

Investigating Cue Selection and Placement in Tutorial Discourse

Megan Moser

Learning Research & Dev. Center,
and Department of Linguistics
University of Pittsburgh
Pittsburgh, PA 15260
moser@isp.pitt.edu

Johanna D. Moore

Department of Computer Science, and
Learning Research & Dev. Center
University of Pittsburgh
Pittsburgh, PA 15260
jmoore@cs.pitt.edu

Abstract

Our goal is to identify the features that predict cue selection and placement in order to devise strategies for automatic text generation. Much previous work in this area has relied on ad hoc methods. Our coding scheme for the exhaustive analysis of discourse allows a systematic evaluation and refinement of hypotheses concerning cues. We report two results based on this analysis: a comparison of the distribution of *SINCE* and *BECAUSE* in our corpus, and the impact of embeddedness on cue selection.

Discourse cues play a crucial role in many discourse processing tasks, including plan recognition (Litman and Allen, 1987), anaphora resolution (Grosz and Sidner, 1986), and generation of coherent multisentential texts (Elhadad and McKewen, 1990; Roesner and Stede, 1992; Scott and de Souza, 1990; Zukerman, 1990). Cues are words or phrases such as *BECAUSE*, *FIRST*, *ALTHOUGH* and *ALSO* that mark structural and semantic relationships between discourse entities. While some specific issues concerning cue usage have been resolved (e.g., the disambiguation of discourse and sentential cues (Hirschberg and Litman, 1993)), our concern is to identify general strategies of cue selection and placement that can be implemented for automatic text generation. Relevant research in reading comprehension presents a mixed picture (Goldman and Murray, 1992; Lorch, 1989), suggesting that felicitous use of cues improves comprehension and recall, but that indiscriminate use of cues may have detrimental effects on recall (Millis et al., 1993) and that the benefit of cues may depend on the subjects' reading skill and level of domain knowledge (McNamara et al., In press). However, interpreting the research is problematic because the manipulation of cues both within and across studies has been very unsystematic (Lorch, 1989). While Knott and Dale (1994) use systematic manipulation to identify functional categories of cues, their method does not provide the description of those functions needed for text

generation.

For the study described here, we developed a coding scheme that supports an exhaustive analysis of a discourse. Our coding scheme, which we call Relational Discourse Analysis (RDA), synthesizes two accounts of discourse structure (Grosz and Sidner, 1986; Mann and Thompson, 1988) that have often been viewed as incompatible. We have applied RDA to our corpus of tutorial explanations, producing an exhaustive analysis of each explanation. By doing such an extensive analysis and representing the results in a database, we are able to identify patterns of cue selection and placement in terms of multiple factors including segment structure and semantic relations. For each cue, we determine the best description of its distribution in the corpus. Further, we are able to formulate and verify more general patterns about the distribution of types of cues in the corpus.

The corpus study is part of a methodology for identifying the factors that influence effective cue selection and placement. Our analysis scheme is coordinated with a system for automatic generation of texts. Due to this coordination, the results of our analyses of "good texts" can be used as rules that are implemented in the generation system. In turn, texts produced by the generation system provide a means for evaluation and further refinement of our rules for cue selection and placement. Our ultimate goal is to provide a text generation component that can be used in a variety of application systems. In addition, the text generator will provide a tool for the systematic construction of materials for reading comprehension experiments.

The study is part of a project to improve the explanation component of a computer system that trains avionics technicians to troubleshoot complex electronic circuitry. The tutoring system gives the student a troubleshooting problem to solve, allows the student to solve the problem with minimal tutor interaction, and then engages the student in a post-problem critiquing session. During this session, the system replays the student's solution step by step, pointing out good aspects of the solution as well as ways in which the solution could be improved.

To determine how to build an automated explanation component, we collected protocols of 3 human expert tutors providing explanations during the critiquing session. Because the explanation component we are building interacts with users via text and menus, the student and human tutor were required to communicate in written form. In addition, in order to study effective explanation, we chose experts who were rated as excellent tutors by their peers, students, and superiors.

1 Relational Discourse Analysis

Because the recognition of discourse coherence and structure is complex and dependent on many types of non-linguistic knowledge, determining the way in which cues and other linguistic markers aid that recognition is a difficult problem. The study of cues must begin with descriptive work using intuition and observation to identify the factors affecting cue usage. Previous research (Hobbs, 1985; Grosz and Sidner, 1986; Schiffrin, 1987; Mann and Thompson, 1988; Elhadad and McKeown, 1990) suggests that these factors include structural features of the discourse, intentional and informational relations in that structure, givenness of information in the discourse, and syntactic form of discourse constituents. In order to devise an algorithm for cue selection and placement, we must determine how cue usage is affected by combinations of these factors. The corpus study is intended to enable us to gather this information, and is therefore conducted directly in terms of the factors thought responsible for cue selection and placement. Because it is important to detect the contrast between occurrence and nonoccurrence of cues, the corpus study must be exhaustive, i.e., it must include all of the factors thought to contribute to cue usage and all of the text must be analyzed. From this study, we are deriving a system of hypotheses about cues.

In this section we describe our approach to the analysis of a single speaker's discourse, which we call Relational Discourse Analysis (RDA). Applying RDA to a tutor's explanation is exhaustive, i.e., every word in the explanation belongs to exactly one element in the analysis. All elements of the analysis, from the largest constituents of an explanation to the minimal units, are determined by their function in the discourse. A tutor may offer an explanation in multiple segments, the topmost constituents of the explanation. Multiple segments arise when a tutor's explanation has several steps, e.g., he may enumerate several reasons why the student's action was inefficient, or he may point out the flaws in the student's step and then describe a better alternative. Each segment originates with an intention of the speaker; segments are identified by looking for sets of clauses that taken together serve a purpose. Segments are internally structured and consist of a

core, i.e., that element that most directly expresses the segment purpose, and any number of contributors, the remaining constituents in the segment each of which plays a role in serving the purpose expressed by the core. For each contributor in a segment, we analyze its relation to the core from an intentional perspective, i.e., how it is intended to support the core, and from an informational perspective, i.e., how its content relates to that of the core. Each segment constituent, both core and contributors, may itself be a segment with a *core:contributor* structure, or may be a simpler functional element. There are three types of simpler functional elements: (1) units, which are descriptions of domain states and actions, (2) matrix elements, which express a mental attitude, a prescription or an evaluation by embedding another element, and (3) relation clusters, which are otherwise like segments except that they have no *core:contributor* structure.

This approach synthesizes ideas which were previously thought incompatible from two theories of discourse structure, the theory proposed by Grosz and Sidner (1986) and Rhetorical Structure Theory (RST) proposed by Mann and Thompson (1988). The idea that the hierarchical segment structure of discourse originates with intentions of the speaker, and thus the defining feature of a segment is that there be a recognizable segment purpose, is due to Grosz and Sidner. The idea that discourse is hierarchically structured by pairwise relations in which one relatum (the *nucleus*) is more central to the speaker's purpose is due to Mann and Thompson. Work by Moore and Pollack (1992) modified the RST assumption that these pairwise relations are unique, demonstrating that intentional and informational relations occur simultaneously. Moser and Moore (1993) point out the correspondence between the relation of dominance among intentions in Grosz and Sidner and the nucleus-satellite distinction in RST. Because our analysis realizes this relation/distinction in a form different from both intention dominance and nuclearity, we have chosen the new terms core and contributor.

To illustrate the application of RDA, consider the partial tutor explanation in Figure 1¹. The purpose of this segment is to inform the student that she made the strategy error of testing inside part3 too soon. The constituent that expresses the purpose, in this case (B), is the core of the segment. The other constituents help to achieve the segment purpose. We analyze the way in which each contributor relates to the core from two perspectives, intentional and informational, as illustrated below. Each constituent may itself be a segment with its own *core:contributor* structure. For example, (C) is a subsegment whose

¹In order to make the example more intelligible to the reader, we replaced references to parts of the circuit with the simple labels *part1*, *part2* and *part3*.

purpose is to give a reason for testing part2 first, namely that part2 is more susceptible to damage and therefore a more likely source of the circuit fault. The core of this subsegment is (C.2) because it most directly expresses this purpose. The contributor in (C.1) provides a reason for this susceptibility, i.e., that part2 is moved frequently.

- ALTHO
- A. you know that part1 is good,
 - B. you should eliminate part2 before troubleshooting in part3.
- THIS IS BECAUSE
- C. 1. part2 is moved frequently
 - AND THUS
 - 2. is more susceptible to damage.

Figure 1: An example tutor explanation

Due to space limitations, we can provide only a brief description of *core:contributor* relations, and omit altogether the analysis of the example into the minimal RDA units of state and action units, matrix expressions and clusters. A contributor is analyzed for both its intentional and informational relations to its core. Intentional relations describe how a contributor may affect the hearer's adoption of the core. For example, (A) in Figure 1 acknowledges a fact that might have led the student to make the mistake. Such a concession contributes to the hearer's adoption of the core in (B) by acknowledging something that might otherwise interfere with this intended effect. Another kind of intentional relation is evidence, in which the contributors are intended to increase the hearer's belief in the core. For example, (C) stands in the evidence relation to (B). The set of intentional relations in RDA is a modification of the presentational relations of RST. Each *core:contributor* pair is also analyzed for its informational relation. These relations describe how the situations referred to by the core and contributor are related in the domain.

The RDA analysis of the example in Figure 1 is shown schematically in Figure 2. As a convention, the core appears as the mother of all the relations it participates in. Each relation is labeled with both its intentional and informational relation, with the order of relata in the label indicating the linear order in the discourse. Each relation node has up to two daughters: the cue, if any, and the contributor, in the order they appear in the discourse.

2 Reliability of RDA application

To assess inter-coder reliability of RDA analyses, we compared two independent analyses of the same data. Because the results reported in this paper depend only on the structural aspects of the analysis, our reliability assessment is confined to these. The

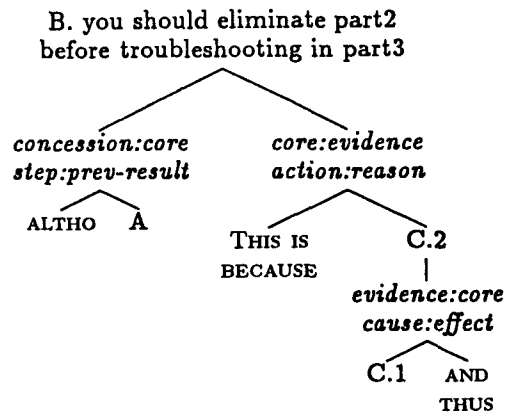


Figure 2: The RDA analysis of the example in Figure 1

categorization of *core:contributor* relations will not be assessed here.

The reliability coder coded one quarter of the currently analyzed corpus, consisting of 132 clauses, 51 segments, and 70 relations. Here we report the percentage of instances for which the reliability coder agreed with the main coder on the various aspects of coding.

There are several kinds of judgements made in an RDA analysis, and all of them are possible sources of disagreement. First, the two coders could analyze a contributor as supporting different cores. This occurred 7 times (90% agreement). Second, the coders could disagree on the core of a segment. This occurred 2 times (97% agreement). Third, the coders could disagree on which relation a cue was associated with. This occurred 1 time (98% agreement). The final source of disagreement reflects more of a theoretical question than a question of reliable analysis. The coders could disagree on whether a relatum should be further analyzed into an embedded *core:contributor* structure. This occurred 8 times (91% agreement).

These rates of agreement cannot be sensibly compared to those found in studies of (nonembedded) segmentation agreement (Grosz and Hirschberg, 1992; Passonneau and Litman, 1993; Hearst, 1994) because our assessment of RDA reliability differs from this work in several key ways. First, the RDA coding task is more complex than identifying locations of segment boundaries. Second, our subjects/coders are not naive about their task; they are trained. Finally, the data is not spoken as in these other studies.

Future work will include a more extensive reliability study, one that includes the intentional and informational relations.

3 Initial results and their application

For each tutor explanation in our corpus, each coder analyzes the text as described above, and then enters this analysis into a database. The technique of representing an analysis in a database and then using database queries to test hypotheses is similar to work using RST analyses to investigate the form of purpose clauses (Vander Linden et al., 1992). Because our analysis is exhaustive, information about both occurrence and nonoccurrence of cues can be retrieved from the database in order to test and modify hypotheses about cue usage. That is, both cue-based and factor-based retrievals are possible. In cue-based retrievals, we use an occurrence of the cue under investigation as the criterion for retrieving the value of its hypothesized descriptive factors. Factor-based retrievals provide information about cues that is unique to this study. In factor-based retrieval, the occurrence of a combination of descriptive factor values is the criteria for retrieving the accompanying cues. In this section, we report two results, one from each perspective: a comparison of the distribution of *SINCE* and *BECAUSE* in our corpus, and the impact of embeddedness on cue selection.

These results are based on the portion of our corpus that is analyzed and entered into the database, approximately 528 clauses. These clauses comprise 216 segments in which 287 relations were analyzed. Accompanying these relations were 165 cue occurrences, resulting from 39 distinct cues.

3.1 Choice of "Since" or "Because"

SINCE and *BECAUSE* were two of the most frequently used cues in our corpus, occurring 23 and 13 times, respectively. To investigate their distribution, we began with the proposal of Elhadad and McKeown (1990). As with our study, their work aims to define each cue in terms of features of the propositions it connects for the purpose of cue selection during text generation. Their work relies on the literature and intuitions to identify these features, and thus provides an important background for a corpus study by suggesting features to include in the corpus analysis and initial hypotheses to investigate.

Quirk et al. (1972) note several distributional differences between the two cues: (i) *SINCE* is used when the contributor precedes the core, whereas *BECAUSE* typically occurs when the core precedes the contributor, (ii) *BECAUSE* can be used to directly answer a *why* question, whereas *SINCE* cannot, and (iii) *BECAUSE* can be in the focus position of an *it*-cleft, whereas *SINCE* cannot. These distributional differences are reflected in our corpus, and the ordering difference (i) is of particular interest. *SINCE* and *BECAUSE* are always placed with a contributor. All but one (22/23) occurrences of *SINCE* accompanied relations in *contributor:core* order, while all (13/13) occurrences of

BECAUSE accompanied relations in *core:contributor* order².

The crucial factor in distinguishing between *SINCE* and *BECAUSE* is the relative order of core and contributor. Elhadad and McKeown (1990) claim that the two cues differ with respect to what Ducrot (1983) calls "polyphony", i.e., whether the subordinate relatum is attributed to the hearer or to the speaker. The idea is that *SINCE* is used when a relatum has its informational source with the hearer (e.g., by being previously said or otherwise conveyed by the hearer). *BECAUSE* is monophonous, i.e., its relata originate from a single utterer, while *SINCE* can be polyphonous. According to Elhadad and McKeown, polyphony is a kind of given-new distinction and thus the ordering difference between the two cues reduces to the well-known tendency for given to precede new. Unfortunately, this characterization of the distinction between *SINCE* and *BECAUSE* is not supported by our corpus study.

As shown in Figure 3, whether or not contributors could be attributed to the hearer did *not* correlate with the choice of *SINCE* or *BECAUSE*. To judge whether a contributor is attributable to the student, mention of an action or result of a test that the student previously performed (e.g., *you tested 30 to ground earlier*) was counted as 'yes', while information available by observation (e.g., *part1 and part2 are connected by wires*), specialized circuit knowledge (e.g., *part1 is used by this test step*) and general knowledge (e.g., *part2 is more prone to damage*) were counted as 'no'.

Is contributor attributable to student?	Cue choice	
	<i>SINCE</i>	<i>BECAUSE</i>
yes	13	9
no	10	4

Figure 3: Polyphony does not underlie the choice between *SINCE* and *BECAUSE*.

This result shows that the choice between *SINCE* and *BECAUSE* is determined by something other than the attributability of contributor to hearer. In future work, we will consider other factors that may determine ordering as possible alternative accounts for this choice. Another factor to be considered in distinguishing the two cues is the embeddedness discussed in the next section. Furthermore, this result demonstrates the need to move beyond small numbers of constructed examples and intuitions formed

²This included answers that begin with *BECAUSE*. In these cases, we took the core to be the presupposition to the question.

from unsystematic analyses of naturally occurring data. Only by an exhaustive analysis such as ours can hypotheses such as the one discussed here be systematically evaluated.

3.2 Effect of Segment Embeddedness on Cue Selection

The second question we report on here concerns whether segment embeddedness affects cue selection. Much of the work on cue usage, e.g., (Elhadad and McKeown, 1990; Millis et al., 1993; Schiffrin, 1987; Zukerman, 1990) has focused on pairs of text spans, and this has led to the development of heuristics for cue selection that take into account the relation between the spans and other local features of the two relata (e.g., relative ordering of core and contributor, complexity of each span). However, analysis of our corpus led us to hypothesize that the hierarchical context in which a relation occurs, i.e., what segment(s) the relation is embedded in, is a factor in cue usage.

For example, recall that the relation between C.1 and C.2 in Figure 2 was expressed as *part2 is moved frequently, AND THUS it is more susceptible to damage*. Now, the relation between C.1 and C.2 could have been expressed, *BECAUSE part2 is moved frequently, it is more susceptible to damage*. However, this relation is embedded in the contributor of the relation between B and C, which is cued by *THIS IS BECAUSE*. Intuitively, we expect that, when a relation is embedded in another relation already marked by *BECAUSE*, a speaker will select an alternative to *BECAUSE* to mark the embedded relation. That is, two relations, one embedded in the other, should be signaled by different cues. Because RDA analyses capture the hierarchical structure of texts, we were able to explore the effect of embedding on cue selection.

We hypothesized that cue selection for one relation constrains the cue selection for relations embedded in it to be a different cue. To test this hypothesis, we paired each cue occurrence with all the other cue occurrences in the same turn. Then, for each pair of cues in the same turn, it was categorized in two ways: (1) the embeddedness of the relations associated with the two cues, and (2) whether the two cues are the same, alternatives or different. Two cues are alternatives when their use with a relation would contribute (approximately) the same semantic content³. The sets of alternatives in our data are {ALSO,AND}, {BUT,ALTHOUGH,HOWEVER} and

³Because it is based on a test of intersubstitutability, the taxonomy proposed by Knott and Dale (1994) does not establish the sets of alternatives that are of interest here. Two cues may be intersubstitutable in some contexts but not semantic alternatives (e.g., AND and BECAUSE), or they may be semantic alternatives but not intersubstitutable because they are placed in different positions in a relation (e.g., SO and BECAUSE).

{BECAUSE,SINCE,SO,THUS,THEREFORE}. The question is whether the choice between the same and an alternate cue correlates with the embeddedness of the two relations.

As shown in Figure 4, we can conclude that, when a relation is going to have a cue that is semantically similar to the cue of a relation it is embedded in, an alternative cue must be chosen. Other researchers in text generation recognized the need to avoid repetition of cues within a single text and devised heuristics such as “avoid repeating the same connective as long as there are others available” (Roesner and Stede, 1992). Our results show that this heuristic is over constraining. The first column of Figure 4 shows that the same cue may occur within a single explanation as long as there is no embedding between the two relations being cued. Based on these results, our text generation algorithm will use embeddedness as a factor in cue selection.

Are relations embedded?	Cue choice	
	Same	Alternate
yes	0	7
no	6	18

Figure 4: Embeddedness correlates with choice between same and alternate cues.

4 Conclusions

We have introduced Relational Discourse Analysis, a coding scheme for the exhaustive analysis of text or single speaker discourse. RDA is a synthesis of ideas from two theories of discourse structure (Grosz and Sidner, 1986; Mann and Thompson, 1988). It provides a system for analyzing discourse and formulating hypotheses about cue selection and placement. The corpus study results in rules for cue selection and placement that will then be exercised by our text generator. Evaluation of these automatically generated texts forms the basis for further exploration of the corpus and subsequent refinement of the rules for cue selection and placement.

Two initial results from the corpus study were reported. While the factor of *core:contributor* order accounted for the choice between *SINCE* and *BECAUSE*, this factor could not be explained in terms of whether the contributor can be attributed to the hearer. Alternative explanations for the ordering factor will be explored in future work, including other types given-new distinctions and larger contextual factors such as focus. Second, the cue selection for one relation was found to constrain the cue selection for embedded relations to be distinct cues. Both of these results are being implemented in our text generator.

Acknowledgments

The research described in this paper was supported by the Office of Naval Research, Cognitive and Neural Sciences Division (Grant Number: N00014-91-J-1694), and a grant from the DoD FY93 Augmentation of Awards for Science and Engineering Research Training (ASSERT) Program (Grant Number: N00014-93-I-0812). We are grateful to Erin Glendening for her patient and careful coding and database entry, and to Maria Gordin for her reliability coding.

References

- O. Ducrot. 1983. *Le sens commun. Le dire et le dit*. Les editions de Minuit, Paris.
- Michael Elhadad and Kathleen McKeown. 1990. Generating connectives. In *Proceedings of the Thirteenth International Conference on Computational Linguistics*, pages 97-101, Helsinki.
- Susan R. Goldman and John D. Murray. 1992. Knowledge of connectors as cohesion devices in text: A comparative study of native-english speakers. *Journal of Educational Psychology*, 44(4):504-519.
- Barbara Grosz and Julia Hirschberg. 1992. Some intonational characteristics of discourse structure. In *Proceedings of the International Conference on Spoken Language Processing*.
- Barbara J. Grosz and Candace L. Sidner. 1986. Attention, intention, and the structure of discourse. *Computational Linguistics*, 12(3):175-204.
- Marti Hearst. 1994. Multi-paragraph segmentation of expository discourse. In *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics*.
- Julia Hirschberg and Diane Litman. 1993. Empirical studies on the disambiguation of cue phrases. *Computational Linguistics*, 19(3):501-530.
- Jerry R. Hobbs. 1985. On the coherence and structure of discourse. Technical Report CSLI-85-37, Center for the Study of Language and Information, Leland Stanford Junior University, Stanford, California, October.
- Alistair Knott and Robert Dale. 1994. Using linguistic phenomena to motivate a set of coherence relations. *Discourse Processes*, 18(1):35-62.
- Diane J. Litman and James F. Allen. 1987. A plan recognition model for subdialogues in conversations. *Cognitive Science*, 11:163-200.
- Robert Lorch. 1989. Text signaling devices and their effects on reading and memory processes. *Educational Psychology Review*, 1:209-234.
- William C. Mann and Sandra A. Thompson. 1988. Rhetorical Structure Theory: Towards a functional theory of text organization. *TEXT*, 8(3):243-281.
- Danielle S. McNamara, Eileen Kintsch, Nancy Butler Songer, and Walter Kintsch. In press. Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*.
- Keith Millis, Arthur Graesser, and Karl Haberlandt. 1993. The impact of connectives on the memory for expository text. *Applied Cognitive Psychology*, 7:317-339.
- Johanna D. Moore and Martha E. Pollack. 1992. A problem for RST: The need for multi-level discourse analysis. *Computational Linguistics*, 18(4):537-544.
- Megan Moser and Johanna D. Moore. 1993. Investigating discourse relations. In *Proceedings of the ACL Workshop on Intentionality and Structure in Discourse Relations*, pages 94-98.
- Rebecca Passonneau and Diane Litman. 1993. Intention-based segmentation: Human reliability and correlation with linguistic cues. In *Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics*.
- Randolph Quirk et al. 1972. *A Grammar of Contemporary English*. Longman, London.
- Dietmar Roesner and Manfred Stede. 1992. Customizing RST for the automatic production of technical manuals. In R. Dale, E. Hovy, D. Rosner, and O. Stock, editors, *Proceedings of the Sixth International Workshop on Natural Language Generation*, pages 199-215, Berlin. Springer-Verlag.
- Deborah Schiffrin. 1987. *Discourse Markers*. Cambridge University Press, New York.
- Donia Scott and Clarisse Sieckenius de Souza. 1990. Getting the message across in RST-based text generation. In R. Dale, C. Mellish, and M. Zock, editors, *Current Research in Natural Language Generation*, pages 47-73. Academic Press, New York.
- Keith Vander Linden, Susanna Cumming, and James Martin. 1992. Expressing local rhetorical relations in instructional text. Technical Report 92-43, University of Colorado. To appear in *Computational Linguistics*.
- Ingrid Zukerman. 1990. A predictive approach for the generation of rhetorical devices. *Computational Intelligence*, 6(1):25-40.