wants, and also perhaps in terms of the mental structures of the conversation partners, exploiting how they process analogies, similes, and classifications. In sum, we need to disassemble the information compiled into rhetorical relations and predicates.

References

[May92] Mark Maybury. Communicative acts for explanation generation. *International Journal of Man-Machine Studies*, 37, 1992.

[McK85] K.R. McKeown. Discourse strategies for generating natural-language text. *Artificial Intelligence*, 27(1):1–41, 1985.

[Moo89] Johanna D. Moore. *A Reactive Approach to Explanation.* PhD thesis, University of California, Los Angeles, 1989.

[MP92] Johanna Moore and Martha Pollack. A problem for rst: The need for multi-level discourse analysis. *Computational Linguistics – Squib*, 18(4), December 1992.

[MS91] Johanna D. Moore and William R. Swartout. A reactive approach to explanation: Taking the user's feedback into account. In Cecile Paris, William Swartout, and William Mann, editors, *Natural Language Generation in Artificial Intelligence and Computational Linguistics*. Kluwer Academic Publishers, Boston, 1991.

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