Rhetorical structure in dialog*

Amanda Stent
Computer Science Department
University of Rochester
Rochester, NY 14627
stent@cs.rochester.edu

Abstract

In this paper we report on several issues arising out of a first attempt to annotate task-oriented spoken dialog for rhetorical structure using Rhetorical Structure Theory. We discuss an annotation scheme we are developing to resolve the difficulties we have encountered.

1 Introduction

In this paper we report on several issues arising out of a first attempt to annotate complex task-oriented spoken dialog for rhetorical structure using Rhetorical Structure Theory (RST):

- Relations needed (section 3.1)
- Identification of minimal units for annotation (section 3.2.2)
- Dialog coverage (section 3.2.3)
- Overlap due to the subject-matter/presentational relation distinction (section 3.3)

We discuss how we are dealing with these issues in an annotation scheme for argumentation acts in dialog that we are developing.

2 Previous work

We are engaged in the construction and implementation of a theory of content-planning for complex, mixed-initiative task-oriented dialogs based on corpus analysis, for use in dialog systems such as the TRIPS system (Allen et al., 2000)[1]. Our basic premise is that a conversational agent should be able to produce whatever a human can produce in similar discourse situations, and that if we can explain why a human produced a particular contribution, we can program a conversational agent to produce something similar. Therefore, in examining our dialogs the question we must answer is "Why did this speaker produce this?".

RST is a descriptive theory of hierarchical structure in discourse that identifies functional relationships between discourse parts based on the intentions behind their production (Mann and Thompson, 1987). It has been used in content planning systems for text (effectively text monolog) (e.g. (Cawsey, 1993), (Hovy, 1993), (Moore and Paris, 1993)). It has not yet been used much in content planning for spoken dialog.

Because the dialogs we are examining are task-oriented, they are hierarchically structured and so provide a natural place to use RST. In fact, in order to uncover the full structure behind discourse contributions, it is necessary for us to use a model of rhetorical structure. Certain dialog contributions are explained by the speaker's rhetorical goals, rather than by task goals. In example 1, utterance 3 is *justification* for utterance 1 but does not directly contribute to completing the task.

Example 1

A	1	They can't fix that power line at five ninety and East
B	2	Well it
A	3	Because you got to fix the tree first

The details of how to apply RST to spoken dialog are unclear. If we mark rhetorical structure only within individual turns (as has generally been the case in annotations of text dialog, e.g. (Moser et al., 1996),(Cawsey, 1993)), we miss the structure in contributions like example 1 or example 2. There is also the question of how to handle dialog-specific behaviors: grounding utterances and back-channels (utterances that maintain the communication), and abandoned or interrupted utterances.

Example 2 (simplified)

A	1	Bus C at Irondequoit broke down.
B	2	Before it even got started?
A	3	Yeah, but we convinced some people to loan us some vans.

* This work was supported by ONR research grant N00014-95-1-1088, U.S. Air Force/Rome Labs research contract no. F30602-95-1-0025, NSF research grant no. IRI-9623665 and Columbia University/NSF research grant no. OPG: 1307. We would like to thank the anonymous reviewers and Dr. Jason Eisner for their helpful comments on earlier drafts of this paper.

[1]We are using the Monroe corpus (Stent, 2000), with reference to the TRAINS corpus (Heeman and Allen, 1995) and the HCRC Maptask corpus (Anderson et al., 1991).

Initial annotation

Dialog-specific	Subtypes of *Elaboration*	Other
Comment	Particularize, Generalize	Comparison
Correction	Instantiate	Counter-expectation
Cue	Exemplify	Agent, Role

New manual

Argumentation acts	Subtypes of *Elaboration*	Schemas
Question-response	Set-member	Joke, List
Proposal-accept	Process-step	Make-plan
Greeting-ack.	Object-attribute	Describe-situation

Figure 1: Examples of other relations

In our first attempt to annotate, we removed abandoned utterances, back-channels, and simple acknowledgments such as "Okay". We used utterances as minimal units; utterances were segmented using prosodic and syntactic cues and speaker changes (see 3.2.2). We did occasionally split an utterance into two units if it consisted of two phrases or clauses separated by a cue word such as "because".

Two annotators, working separately, marked one complete dialog using Michael O'Donnell's RST annotation tool (1997). They used the set of relations in (Mann and Thompson, 1987), and some additional relations specific to dialog or to our domain. Examples of the additional relations are given in figure 1. When we compared the results, the tree structures obtained were similar, but the relation labels were very different, and in neither case was the entire dialog covered. Also, the annotators found structure not covered by the relations given. As a result, we stopped the annotation project and started developing an annotation scheme that would retain rhetorical relations while dealing with the difficulties we had encountered. The rest of this paper describes this new annotation scheme. An example of the type of analysis we are looking for appears in figure 3.

3 Issues and proposals

The issues we encountered fall into three areas, which we will examine in turn: issues related to individual relations, dialog-specific issues, and issues related to the well-known presentational/subject-matter distinction in RST.

3.1 Relations

The key in any annotation project is to have a set of tags that are mutually exclusive, descriptive, and give a useful distinction between different behaviors. The set of relations we used failed this test with respect to our corpus.

As in earlier work (Moore and Paris, 1992), our annotators found some of the relations ambiguous. In particular, the differences between the *motivate* and *justify* relations and between the *elaboration* and *motivation* relations were unclear (partly because we did not distinguish between presentational and subject-matter relations).

Some of the relations we used overlapped. The *elaboration* relation is too broad; in some sections of our dialogs almost every utterance is an elaboration of the first one, but the utterances cover a wide variety of different types of elaborations. Anticipating this, we had given the annotators several more specific relations (see figure 1), but we also allowed them to use the *elaboration* tag in case a type of elaboration arose for which there was no subtype. As a result of the overlap, use of the *elaboration* tag was inconsistent. The *joint* relation is also too broad.

Other relations were never used, although one annotator went on to look at several more dialogs. In short, the set of relation-tags we used did not effectively partition the set of relations we saw.

In our annotation scheme, we are taking several steps to define relations more clearly, reduce overlap, and eliminate too-broad relations. Instead of giving annotators an semi-ordered set of relations with their definitions, we are giving them decision trees, with questions they can use to clarify the distinctions between relations at each point (figure 2). The annotators did not find the relation definitions in (Mann and Thompson, 1987) particularly helpful, but we are including simplified definitions, and annotators are instructed to test against the definitions before labeling any relation. We are including several examples with each definition, so that annotators can obtain an intuitive understanding of how the relations appear. Finally, we are providing any useful discourse cues that signal the existence of a relation.

We are eliminating relations that overlap with others. Where a relation appears to cover a variety of different phenomena, as in the case of *elaboration*, we are using more specific relations instead. We are eliminating the *joint* relation, as it gives no helpful information from a content-planning perspective and annotators are tempted to over-use it.

One of the criticisms of RST is that there is an infinite set of relations (Grosz and Sidner, 1986). The goal is to arrive at a mutually-exclusive, clearlydefined set of relations with discriminatory power in each domain, so we expect that for each new domain, it may be necessary to start with an initial set of high-level relations selected from different categories, examine a small set of texts or dialogs in that domain, and then revise the set of relations by making relevant high-level relations more specific. We used this process to develop our annotation scheme. In the manual we include instructions for moving to new domains. Our examples come from a variety of domains and types of discourse, to add generality.

3.2 Dialog-specific issues

3.2.1 Dialog-specific relations, schemas and conversational games

Task-oriented dialog is a complex behavior, involving two participants, each with their own beliefs and intentions, in a collaborative effort to interact to solve some problem. There is a whole set of behaviors related to maintaining the collaboration and synchronizing beliefs that does not arise in monolog [(Clark, 1996), (Traum and Hinkelman, 1992)]. These include answering questions, agreeing to proposals, and simply acknowledging that the other participant has spoken.

In example 3, utterance 3 provides *motivation* for utterance 1. However, A would not have produced utterance 3 without B's question. If we simply mark a *motivation* relation between utterances 1 and 3 we will be losing dialog coverage, the spans involved in the relation will not be adjacent, and we will be ignoring the important relationship between utterances 2 and 3. A better analysis would be to mark a *question-answer* relation between utterances 2 and 3, and a *motivation* relation between utterance 1 and the unit consisting of utterances 2 and 3.

Example 3

A	1	Then they're going to have to basically wait
B	2	Why?
A	3	Because the roads have to be fixed before electrical lines can be fixed

The *question-answer* relation is not in Mann and Thompson's original list of relations[2]. It is an "adjacency pair"[3], and is a type of conversational game (Clark, 1996). Adjacency pairs, like other relations, are functional relationships between parts of discourse, but they are specific to multi-party discourse.

In our annotation scheme, we include relations for different kinds of adjacency pairs (figure 1). We have tentatively categorized adjacency pairs with subject-matter relations, although they may eventually become a third category of relation.

Some of these relations are bi-nuclear. For instance, although usually the answer is the only part required for discourse coherence, at times both question and answer may be needed, as in example 4.

Example 4

A	1	And the last one was at the where on the loop?
B	2	Four ninety.

It would seem that these relations can only apply at the lowest levels of an RST analysis, with a different speaker for each span. However, example 5, in which turns 2–7 are the answer to the question in utterance 1, shows that this is not the case.

Example 5 (slightly simplified)

A	1	What's "close"?
B	2	"Close". Um I don't know. I I'm pretty sure that
A	3	So Mount Hope and Highland would be.
B	4	Yeah.
A	5	Well what about like 252 and 383?
B	6	It says "next".
A	7	Okay. So I guess it has to be adjacent.

It might seem that the simplest approach would be to annotate adjacency pairs between turns, and mark other rhetorical relations only within turns. However, we have found many instances of rhetorical relations, or even units (section 3.2.2), spanning turns. The two examples below illustrate a cross-speaker *elaboration* and a cross-speaker *sequence* relation.

Example 6

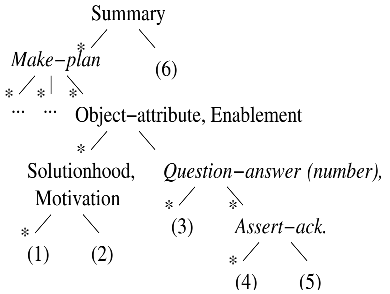
A	1	So that takes care of the ill guy and the handicapped guy.
B	2	Okay
B	3	And that takes two hours.

1. In this set of spans, is the speaker attempting to affect the hearer's:
   - **belief** – go to question 2
   - **attitude** – go to question 3
   - **ability** to perform an action – *enablement*
2. Is the speaker trying to increase the hearer's **belief** in some fact, or enable the hearer to better **understand** some fact?
   - Belief – *evidence*
   - Understanding – *background*
3. ...

Figure 2: Partial decision tree for presentational relations, expressed as a list of questions

[2] They do, however, include requests for information in the *solutionhood* relation

[3] An adjacency pair is a pair of utterances, the first of which imposes a cognitive preference for the second, e.g. question-answer, proposal-accept.



A	1	We have to send buses to the Lake.
A	2	There are people there to evacuate.
B	3	How many are we sending?
A	4	Two.
B	5	Okay.
B	6	So 1 ambulance to Pittsford and 2 buses to the Lake.

Figure 3: Sample analysis of part of a constructed dialog. Nuclei are marked with *; non-RST relations are in italics.

Example 7

A	1	So they can ta- ta- take out the power.
B	2	And then we have to wait ...

With a model of adjacency pairs, we can now handle grounding acts such as acknowledgments. If an utterance is clearly a back-channel or abandoned, annotators are instructed to so mark it and leave it out of further annotation.

RST in its original formulation does not cover enveloping or parallel structures or conventional forms. However, even in task-oriented dialogs speakers occasionally tell jokes. Furthermore, there are fixed, structural patterns in dialog, such as form-filling behaviors. These are frequently domain-specific, and resemble schemas [(McKeown, 1985), (Cawsey, 1993)]. While it may be possible to give an RST analysis for some of these, it is more accurate to identify what is actually going on. Our annotation scheme includes four of these, *make-plan*, *describe-situation*, *list* and *joke*. It also includes an adjacency pair for greetings, a conventional form.

An annotated dialog extract illustrating most of these issues is shown in figure 3.

3.2.2 Identifying and ordering units

In spoken dialog, both participants often speak at once, or one speaker may complete what another speaker says, as in examples 8 and 9.

Example 8 (+'s mark overlapping speech)

A	1	And + he's done + with that at one thirty
B	2	+ Okay +

Example 9

A	1	So it'll take them
B	2	Two more hours

Our original use of utterances as minimal units splits a cross-turn completion from the utterance it completes (example 9) , and says nothing about how to order units when one overlaps with another. We have altered our segmentation rules to take care of these difficulties. Our definition is that a minimal unit must be one of the following, with earlier possibilities taking precedence over later ones:

1. A syntactic phrase separated from the immediately prior phrase by a cue word such as "because" or "since"
2. A syntactically complete clause
3. A stretch of continuous speech ended by a pause, a prosodic boundary or a change of speaker

One unit will be considered to succeed another if it starts after the other.

This means that the standard segmentation of a dialog into utterances may have to be modified for the purposes of an RST analysis, although a segmentation into utterances and one into minimal units will be very similar. Annotators will start with a dialog segmented into turns and utterances, and are encouraged to re-segment as needed.

3.2.3 Dialog coverage

When one gets higher in the tree resulting from an RST annotation, the spans typically begin to follow the task structure or the experimental structure. In the Monroe corpus, usually one partner tells the other about the task, then the two collaborate to solve it, and finally one partner summarizes the solution (following the experimental structure). In the TRAINS corpus usually one subtask in the plan is discussed at a time (following the task structure).

Given the length and complexity of a typical dialog, it may not be possible to achieve complete coverage, even with our expanded relation set and the use of schemas. If we can identify useful sub-dialogs or can associate parts of a dialog with parts of the task, finding annotations for each part may suffice. For our domain, we have established heuristics about when an annotator can stop trying to achieve coverage. An annotator can stop when:

- The top level of the annotation tree has one relation label covering the whole dialog.
- The structure between the spans at the top level is identical to the task structure.
- The structure between the spans at the top level is identical to a domain-dependent or experiment-dependent schema.
- There is consensus between annotators that no more relations can be marked.

3.3 The subject-matter/presentational relation distinction

The relations in RST fall into two classes. Subject-matter relations such as *summary* are intended to be recognized by the hearer. Presentational relations such as *motivation* are supposed to "increase some inclination" in the hearer, such as the inclination to act (Mann and Thompson, 1987). As Moore and associates have explained in (1992) and (1993), while the intentions of the speaker are adequately represented in the case of presentational relations by the relations themselves, in the case of subject-matter relations the intentions of the speaker may vary. Furthermore, these two types of relations actually come from different levels of relationship between discourse elements: the informational level (subject-matter relations), and the intentional level (presentational relations). RST conflates these two levels.

Mann and Thompson said that, in the case where a presentational relation and a subject-matter relation were both applicable, the subject-matter relation should take precedence. However, we would like to have information about both levels when possible. In our annotation scheme the presentational relations are split from the subject-matter relations and annotators are instructed to consider for each set of spans whether there is a subject-matter relation, and also whether there is a presentational relation. If there are two relations, both are marked. If one covers a slightly different span than the other, at the next level of annotation the span that seems more appropriate is used.

In the following example, utterance 3 is *justification* (presentational) for utterance 1, but it is also in a *non-volitional cause* (subject-matter) relationship with utterance 1. The annotator would be instructed to label both relations.

Example 10 (slightly simplified)

A	1	I can't find the Rochester airport
B	2	+ I- it's +
A	3	+ I think I have + a disability with maps

We would also like more information, at times, about the subject matter in the spans of a relation. The relation between a "When" question and answer is *question-answer*, as is that between a "Why" question and answer; but the first question-answer forms part of an *elaboration* and the second forms part of a *justification* or *motivation*. In our annotation scheme, we supply a list of content types, such as time, location and number. The annotator adds the content type in parentheses after the relation tag when required. This means that the annotator may have to mark three items for a given set of spans: the presentational relation (if any), the subject-matter relation, and the content type (if required). We find this approach preferable to expanding the set of relations to include, for instance, *temporal-question-answer* and *spatial-question-answer*. Cawsey used a similar method in (1993).

4 Current and future work

We have an annotation manual that we are refining using TRAINS-93 dialogs[4]. Shortly, we will begin annotating the Monroe corpus with the new manual and different annotators. We will also annotate a few dialogs from a different corpus (e.g. Maptask) to ensure generality. We plan to use the results of our annotation in the construction (ongoing) of new generation components for the TRIPS system at the University of Rochester (Allen et al., 2000).

5 Related Work

In recent years there has been much research on annotation schemes for dialog. Traum and Hinkelman outline four levels of "conversational acts" in (1992). "Argumentation acts", including rhetorical relations, form the top level, but this level is not described in detail. DAMSL (Core and Allen, 1997) includes speech acts and some grounding acts, but not rhetorical relations. The HCRC Maptask project annotation scheme includes adjacency pairs, but not rhetorical relations (Carletta et al., 1996).

The COCONUT project annotation manual allows the annotator to mark individual utterances as *elaboration*, and segments as *summary*, *act:condition*, *act:consequence* or *otherinfo* (DiEugenio et al., 1998). This annotation scheme does not treat rhetorical structure separately from other types of dialog behavior. We have observed enough structure in the corpora we have looked at to justify treating rhetorical structure as a separate, important phenomenon. For instance, in a DAMSL-tagged set of 8 dialogs in our corpus, 40% of the utterances were statements, and many of these appeared in sequences of statements. The relationships between many of these statements are unclear without a model of rhetorical structure.

In (1999), Nakatani and Traum describe a hierarchical annotation of dialog for *I-units*, based on the *domination* and *satisfaction-precedence* relations of (Grosz and Sidner, 1986). Other researchers have shown that Grosz and Sidner's model of discourse structure (GST) and RST are similar in many respects [(Moser and Moore, 1996), (Marcu, 1999)]. However, RST provides more specific relations than GST, and this is useful for content planning. As well as helping to specify generation goals, content and ordering constraints, the rhetorical information is needed in case the system has to explain what it has said.

[4]A rough draft is available from the author.

RDA is an annotation scheme for identifying rhetorical structure in explanatory texts in the SHERLOCK domain (Moser et al., 1996). We follow RDA in requiring annotators to consider both intentional and informational relations. However, because of the dialog issues previously described, RDA is not sufficient for dialog.

Marcu uses discourse cues to automatically uncover rhetorical relations in text (1997). Much of this work is applicable to the problem of uncovering rhetorical relations in dialog; however, many cues in dialog are prosodic and it is not yet possible to obtain accurate information about prosodic cues automatically.

6 Conclusions

We have examined several issues arising from a first attempt to annotate spoken dialog for rhetorical structure. We have proposed ways of dealing with each of these issues in an annotation scheme we are developing. Much future work is certainly needed in this area; we hope that the results of our annotation may form a quantitative baseline for comparison with future work.

References

J. Allen, D. Byron, M. Dzikovska, G. Ferguson, L. Galescu, and A. Stent. 2000. An architecture for a generic dialogue shell. *upcoming in the Natural Language Engineering Journal special issue on Best Practices in Spoken Language Dialogue Systems Engineering*.

A. Anderson, M. Bader, E. Bard, E. Boyle, G. Doherty, S. Garrod, S. Isard, J. Kowtko, J. McAllister, J. Miller, C. Sotillo, H. Thompson, and R. Weinert. 1991. The HCRC Maptask corpus. *Language and Speech*, 34:351–366.

J. Carletta, A. Isard, S. Isard, J. Kowtko, and G. Doherty-Sneddon. 1996. HCRC dialog structure coding manual. Technical Report HCRC/TR–82, HCRC, Edinburgh University.

A. Cawsey. 1993. Planning interactive explanations. *International Journal of Man-Machine Studies*, 38:169–199.

H. Clark. 1996. *Using Language*. Cambridge University Press.

M. Core and J. Allen. 1997. Coding dialogs with the DAMSL annotation scheme. In *AAAI Fall Symposium on Communicative Action in Humans and Machines*, pages 28–35, November.

B. DiEugenio, P. Jordan, and L. Pylkkänen. 1998. The COCONUT project: Dialogue annotation manual. Technical Report 98–1, ISP, University of Pittsburgh.

B. Grosz and C. Sidner. 1986. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3).

P. Heeman and J. Allen. 1995. The TRAINS–93 dialogs. Technical Report Trains TN 94-2, Computer Science Dept., U. Rochester, March.

E. Hovy. 1993. Automated discourse generation using discourse structure relations. *Artificial Intelligence*, 63(1-2):341–385.

W. Mann and S. Thompson. 1987. Rhetorical structure theory: a theory of text organisation. In L. Polanyi, editor, *The Structure of Discourse*. Ablex, Norwood, NJ.

D. Marcu. 1997. The rhetorical parsing, summarization, and generation of natural language texts. Technical Report CSRG–371, Department of Computer Science, University of Toronto.

D. Marcu. 1999. A formal and computational synthesis of Grosz and Sidner's and Mann and Thompson's theories. In *The Workshop on Levels of Representation in Discourse*, Edinburgh, Scotland.

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

J. Moore and C. Paris. 1992. Exploiting user feedback to compensate for the unreliability of user models. *UMUAI*, 2(4):331–365.

J. D. Moore and C. L. Paris. 1993. Planning text for advisory dialogues: Capturing intentional and rhetorical information. *Computational Linguistics*, 19(4):651–695.

J. Moore and M. Pollack. 1992. A problem for RST: The need for multi-level discourse analysis. *Computational Linguistics*, 18(4):537–544.

M. G. Moser and J. D. Moore. 1996. Toward a synthesis of two accounts of discourse structure. *Computational Linguistics*, 22(3):409–420.

M. Moser, J. Moore, and E. Glendening. 1996. Instructions for coding explanations: Identifying segments, relations and minimal units. Technical Report 96–17, University of Pittsburgh, Department of Computer Science.

C. Nakatani and D. Traum. 1999. Coding discourse structure in dialogue (version 1.0). Technical Report UMIACS-TR-99-03, University of Maryland.

Michael O'Donnell. 1997. RST–Tool: An RST analysis tool. In *Proceedings of the 6th European Workshop on Natural Language Generation*, Gerhard-Mercator University, Duisburg, Germany.

A. Stent. 2000. The Monroe corpus. Technical Report TR728/TN99-2, University of Rochester.

D. Traum and E. Hinkelman. 1992. Conversation acts in task-oriented spoken dialogue. *Computational Intelligence*, 8(3):575–599.