

Two Views of Rhetorical Structure Theory

William C. Mann and Sandra A. Thompson

Abstract

Rhetorical Structure Theory (RST) is about the structure of language uses, specifically the discourse structure of written texts. In particular, it provides a way to talk about the connectedness and integrity of written monologues, and how texts serve purposes of their authors. RST analysis is illustrated below using a natural text.

Use of RST has uncovered various issues and possibilities for misunderstanding. Misunderstandings can arise from taking an inappropriate view of the scientific status of RST. This paper seeks to resolve some of these problems by identifying two different views that an RST user might take, and describing what can be expected from RST in each view. One view is more empirical and the other more causal.

1. An Introduction

Rhetorical Structure Theory (RST) was created as a way to attribute structure to text. It was developed in the 1980s, and the most widely known form is represented in a paper by Mann and Thompson (Mann and Thompson 1988). Since that publication appeared, RST has been used in both linguistics and computational linguistics. Although many sorts of changes in RST have been proposed, we will use the 1988 version for discussion because it is the most accessible and widely known.

RST has sometimes been misunderstood, in several different ways, probably in part because people have brought to it expectations about discourse structure, and about what a description of discourse structure must include. A major purpose of this paper is to correct or prevent inappropriate expectations, so that some avoidable difficulties will be avoided in fact. RST is a way to explore the structure of written monologues. An example of RST analysis is given below. Typically an RST analysis starts by dividing a text into some minimal units of interest, such as independent clauses. A role in the text is then assigned to each such unit, primarily by linking parts of the text together using *relations*, and by aggregating related parts into *spans*. Spans can be linked to other units or spans, so that the text is connected together into a hierarchic structure. Most of the relations are asymmetric, linking a *nucleus* to a *satellite*. The largest span created in this manner encompasses the whole text. In some cases (identified later below) other structural information is also added.

Many of the misunderstandings of RST focus on the relations mentioned above. RST deals with all of its relations in a uniform way, and so readers have expected that the relations themselves would be very similar. Some readers have also expected that the various relations would be documented in a uniformly complete way. Some have expected that the underlying principles that govern and differentiate the relations and other structural elements would be obvious or at least evident, and therefore that RST can be used in detail as a guide to programming natural language processes for computers. Unfortunately, these expectations are excessive. RST makes some contributions in these directions, but does not fulfill these expectations. Sometimes such expectations have been used as integrity tests, and when RST is found to not meet them, RST is seen as defective and the simplicity and arbitrariness of the tests were not blamed.

Possibly some of these problems originate in assumptions about what RST is. To remedy this, we can provide a more accurate view of the nature of RST, and thereby indicate a way that readers can think about RST and form more appropriate expectations. Doing this is the major purpose of this paper.

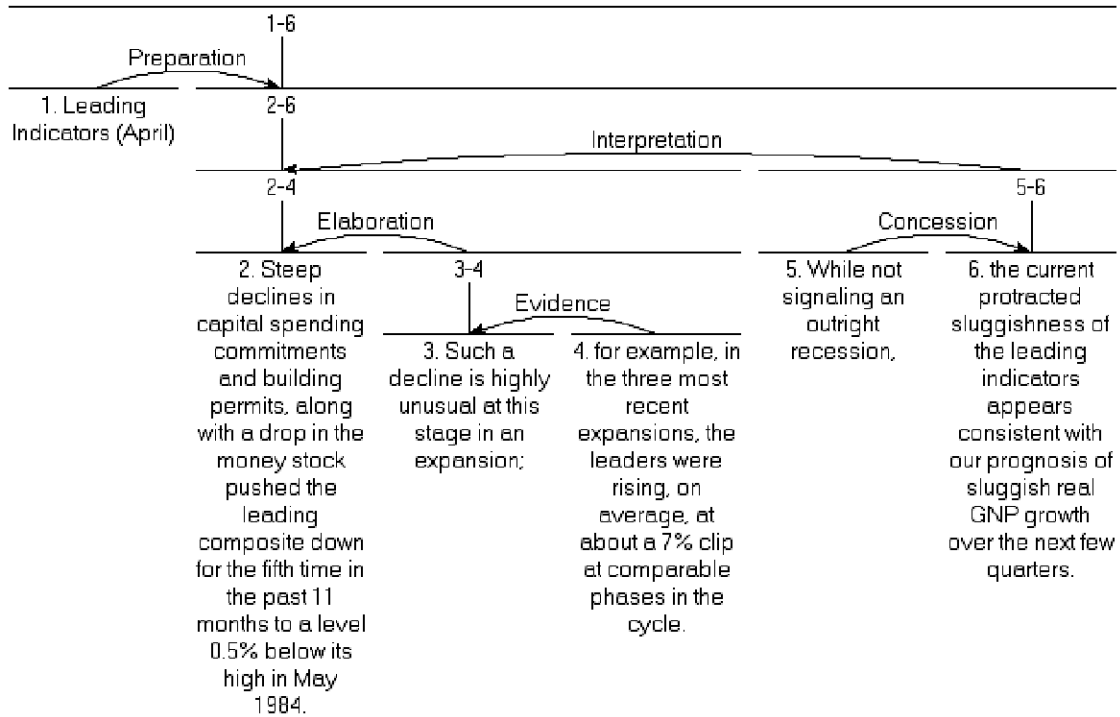
We explore the nature of RST by exhibiting two views of it that a user could take. These two are chosen because they seem to represent the two major kinds of views that have actually been taken. Also, they represent a contrast of positive and negative aspects of the subject.

2. RST Analysis

Below is an example of a text and the corresponding RST analysis diagram. The text is a commentary published in May 1985 by the U.S. Federal Reserve Board as part of a large report of economic statistics. It is divided into six units, beginning with the title, indicated by numbers which have been added.

A Commentary on Some Financial Statistics

1. Leading Indicators (April)
2. Steep declines in capital spending commitments and building permits, along with a drop in the money stock pushed the leading composite down for the fifth time in the past 11 months to a level 0.5% below its high in May 1984.
3. Such a decline is highly unusual at this stage in an expansion;
4. for example, in the three most recent expansions, the leaders were rising, on average, at about a 7% clip at comparable phases in the cycle.
5. While not signaling an outright recession,
6. the current protracted sluggishness of the leading indicators appears consistent with our prognosis of sluggish real GNP growth over the next few quarters.



Working bottom up, unit 4, about previous expansions, is taken as evidence that the present declines are unusual, as stated in unit 3. Units 3 and 4 together form a span which elaborates the “steep declines” mentioned in unit 2. The span of units 2-4 is interpreted by units 5-6, where unit 5 is conceding limits on the degree of the interpretation, and unit 6 states the interpretation. Unit 1 is a title that prepares for reading the entire remainder.

These perceptions are expressed by the Preparation, Interpretation, Elaboration, Evidence and Concession relations seen in the diagram. Each arrow points from a satellite to a nucleus, and the relation names appear as labels on the arrows.

RST analyses are created by people (called *observers*) who make a systematic set of judgments about the text, expressed in diagrams. An RST diagram is equivalent to affirmation of a particular set of judgments. Because they are based on definitions, the judgments can be identified without consulting the observer, so that each analysis has a definite interpretation.

The set of relations is open. A researcher can create new relations by changing the set of relation definitions used.¹

3. Views and Consequences

The development and use of RST has identified certain patterns of events that accompany RST analysis, involving analyzability of texts, ambiguity of discourse structure and agreement between observers. The scientific status of these patterns, and of the structures posited by RST, has been unclear.

This paper explores the consequences of taking various points of view about the scientific status of RST. In particular, two points of view are compared, one more empirical and descriptive, the other more abstract and theory oriented. To anticipate, we see RST as primarily of value as an extension of current descriptive knowledge of language use, and only of suggestive value for theory building or model building. We have constructed the two views discussed here in order to make it easier for researchers to examine their own views and their expectations of RST.

First we identify various phenomena that are somehow associated with RST. These are then selectively reexamined from the two viewpoints, first discussing aspects that are distinctive for one or the other viewpoint and then aspects that are involved in both viewpoints. For each of these examinations, we discuss the credibility and consequences of the viewpoint taken.

What phenomena associated with RST need to be examined? Its course of development indicates several that are discussed below, beginning with aspects of individual analyses and then moving to aspects of collections of analyses.

The development of RST was based on work with several hundred natural texts produced by educated adults. Such texts tend to be coherent and also to be understandable with low levels of apparent ambiguity.

There is nothing in RST which, given a text, forces an analysis to be found. Just as a word sequence can be found ungrammatical, a sequence of text units can be found unanalyzable in RST. However, the most frequent outcome of analysis of a text was that:

¹ More detail is presented in (Mann and Thompson 1988) and in (Mann and Thompson 1992). An introduction and other information, including further references, can be found at www.sil.org/linguistics/RST.

- an analysis was found spanning the whole text;
- there was little or no ambiguity in the analysis;
- there were high levels of agreement between observers working separately.

4. Scientific Status

RST has been used more or less directly in a wide range of work. In addition to studies involving coherence, it has been used to study interclausal relations, conjunctions, implicit communication, clause combining, writing style and genre. In computation it has been used as a design orientation in text generation programming, summarization, text indexing and modeling of comprehension.

Based on publications and comments about RST, it seems that the nature of RST, in scientific terms, has sometimes been misjudged, and misjudgment has led to inappropriate expectations. A better understanding of the nature of RST would clarify notions of success, progress, appropriate rigor, completeness of accounts, degree of demonstration, prediction, confirmation and significance. This paper is intended to contribute directly to the understanding of some, but not all, of these matters.

To develop the two views discussed below, we begin by a step of classification. Particular scientific ideas can be classified on the basis of whether they seek to identify phenomena or instead seek to explain previously identified phenomena. Depending on expository purpose, the first kind may be called questions, descriptions, phenomena or observations, and the second kind may be called answers, explanations, models, claims or theories. The first kind of ideas is often called empirical or descriptive. Ideas of the second kind are often called explanatory. Theories, (in the second group,) are ordinarily expected to be causal, systematic and abstract, and they may or may not include processes. Descriptions, (in the first group,) are generally not causal, are less systematic and abstract and seldom include processes. Our aim here is not to define with precision any one of these terms, but rather to identify and use the points of contrast between the two collections of terms.

We will call the first kind the *Data Gathering* view (DG), since it is focused on collecting and annotating texts taken as data. We will call the second kind the *Theory Component* view (TC), since it attempts to view RST as a component of a theory of language function.

Views of the scientific status of RST can be divided in terms of whether they are the first kind (questions) or the second (answers). Each kind has its own characteristic roles. In particular, identification with one group or the other leads to very different expectations about what will be provided. RST has been identified by various users with each of the two groups, which has sometimes led to confusion or misunderstanding.

To anticipate, we will show ways in which these two views of RST differ, discuss the sorts of significance and application that each one can have, and what expectations of RST are plausible for each. We conclude that each view has a legitimate role, but that RST is much more directly represented by the DG view. We also conclude that a view which does not differentiate between the two kinds of ideas is inappropriate.

5. Views and Elements

Below we discuss each view. We begin with an overview, then discuss the distinctives of each and then present what they share. In essence, the views are these:

The Data Gathering (DG) View says: RST is fundamentally a data collection method. It is intended to provide the basis for a clear statement of particular phenomena of language which theories might account for. It is not itself such a theory.

Employing this data collection method, some patterns arise. (Several were anticipated above and are identified in section 5.1 below.) The patterns can be significant in characterizing the data, and so they can also constitute data. Essentially, in the DG view, the possibility of RST analysis raises among others these two general questions which have to do with the significance of RST:

1. On single analyses: How can it happen that readers of text (both observers and general readers) can identify the sorts of discourse structures that observers regularly find?
2. On collections of analyses: What significant regularities are there in collections of RST analyses?

These are questions that the DG view of RST can raise, but answers are outside of its scope. (Some partial answers to the second question, for particular patterns, are discussed in section 5.1.)

In contrast:

The Theory Component (TC) View says: RST is a systematic causal account (i.e. a possibly testable theory or model) for a certain kind of text structure. It identifies text structures that may occur, and gives conditions under which they may occur. Notes on how this could be done, and the resulting challenges, are in section 5.2.

In a causal account of text function, (creation or reception) there would be many aspects that RST says nothing about. Lexical items, sentential semantics, authors' roles, genre, various processes and many other matters must be provided for in any such account. The account would also need to specify how various such elements interact. One role of RST in such an account could be to guide or constrain the satisfaction of some of the authors' goals by use of the discourse structures of texts.

To use RST in accounting for text creation, the account would include processes that, in effect, develop the relational structure of the text so that the end result conforms to some edition of RST. Similarly, a causal account of how text is received would include processes that, in effect, discover the relational structure of the text in a way that conforms to some edition of RST. In both cases the goals of the writer of the text will be part of the information. Either theory, when validated, would be in part an answer to questions raised by the DG view.

The DG view does not produce a causal account, whereas accounts built following the TC view generally will be causal.

5.1 Specifics of RST seen as Data Gathering (DG)

In developing RST, the DG view has been primary, partly because the team that created RST felt that it was important to be very explicit about what phenomena were being represented in analysis. So, the RST descriptions of the observer and the interpretation to be placed on an analysis are attempts to be specific about the textual phenomena being identified.

Another reason for this emphasis has been a desire to make the role of human judgment in analysis explicit.

Yet another reason was to “de-personalize” discourse analysis. Text analyses in some earlier frameworks cannot be interpreted with certainty without consulting the analysts. Key terms may be

undefined, key assumptions unstated or the significance of the analysis may be left to be identified by the reader. We wanted to avoid these conditions.

Thus the DG view is an attempt to benefit from a very explicit statement of how texts have been analyzed. The 1988 paper was not always fully explicit concerning the underlying nature of RST; this paper is an attempt to fill some of the gaps.

First below we discuss the nature of the process of observing, and then the nature of the outcomes of that process, which are individual RST analysis diagrams. Then we discuss eight particular aspects of collections of RST analyses: analyzability, text coherence, stability of the set of relations and schemas, agreement among observers, recognition of un signaled discourse structure, cross language transferability of RST, apparent ambiguity and “theory-neutral” data gathering.

The Process of Observation:

Observation produces diagrams representing the observer’s judgments. Observation is designed to work on natural texts, texts that were created without any influence from the analysis process. By minimizing use of the assumptions of any particular variety of linguistics, we hope that the results will be seen as having broad significance. RST observation makes explicit various aspects of reading text that might otherwise be missed or simply recognized only tacitly.

Observing is guided reading, seeking instances of a prescribed set of conditions. It is not a technical exercise under the regulation of any elaborate formal theory of linguistics or communication. For ordinary text, the interpretive capabilities of the observer are not distinguished from the capabilities of readers, so RST analyses should be seen as representing understandings that are tacitly available to readers as well.

The Nature of Analysis Results:

An analysis (normally, a diagram) is in effect an expression by the observer, saying “Here is what I found by following the diagram building method.” It does not record how the method was followed; the method does not tell the observer how to proceed but only demands obeying specified constraints on the result.

An analysis diagram is a shorthand notation for the affirmation by the observer, for every element of the diagram, that the observer agrees with the (applied form of the) definition of that element as applied to the particular parts of the text involved. In the usual case -- that the diagram forms a connected whole -- there is an affirmed role for every element of the text. The particular assertions represented by the diagram can be identified from the definitions, the diagram and the text, and so it is not necessary to consult the analyst to decide what the analysis says. Thus every analysis has a definite interpretation.

Some Questions and Answers about Collections of Analyses:

Collections of analyses can show things that single analyses cannot. A certain amount of significance (meaningfulness) of the method is achieved if one observer is consistently able to analyze texts drawn from a wide variety of text types. Another sort of significance is achieved if many observers (with suitable training and practice) can do so. Additional significance is achieved if these observers have a strong tendency toward agreement, either by part-by-part identity of their analyses or by easy acceptance of analyses of others as analyses that they themselves could affirm. Below we identify particular points of significance that arise from collections of analyses done by one or more analysts.

In general, these are informal rather than statistically based findings. Beyond informal sorts of significance, it would be possible to perform statistical tests of levels of agreement between observers. RST has achieved the informal levels of significance cited here, and a small amount of work has been done by others at a more formal level (den Ouden, van Wijk et al. 1998; Marcu, Amorrotu et al. 1999; Marcu, Romera et al. 1999; Marcu 2000).

There are patterns in collections of analyses, and by recognizing and interpreting these patterns we add to the detail and richness of the analyses as data.

A collection of analyses by many observers, of many texts of diverse kinds, with comparability of multiple observers on single texts, would permit the following questions:

- Is there a strong likelihood that an analysis will be found?
- Is there a strong tendency toward agreement between observers of the same text working independently?
- Is there a strong likelihood that any analysis found will be unambiguous or nearly so?

The experience of the developers of RST yields these answers: When a new text was selected for study, there was in fact a very strong tendency for an analysis to be found. However, for arbitrarily chosen spans of language use, e.g. an arbitrarily chosen half of a magazine article, or more generally spans that were not texts, no analyses were found.

There was also a very strong tendency for the same analysis to be found by all of us who attempted to analyze the text independently, and where there were differences, it was generally very easy to find a consensus -- an analysis affirmed by all who analyzed that text. The levels of ambiguity found were extremely low.

In addition to these answers, there are other aspects of collections of analyses that add significantly to our understanding of RST. We examine eight specific aspects below, describing them and assessing their significance.

Significance of aspects of collections of analyses:

- **Analyzability:** High likelihood of producing an analysis is significant only if, as in RST, the method does not force an analysis to be found. The significance of high likelihood of finding an analysis for texts, together with low likelihood of finding an analysis for non texts, is simply that there is some close connection between RST analysis and texthood. This pattern is itself a phenomenon of a different sort, one which we hope will someday be accounted for. By study of the analyses we expect to illuminate it.
- **Coherence:** RST analysis assigns a role to every part of a text. This is significant partly because the roles assigned are related to the coherence of the text read.

One definition of coherence of a text is this:

1. Every part of the text has an evident, discernible role in the text, and
2. Reading the text does not suggest that there are gaps or missing parts.

RST deals only with the first condition.

If coherence carries a requirement that every part of a text has an evident role, and RST analysis is a way of assigning a role to every part of a text, then it is possible (but not forced) that the roles assigned by RST help explain the coherence of the text.

So, the coherence of the financial text above is represented in the analysis by the fact that each of the six units is linked in a structured way to all of the others.

Such an account of text structure must be partial, of course, since RST is working only on certain scales, primarily ranging from the size of a single clause up to the full text. It is also partial in that it does not deal with many of the aspects of text understanding, such as cohesion and information flow, which sometimes operate at the same scale as RST.

The relationship between RST's discourse structures and coherence was recognized even in very early work (Mann and Thompson 1983; Mann and Thompson 1986b). We continue to see part of the significance of RST in this link between coherence and RST analysis, which has been found in large collections of analyses.

- **Stability of the Set of Relations and Schemas:** Extending the set of relations and schemas is seldom necessary. RST has been applied to a very diverse collection of written monologues. The present stability of the defined set of structures suggests that it is approaching a certain sort of complete coverage of communicative texts. How complete must the coverage be to be significant? Complete coverage in any exhaustive sense may not be reached, any more than an unabridged dictionary is a complete list. For linguistic purposes it is enough that the set be thoroughly representative.
- **Agreement among Observers:** Broad agreement among observers shows to a certain degree that RST analyses are based on widely shared knowledge of the conventions of the language, of the subculture of the speaker/writer, of the worlds that the analyzed texts engage and of how texts work. It also shows that the discriminations that RST requires are widely accessible culturally, and that the RST definitions are clear enough to be followed in a uniform way.²

However, the level of agreement should not be overestimated. Experience in teaching RST indicates that people's analyses do not show much uniformity until they have had significant amounts of training and practice at analysis, preferably with coaching. Trained, experienced observers are the only ones whose performance is relevant in judging RST. Yet training, experience and coaching all reduce the effective diversity of the group, and may suggest that agreement comes from the training as well as from the widely shared knowledge of the conventions of the language. This reduces the confidence that the high agreement of observers is coming from independent judgments of text. Happily, the reduction seems small.

- **Recognition of Discourse Structure in the Absence of Signals:** Conjunctions and other discourse markers often signal the sorts of structure that RST recognizes. However, analysts find far more structure than is signalled (Mann and Thompson 1985; Mann and Thompson 1986a). This suggests that readers can recognize and think about part-to-part relational aspects of text without reliance on such signals. Because some of the relations are social and involve intention, this suggests that text reception depends on knowledge about authors, about social facts, and about purposes of texts; this knowledge creates additional communicative effects, effects that

² As noted earlier, the development of RST was shaped in part by the desire to maintain high levels of individual satisfaction with choices and high levels of agreement between observers, in this case the developers. One consequence is that the observers' instructions were written so that particular choices that are characteristically difficult were avoided. (The multiple cases in the definition of Elaboration are an example of this.) This has caused the level of detail available in RST analyses to vary from one relation and group to another. This variability compounds with natural variability of the language.

do not arise from text form. Form-based or signal-based semantics of texts does not capture these communicative effects.

While the above aspects represent relatively firm knowledge, there are other aspects that we still know little about. Three of them are identified here.

- **Cross-language Transferability:** RST has been applied to texts in English, as well as in a wide variety of other languages, including Arabic, French, Hebrew, Kalasha, Kuna, Spanish and Quechua. There is a widely shared impression that direct use of existing definitions is effective. Beyond such an impression, nothing has been established. The potential for surprises is very large, and no firm claims of significant transferability can be made.
- **Apparent Ambiguity:** It is also widely reported that the RST analysis process finds very little ambiguity of discourse structure.³ But why would this ambiguity level be very low? No one is sure. Here are a few possibilities.

It is possible that the source of low ambiguity is the ability of readers to judge stories in terms of motivations and intentions. RST forces the analyst to think about the writer's intentions. The plausibility of particular intentions and of the reasons for including a particular portion of text are directly examined. When one account seems particularly plausible, it has the effect of reducing the plausibility of other accounts, and thus of reducing whatever analytic ambiguity is present. Such ambiguity is not resolved on a basis of form, but rather on the basis of culturally informed judgments about how people think and how they make choices while writing.⁴

So, just as for cross-language transferability, no firm claims of a significant finding concerning apparent discourse structure ambiguity can be made.

- **“Theory-neutral” Data Gathering:** It is generally desirable in scientific endeavors to avoid building into data gathering methods (question building) a bias toward particular theories (covert answer selection.) This has been a persistent problem in all science, and it is particularly acute in the social sciences. That seems inevitable.

In RST we would like the data gathered to be as unbiased, as “theory-neutral” as possible. Even with the best of intentions, there are limits on how well this can be achieved. RST was built with an awareness of the issue. Yet orientations and assumptions are necessary for getting anything in particular done.

RST surely has more built in assumptions than we can enumerate. Assumptions can sometimes turn out to be biases or potential needs for adjustment, (just as in the current discussions of the role of set theory.) Some of the more evident assumptions of RST are suggested below, divided into

³ There are also erroneous stronger statements in the literature that say that RST denies discourse structure ambiguity. Other statements in the literature say that RST does not allow forms of analysis other than single trees, forms such as occur with parallelism. Both ambiguity and non-tree structure are explicitly permitted, along with multiple relations between one pair of units or spans. All of these are expressed by the observer using more than one diagram for a single text.

⁴ This argument parallels an argument about why apparent ambiguity of ordinary sentences is so much lower than the ambiguity levels that a computer program can find.

assumptions about language, about observers, about authors and about interpretation of patterns in collections of analyses.

About language: RST assumes that edited monologue text is essentially linear, so that linear structures can faithfully represent this variety of discourse structure. This is a convenient assumption, but any richer account that also takes into account layout, formatting and other visual issues will have to give it up.

About observers: there are assumptions that certain elements of the written relation definitions can be applied without difficulty by observers. These aspects include degrees of positive regard, and also positive regard as a shorthand for (Choose one among: belief; approval; desire to act). They include the ideas of nuclearity, use of multiple diagrams and structural ambiguity. They also include details of particular definitions.

About authors: there are assumptions that authors have goals when they write, and that those goals are of a kind representable by desired outcomes, and that those outcomes can be seen as changes within individuals comprising their expected audience.⁵

About patterns in collections of analyses: the notions of a text and of coherence are treated as unproblematic.

Certainly some of these assumptions are already controversial, and others could be made so. This is not troublesome, since if RST is effective then it will be enriched and corrected by those researchers who wish to do so.

Summarizing this section, the DG view is an attempt to be specific about identifying particular discourse phenomena. Because RST is examinable, it becomes possible to ask questions about how or why the analysis process produces results. Analysts can very reliably produce analyses for entire texts in a wide diversity of text types, which is significant in accounting for the widespread applicability of the idea of text coherence. Informally, agreement between trained observers is surprisingly high, showing that analysis has its basis in widely shared perceptions. There is clearly some cross-language transferability of RST, but the degree and the basis of it are unknown. And despite an effort to make it “theory-neutral,” RST uses a substantial array of assumptions.

5.2 Specifics of RST seen as a Component of Theories (TC view)

In this view of RST, it is seen as specifying some portion of a model of language use. Such portions of models may be static, in the same way that some grammars are static, or it may include abstract processes as well. Such models are sometimes tested by representing them as computer programs, with computations showing the effects of the structure and details of the model. Others are tested primarily by abstract reasoning. With either sort of test, research is facilitated if the parts of the model are very similar to one another, if the same principles apply to all similar parts, and if similar parts are all known to the same degree of detail or precision. In some kinds of theory or model building, these uniformities are regularly achieved by a formal fiat, simply declaring that members of a collection of similar items are all assumed to fit a single mold.

It is thus easy to come to expect these sorts of uniformity in accounting for aspects of language use.

⁵ In this paper we talk about authors’ goals and authors’ intentions interchangeably. In the literature, both terms are used extensively representing the same concepts.

When such expectations are applied to RST they fit badly. The definitional framework for relations is fairly uniform, but the relations are not. The bad fit can be blamed on specific differences between the relations themselves, on arbitrary or necessary differences in how particular relations are defined, on a lack of certain conceptual distinctions or on the framework of assumptions and other constructs in which the relations are embedded. Different researchers would focus on different ones of these aspects.

Even without any further detail, it is clear that expectations of substantial uniformities among the RST relations are not warranted.

Below we explore how RST might be used in an explanatory or formal model of discourse structure in language use. Our purpose here is to show how it is that RST fits these uses badly. These ideas may suggest remedies, or they may simply discourage inappropriate expectations.

In linguistics and related fields, text structure has been seen in many ways, sometimes in isolation and sometimes as part of a larger linguistic scope such as semantic structure, pragmatic structure or communication. There have been many different suggestions of what the essential, underlying abstract nature of text structure -- its theory -- really is. It has been suggested by some to be a well-formedness notion, akin to a text grammar. Another suggestion is to see text structure as aggregations of cohesive links. Compositions of speech acts, expression of particular genre organizations, the semantics of discourse markers, or an outworking of necessary rhetorical features -- all of these have been suggested. RST does not follow any of these suggestions.

Another group of suggestions, which are more suggestive of RST's methods, says that readers or hearers seek to perceive the goals or intentions of the producers of language.

An abstract linguistic model is often easiest to interpret relative to a particular direction of operation. Generative grammars are defined in the text generation or creation direction. Most traditional grammars are defined in the text reception direction. These orientations are not absolutely necessary, but they are widely used.

Of these two directions, text creation and text reception, the TC view of RST is probably most easily conceived of as a small portion of a model of text creation.

RST cannot be a whole model of text creation, of course, since it does not deal with the roles of words, grammar or semantics, and it does not touch many parts of pragmatics. For these reasons, it is not meaningful to try to see RST in isolation as a theory of text creation. Instead, it is better seen as accompanied by other components, compatible fragments of an overall model.

Thinking about possible roles for RST in theory building, it is interesting to consider what would come of making a partial model (module) or theory of text creation that conforms as closely as possible to what RST posits. After sketching such a model below, we evaluate some of its aspects, finding reasons why it is insufficient and why it would need (if taken seriously) to be augmented and revised. We will call this model the Conforming Model (CM.)

The investigative strategy here involves showing that even within its scope of concerns, RST leaves an enormous number of major orientations and details unspecified, and that if those details are chosen in a simple way the result is not credible. The general conclusion from this is that as a component of theory, there is presently no way to incorporate or evaluate RST in any theory.

To support this conclusion, we briefly examine one approach to RST-as-theory below. The choice of this example is not because it is practical or attractive, but rather that it is instructive.

A Sketch of a Conforming Model of Text Structure (CM)

CM necessarily makes various assumptions about authors and what they do. These assumptions, which are added to those of RST, must for our purposes be sufficient to provide a definite, examinable role for RST in text creation. Below we add enough assumptions so that a text creation model can be designed, but avoid any further assumptions where RST does not say anything. In particular, CM assumes:

- When an author creates a text, it is always done in pursuit of a goal (or goals) to be satisfied by the text as a whole. (This assumption is inherent in RST and is included so that it can be used below.)
- People, specifically authors, have non-topical knowledge structures (memories or skills) that in effect correspond to the schemas of RST.
- Creating a text includes identifying a configuration of these knowledge structures that might collectively achieve the author's goal. (In computational linguistics this is sometimes called a text plan.) Building such a configuration is based in part on knowing which RST-like structures can satisfy which goals.
- Working from the whole text to the parts, when a plan element is too big to be expressed in a single sentence, that element is broken into parts. Then, just as for the text as a whole, structures are added for each part. Thus CM is top-down, hierarchic and recursive.
- Creating a text also includes mapping the non-topical structures (mentioned in the second assumption) into tokens that specify what is to be said in each part of the text. Necessarily this involves lexical, syntactic and semantic choices.
- When the whole configuration has grown large enough, some process outside of CM creates for each unit a sentence that express what CM has specified. (It must be outside of CM, since CM is to conform to the conceptual scope of RST, and RST does not specify how to create sentences.)

Much more could be said about how CM could conform to RST, but the above is enough to allow us to state some criticisms of the model.

There are a number of apparent flaws of this approach as a model or theory of text creation. Many of them arise from the fact that CM does not provide for what RST does not specify. The list below describes direct consequences of the assumptions above. In particular, CM entails or assumes that:

- All sentence specifications can be successfully expressed in sentences.
- No checking is necessary.
- No reference is made by the model to the needs of the reader or the changes accomplished in the reader through reading.
- No provision is needed for accomplishing multiple goals through single actions.
- All structures will always accomplish the goals associated with them.
- The non-topical structures together with particular goals are sufficient to choose particular structures and to express those choices in sentences.
- Top-down development is sufficient.

Rather than criticize these assumptions and their consequences individually, we can summarize their attributes. They are hopelessly optimistic and rigid, inadequately self aware, weak at representing the reader and extremely unresponsive to the realities of using any particular language of expression.

Furthermore, they presume an unrealistic amount of detail in RST-like knowledge structures. Correcting such flaws would require creativity that works far beyond the boundaries of RST. Thus CM has no credibility as an independent model of text creation. (Obviously, a parallel conclusion could be drawn for text reception.)

Expectations in Model Building

To judge by some publications that discuss RST, it has raised, and then failed to meet, a number of expectations. It has been criticized for not being organized enough, not being taxonomic enough, not being precise enough, ... (there is a longer list.) But “not enough” for what? Often the driving purpose has been a desire to use RST in designing a computer program that would do a text-related task such as text analysis or text generation. (Programming shares many of the attributes of theory creation.) For such purposes, there is disappointment that RST does not tell all that needs to be told, even on topics that it raises. A hoped-for level of detail and regularity is not found. Criticism that RST is not ready for such applications have often been sound.

The silences of RST give the TC model builder a great deal of flexibility and room for innovation. A text creation model (or a comparable text reception model) is in effect a large, complex hypothesis about text and human use of language. From this complexity abundant opportunities arise for creativity and insight, and many opportunities for disciplines to influence their neighbors. There is certainly the opportunity to reorganize, recharacterize or taxonomize the RST relations for various purposes.

Some of these reorganizations and augmentations have already been described in the literature. They include enrichments of the notion of author’s goals, new treatments of the “Subject-matter” and “Presentational” groups of relations defined in the 1988 paper, and new ways to identify, organize and apply relations (Moore and Pollack 1992). More will surely follow, even independent of any effort to improve RST in the DG view.

Summarizing this section, we see that for building models of linguistic processes and structures, RST gives some orientation but is quite underspecified and so is not a “theory of text structure” per se. Turning these gaps into an advantage, there is thus a large opportunity to create and test models or hypotheses about how text structure arises and is used. More directly, we see that CM, and thus the set of things that RST specifies, are quite inadequate as a specification of how text structure does or can work.

5.3 Specifics of RST Shared by the DG and TC views

For completeness, we consider here aspects that are not unique to one of the two views above, but rather are present in both. Each of these aspects presents some sort of contrast between the two views. The first of these is the concept of intention, the second is implicit communication effects, and the third is the elementary units of analysis.

Both the DG and the TC views involve intention. Intention (here meaning the desires of the writer to produce particular effects) is a controversial topic in academic circles. Here we simply assume that

intentions are involved in communication. It does not seem to be a risky assumption, but we cannot justify it here, since this paper is focused on other matters.⁶

The intentions represented by an RST analysis are a very incomplete set. In either view, many additional intentions would have to be identified as necessary parts of a consistent view of communication. The ones identified by RST are foundational, rudimentary constructs, enough to indicate the reasons for the presence of the parts of the text, enough to support an interesting account of text coherence, but not all that could be identified with the text using RST-observer-like methods. Beyond these, many more would be needed to construct a full theoretical (TC) account of intentions in text creation.

Another aspect of text that touches both of these views involves the implicit communication effects of discourse structure. Clauses communicate, and discourse structure communicates as well, providing part of the communicative effect of a text. The discourse structure that RST recognizes communicates; the term “Relational Propositions” has been applied to this. (For additional detail see (Mann and Thompson 1985; Mann and Thompson 1986a).)

In analysis, the observer is a reader, recognizing what is communicated. There is no special place for Relational Propositions in the DG view. Rather, they contribute to the observer’s judgments about what is plausible.

In contrast, in a TC model of text creation, it is necessary to identify the structure that a reader will impute to the text, so that the Relational Propositions of a text being created can be checked for appropriateness -- whether they express, agree with or contradict the writer’s knowledge and goals.

Another aspect for comparison involves RST’s “units.” In both the DG view and the TC view, there appears to be an issue of what will be units, i.e. of what is the smallest element to which RST will be applied or used as a guide. This appears to demand a technical specification, or perhaps two, one for each view. Actually, the identification of units does not really create technical issues for either view.

In DG, the units are chosen for convenience rather than theoretical significance. They are intended to be easily identified, uncontroversial, a starting point for analysis. They are defined in this way so that the danger of circularity is avoided -- of analysis depending on units, and units depending on analysis. Most often, independent clauses together with their dependent material have been chosen as units. (Exceptions have been made for certain short items such as titles and section heads or addresses and phone numbers in advertising.) For some purposes, orthographic paragraphs have been used. Conceivably there is some analyst’s purpose for which book chapters could be the units.

In the TC view, the issue of unit size is not particularly significant, since for modeling either text creation or text reception, aggregates both larger and smaller than traditional units must be used. The unit level is not particularly distinctive. Furthermore, the RST creators strongly suggested that coordination methods that apply at the multisentence levels also apply within the sentence, and that there is a conceptual continuity from RST to clause combining (Thompson and Mann 1986; Mann and Thompson 1987; Matthiessen and Thompson 1987).

Finally, there is the question of whether the two views might somehow be combined. There seems to be no consistency or potential purpose in taking the DG and TC views simultaneously, in seeing

⁶ A strong description of the controversy over intention, along with a strong defense for using it in cognitive linguistics, is found in (Gibbs 1999).

RST somehow as both question and answer. From either viewpoint, there is no sense to incorporating the other.

6. Conclusions

We have seen that RST can be seen in two complementary ways, and that those ways lead to different expectations and uses of RST. By far the most complete and well developed view is the Data Gathering view, in which RST is seen as posing questions, identifying phenomena, revealing patterns in texts and their reading, and raising issues about the nature and function of language.

The other view, of Theory Components, puts RST in a role of suggesting what some small portions of models of language use might look like. Even in that limited role, RST is extremely underspecified. It might be usefully suggestive, but not definitive.

The mixture of the two views seems untenable.

So, each view has its own kind of usefulness. The Data Gathering view is the one for which we can expect recognizable development in years to come.

References

den Ouden, J. N. , C. H. van Wijk, J. M. B. Terken and L. G. M. Noordman (1998). Reliability of Discourse Structure Annotation. *IPO Annual Progress Report 33*: 129-138

Gibbs, Raymond W. Jr. (1999). *Intentions in the Experience of Meaning*. Cambridge: Cambridge University Press.

Mann, William C. and Sandra A. Thompson, (1983). *Relational Propositions in Discourse*, ISI: Information Sciences Institute of University of Southern California, Los Angeles, ISI/RR-83-115, 1-28.

Mann, William C. and Sandra A. Thompson (1985). Assertions from Discourse Structure In (eds.), *Proceedings of the Eleventh Annual Meeting* Berkeley, CA: Berkeley Linguistics Society, .

Mann, William C. and Sandra A. Thompson (1986a). Relational Propositions in Discourse. *Discourse Processes* **9**(1): 57-90

Mann, William C. and Sandra A. Thompson, (1986b). *Rhetorical Structure Theory: description and construction of text structures*, ISI: Information Sciences Institute of University of Southern California, Los Angeles, ISI/RS-86-174, 1-15. Presented at Third International Workshop on Text Generation, Nijmegen, The Netherlands.

Mann, William C. and Sandra A. Thompson (1987). Antithesis: A Study in Clause Combining and Discourse Structure In R. Steele and T. Threadgold (eds.), *Language topics: Essays in Honour of Michael Halliday II*: Amsterdam: John Benjamins, 359-384.

Mann, William C. and Sandra A. Thompson (1988). Rhetorical Structure Theory: Toward a functional theory of text organization. *Text* **8**(3): 243-281

Mann, William C. and Sandra A. Thompson, eds. (1992). *Discourse Description: Diverse linguistic analyses of a fund-raising text*. Pragmatics & Beyond, New Series. Amsterdam: John Benjamins.

Marcu, Daniel (2000). The Rhetorical Parsing of Unrestricted Texts: a surface-based approach. *Computational Linguistics* **26**(3): 395-448

Marcu, Daniel, Estibaliz Amorrutu and Magdalena Romera (1999). Experiments in Constructing a Corpus of Discourse Trees. *ACL'99 Workshop on Standards and Tools for Discourse Tagging Maryland*: 48-57.<http://www.isi.edu/~marcu/papers.html>

Marcu, Daniel, Magdalena Romera and Estibaliz Amorrutu (1999). Experiments in Constructing a Corpus of Discourse Trees: Problems, Annotation Choices, Issues. *Workshop on Levels of Representation in Discourse, Edinburgh*: 71-78.<http://www.isi.edu/~marcu/papers.html>

Matthiessen, Christian M. I. M. and Sandra A. Thompson (1987). The Structure of Discourse and "Subordination" In J. Haiman and S. A. Thompson (eds.), *Clause Combining in Discourse and Grammar* Amsterdam: John Benjamins, .

Moore, Johanna D. and Martha E. Pollack (1992). A Problem for RST: The Need for Multi-Level Discourse Analysis. *Computational Linguistics* **18**(4)

Thompson, Sandra A. and William C. Mann (1986). A Discourse View of Concession in Written English In S. Delancy and R. Tomlin (eds.), *Proceedings of the Second Annual Meeting of the Pacific Linguistics Conference 2*: Pacific Linguistics Conference, .