A Problem for RST: The Need for Multi-Level Discourse Analysis

Johanna D. Moore and Martha E. Pollack
Department of Computer Science
University of Pittsburgh
Pittsburgh, PA 15260
jmoore,pollack@cs.pitt.edu

Rhetorical Structure Theory (RST)[8], argues that in most coherent discourse, consecutive discourse elements are related by a small set of *rhetorical relations*. Moreover, RST suggests that the information conveyed in a discourse over and above what is conveyed in its component clauses can be derived from the rhetorical relation-based structure of the discourse. A large number of natural-language generation systems rely on the rhetorical relations defined in RST to impose structure on multi-sentential text [4, 5, 10, 12]. In addition, many descriptive studies of discourse have employed RST [2, 7, 9]. However, recent work by Moore and Paris noted that RST cannot be used as the sole means of controlling discourse structure in an interactive dialogue system [11], because RST representations provide insufficient information to support the generation of appropriate responses to "follow-up questions". The basic problem is that an RST representation of a discourse does not fully specify the intentional structure [3] of that discourse. Intentional structure is crucial for responding effectively to questions that address a previous utterance: without a record of what an utterance was intended to achieve, it is impossible to elaborate or clarify that utterance.[1]

Further consideration has led us to conclude that the difficulty observed by Moore and Paris stems from a more fundamental problem with RST analyses. RST presumes that, in general, there will be a single, preferred rhetorical relation holding between consecutive

[1]In addition, intentional structure is needed to make certain types of choices during the generation process, e.g., how to refer to an object [1].

discourse elements. In fact, as has been noted in other work on discourse structure [3], discourse elements are related *simultaneously* on multiple levels. In this paper, we focus on two levels of analysis. The first involves the relation between the information conveyed in consecutive elements of a coherent discourse. Thus, for example, one utterance may describe an event that can be presumed to be the cause of another event described in the subsequent utterance. This causal relation is at what we will call the *informational level.* The second level of relation results from the fact that discourses are produced to effect changes in the mental state of the discourse participants. In coherent discourse, a speaker is carrying out a consistent plan to achieve the intended changes, and consecutive discourse elements are related to one another by means of the ways in which they participate in that plan. Thus, one utterance may be intended to increase the likelihood that the hearer will come to believe the subsequent utterance: we might say that the first utterance is intended to provide evidence for the second. Such an evidence relation is at what we will call the *intentional level.*

RST acknowledges that there are two types of relations between discourse elements, distinguishing between *subject matter* and *presentational* relations. According to Mann and Thompson, "[s]ubject matter relations are those whose intended effect is that the [hearer] recognize the relation in question; presentational relations are those whose intended effect is to increase some inclination in the [hearer]" [8, p.18].[2] Thus, subject-matter relations are informational; presentational relations are intentional. However, RST analyses presume that, for any two consecutive elements of a coherent discourse, one rhetorical relation will be primary. This means that in an RST analysis of a discourse, consecutive elements will either be related by an informational or an intentional relation.

[2] Mann and Thompson analyzed primarily written texts, and so speak of the "writer" and "reader". For consistency with much of the rest of the literature on discourse structure, we use the terms "speaker" and "hearer" in this paper, but nothing in our argument depends on this fact.

In this paper, we argue that a complete computational model of discourse structure cannot depend upon analyses in which the informational and intentional levels of relation are in competition. Rather, it is essential that a discourse model include both levels of analysis. We show that the assumption of a single rhetorical relation between consecutive discourse elements is one of the reasons that RST analyses are inherently ambiguous.[3] We also show that this same assumption underlies the problem observed by Moore and Paris. Finally, we point out that a straightforward approach to revising RST by modifying the definitions of the subject matter relations to indicate associated presentational analyses (or vice versa) cannot succeed. Such an approach presumes a one-to-one mapping between the ways in which information can be related and the ways in which intentions combine into a coherent plan to affect a hearer's mental state—and no such mapping exists. We thus conclude that in RST, and, indeed, in any viable theory of discourse structure, analyses at the informational and the intentional levels must coexist.

An example

To illustrate the problem, consider the following example.

1. (a) George Bush supports big business.
   (b) He's sure to veto House bill 1711.

A plausible RST analysis of (1) is that there is an EVIDENCE relation between utterance (b), the nucleus of the relation, and utterance (a), the satellite. This analysis is licensed by the definition of this relation [8, p. 10]:

Relation name: EVIDENCE
Constraints on Nucleus: H might not believe Nucleus to a degree satisfactory to S

[3] It is not the only reason for ambiguity in RST analyses, but it is the only one we will comment on in this paper. Another well-known problem involves the underspecificity of the rhetorical relation definitions.

Constraints on Satellite: H believes Satellite or will find it credible
Constraints on Nucleus + Satellite combination:
H's comprehending Satellite increases H's belief of Nucleus.
Effect: H's belief of Nucleus increased.

However, an equally plausible analysis of this discourse is that utterance (b) is the nucleus of a VOLITIONAL CAUSE relation, as licensed by the definition [8, p. 58]:

Relation name: VOLITIONAL-CAUSE
Constraints on Nucleus: presents a volitional action or else a situation that could have arisen from a volitional action
Constraints on Satellite: none
Constraints on Nucleus + Satellite combination:
Satellite presents a situation that could have caused the agent of the volitional action in Nucleus to perform that action;
without the presentation of Satellite, H might not regard the action as motivated or know the particular motivation;
Nucleus is more central to S's purposes in putting forth the Nucleus–Satellite combination than Satellite is.
Effect: H recognizes the situation presented in Satellite as a cause for the volitional action presented in Nucleus.

It seems clear that example 1 satisfies both the definition of EVIDENCE, a presentational relation, and VOLITIONAL CAUSE, a subject matter relation. In their formulation of RST, Mann and Thompson note that potential ambiguities such as this can arise in RST, but they argue that one analysis will be preferred, depending on the intent that the analyst ascribes to the speaker:

> Imagine that a satellite provides evidence for a particular proposition expressed in its nucleus, and happens to do so by citing an attribute of some element expressed in the nucleus. Then ...the conditions for both EVIDENCE and ELABORATION are fulfilled. If the analyst sees the speaker's purpose as increasing the hearer's belief of the nuclear propositions, and not as getting the hearer to recognize the `object:attribute` relationship, then *the only analysis* is the one with the EVIDENCE relation ([8, p. 30], emphasis ours).

This argument is problematic. The purpose of all discourse is, ultimately, to affect a change in the mental state of the hearer. Even if a speaker aims to get a hearer to recognize some `object:attribute` relationship, she has some underlying intention for doing that: she wants to enable the hearer to perform some action, or to increase the hearer's belief in some proposition, etc. Taken seriously, Mann and Thompson's strategy for dealing with potential ambiguities between presentational (i.e., intentional) and subject matter (i.e., informational) relations would result in analyses that contain only presentational relations, since these are what most directly express the speaker's purpose. But, as we argue below, a complete model of discourse structure must maintain both levels of relation.

The argument from interpretation

We begin by showing that in discourse interpretation, recognition may flow from the informational level to the intentional level or vice versa. In other words, a hearer may be able to determine what the speaker is trying to do because of what the hearer knows about the world or what she knows about what the speaker believes about the world. Alternatively, the hearer may be able to figure out what the speaker believes about the world by recognizing what the speaker is trying to do in the discourse. This point has previously been made by Grosz and Sidner [3, pp. 188-190].[4]

Returning to our initial example

1. (a) George Bush supports big business.
   (b) He's sure to veto House Bill 1711.

suppose that the hearer knows that House Bill 1711 places stringent environmental controls

[4]In [3], *dominates* and *satisfaction-precedence* are the intentional relations, while *supports* and *generates* are the informational relations.

on manufacturing processes.[5] From this she can infer that supporting big business will cause one to oppose this bill. Then, because she knows that one way for the speaker to increase a hearer's belief in a proposition is to describe a plausible cause of that proposition, she can conclude that (a) is intended to increase her belief in (b), i.e., (a) is evidence for (b). The hearer reasons from informational coherence to intentional coherence.

Alternatively, suppose that the hearer has no idea what House Bill 1711 legislates. However, she is in a conversational situation in which she expects the speaker to support the claim that Bush will veto it. For instance, the speaker and hearer are arguing and the hearer has asserted that Bush will not veto any additional bills before the next election. Again using the knowledge that one way for the speaker to increase her belief in a proposition is to describe a plausible cause of that proposition, the hearer in this case can conclude that House Bill 1711 must be something that a big business supporter would oppose—in other words that (a) may be a cause of (b). Here the reasoning is from intentional coherence to informational coherence. Note that this situation illustrates how a discourse can convey more than the sum of its parts. The speaker not only conveys the propositional content of (a) and (b), but also the implication relation between (a) and (b): supporting big business entails opposition to House Bill 1711.[6]

It is clear from this example that any interpretation system must be capable of recognizing both intentional and informational relations between discourse elements, and must be able to use relations recognized at either level to facilitate recognition at the other level. We are not claiming that interpretation always depends on the recognition of relations at both levels, but rather that there are obvious cases where it does. An interpretation system therefore needs the capability of maintaining both levels of relation.

[5] The hearer also needs to believe that it is plausible the speaker holds the same belief; see [6].

[6] This is thus an example of what Sadock call *modus brevis* [13].

The argument from generation

It is also crucial that a generation system have access to both the intentional and informational relations underlying the discourses it produces. For example, consider the following discourse:

S: (a) Come home by 5:00. (b) Then we can go to the hardware store before it closes.

H: (c) We don't need to go to the hardware store. (d) I borrowed a saw from Jane.

At the informational level, (a) specifies a CONDITION for doing (b): getting to the hardware store before it closes depends on H's coming home by 5:00.[7] How should S respond when H indicates in (c) and (d) that it is not necessary to go to the hardware store? This depends on what S's intentions are in uttering (a) and (b). In uttering (a), S may be trying to increase H's ability to perform the act described in (b): S believes that H does not realize that the hardware store closes early tonight. In this case, S may respond to H by saying:

S: (e) OK, I'll see you at the usual time then.

On the other hand, in (a) and (b), S may be trying to motivate H to come home early, say because S is planning a surprise party for H. Then she may respond to H with something like the following:

S: (f) Come home by 5:00 anyway. (g) Or else you'll get caught in the storm that's moving in.

What this example illustrates is that a generation system cannot rely only on informational level analyses of the discourse it produces. This is precisely the point that Moore

[7] See [8] for definitions of the RST relations used throughout this example.

and Paris have noted [11]. If the generation system is playing the role of S, then it needs a record of the intentions underlying utterances (a) and (b) in order to determine how to respond to (c) and (d). Of course, if the system can recover the intentional relations from the informational ones, then it will suffice for the system to record only the latter. However, as Moore and Paris have argued, such recovery is not possible because there is not a one-to-one mapping between intentional and informational relations.

The current example illustrates this last point. At the informational level, utterance (a) is a CONDITION for (b), but on one reading of the discourse there is an ENABLEMENT relation at the intentional level between (a) and (b), while on another reading there is a MOTIVATION relation. Moreover, the nucleus/satellite structure of the informational level relation is maintained only on one of these readings. Utterance (b) is the nucleus of the CONDITION relation, and, similarly, it is the nucleus of the ENABLEMENT relation on the first reading. However, on the second reading, it is utterance (a) that is the nucleus of the MOTIVATION relation.

Just as one cannot always recover intentional relations from informational ones, neither can one always recover informational relations from intentional ones. In the second reading of the current example, the intentional level MOTIVATION relation is realized first with a CONDITION relation between (a) and (b), and, later, with an OTHERWISE relation in (f) and (g).

Discussion

We have illustrated that natural-language interpretation and natural-language generation require discourse models that include both the informational and the intentional relations between consecutive discourse elements. RST includes relations of both types, but commits to discourse analyses in which a single relation holds between each pair of elements.

One might imagine modifying RST to include multi-relation definitions, i.e., definitions that ascribe both an intentional and an informational relation to consecutive discourse elements. Such an approach was suggested by Hovy [4], who augmented rhetorical relation definitions to include a "results" field. Although Hovy did not cleanly separate intentional from informational level relations, a version of his approach might be developed in which definitions are given only for informational (or, alternatively, intentional) level relations, and the results field of each definition is used to specify an associated intentional (informational) relation. However, this approach cannot succeed, for several reasons.

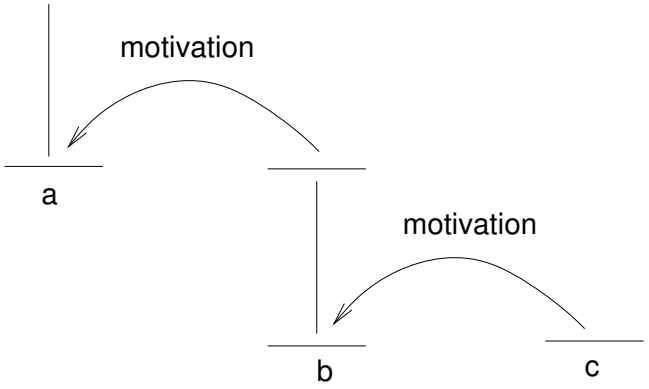
First, as we have argued, there is not a fixed, one-to-one mapping between intentional and informational level relations. We showed, for example, that a CONDITION relation may hold at the informational level between consecutive discourse elements at the same time as either an ENABLEMENT or a MOTIVATION relation holds at the intentional level. Similarly, we illustrated that either a CONDITION or an OTHERWISE relation may hold at the informational level at the same time as a MOTIVATIONAL relation holds at the intentional level.

Thus, an approach such as Hovy's that is based on multi-relation definitions will result in a proliferation of definitions. Indeed, there will be potentially $n \times m$ relations created from a theory that initially includes n informational relations and m intentional relations. Moreover, by combining informational and intentional relations into single definitions, one makes it difficult to perform the discourse analysis in a modular fashion. As we showed earlier, it is sometimes useful first to recognize a relation at one level, and to use this relation in recognizing the discourse relation at the other level.

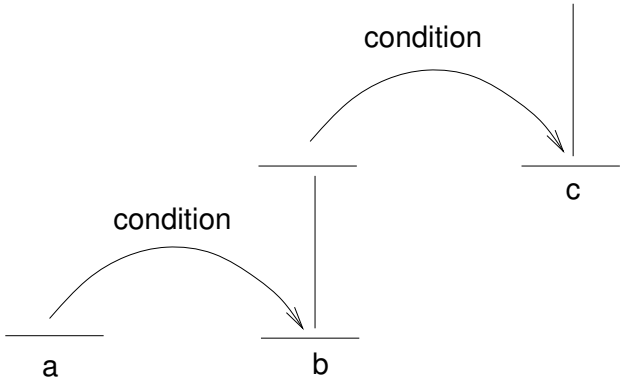
In addition, the multi-relation definition approach faces an even more severe challenge. In some discourses, the intentional structure is not merely a relabeling of the informational structure. A simple extension of our previous example illustrates the point:

S: (a) Come home by 5:00. (b) Then we can go to the hardware store before it closes. (c) That way we can finish the bookshelves tonight.

A plausible intentional level analysis of this discourse, which follows the second reading we gave earlier, is that finishing the bookshelves (c) motivates going to the hardware store (b), and that (c) and (b) together motivate coming home by 5:00 (a). Coming home by 5:00 is the nucleus of the entire discourse: it is the action that S wishes H to perform (recall that S is planning a surprise party for H). This structure is illustrated below:



At the informational level, this discourse has a different structure. Finishing the bookshelves is the nuclear proposition. Coming home by 5:00 (a) is a condition on going to the hardware store (b), and together these are a condition on finishing the bookshelves (c):



The intentional and informational structures for this discourse are not isomorphic. Thus, they cannot be produced simultaneously by the application of multiple-relation definitions that assign two labels to consecutive discourse elements. The most obvious "fix" to RST

will not work. RST's failure to adequately support multiple levels of analysis is a serious problem for the theory, both from a computational and a descriptive point of view.

Acknowledgements

We are grateful to Barbara Grosz, Kathy McCoy, Cécile Paris, Donia Scott, Karen Sparck Jones, and an anonymous reviewer for their comments on this research. Johanna Moore's work on this project is being supported by grants from the Office of Naval Research Cognitive and Neural Sciences Division and the National Science Foundation.

References

[1] Douglas E. Appelt. *Planning english sentences.* Cambridge University Press, Cambridge, England, 1985.

[2] Barbara Fox. *Discourse structure and anaphora: written and conversational English.* Cambridge University Press, Cambridge, England, 1987.

[3] Barbara J. Grosz and Candace L. Sidner. Attention, intention, and the structure of discourse. *Computational Linguistics*, 12(3):175–204, 1986.

[4] Eduard H. Hovy. Approaches to the planning of coherent text. In Cécile Paris, William Swartout, and William Mann, editors, *Natural Language Generation in Artificial Intelligence and Computational Linguistics*, pages 83–102. Kluwer Academic Publishers, Boston, 1991.

[5] Alistair Knott. New strategies and constraints in RST-based text planning. MsC Thesis, Department of Artificial Intelligence, University of Edinburgh, 1991.

[6] Kurt Konolige and Martha E. Pollack. Ascribing plans to agents: Preliminary report. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*, pages 924–930, Detroit, MI, 1989.

[7] Keith Vander Linden, Susanna Cumming, and James Martin. Using system networks to build rhetorical structures. In *Proceedings of the Sixth International Workshop on Natural Language Generation*, pages 183–198, Berlin Heidelberg, Germany, 1992. Springer-Verlag.

[8] William C. Mann and Sandra A. Thompson. Rhetorical structure theory: A theory of text organization. USC/Information Sciences Institute Technical Report Number RS-87-190, Marina del Rey, CA, 1987. Also published in Livia Polanyi, editor, *The Structure of Discourse*. Ablex Publishing Corporation, Norwood, N.J., 1987.

[9] Christian Matthiessen and Sandra Thompson. The structure of discourse and 'subordination'. In J. Benjamins, editor, *Clause Combining in Grammar and Discourse*. Amsterdam, 1988.

[10] Johanna D. Moore and Cécile L. Paris. Planning text for advisory dialogues. In *Proceedings of the Twenty-Seventh Annual Meeting of the Association for Computational Linguistics*, pages 203–211, Vancouver, B.C., 1989.

[11] Johanna D. Moore and Cécile L. Paris. Planning text for advisory dialogues: Capturing intentional, rhetorical and attentional information, University of Pittsburgh Computer Science Department Technical Report 92-22, Pittsburgh, PA, 1992.

[12] Dietmar Rosner and Manfred Stede. Customizing RST for the automatic production of technical manuals. In *Proceedings of the Sixth International Workshop on Natural Language Generation*, pages 199–215, Berlin Heidelberg, Germany, 1992. Springer-Verlag.

[13] Jerrold M. Sadock. Modus brevis: The truncated argument. In *Proceedings of the Chicago Linguistics Society*, pages 545–554, 1977.