

Chapter 8

Implicit and explicit coherence relations

Maite Taboada
Simon Fraser University

Jan Renkema, in his *Introduction to Discourse Studies*, describes current research in coherence relations (also called rhetorical or discourse relations), and lists a few open questions in that area. One of the most important issues, from both a theoretical and applied point of view, is the signaling of relations, that is, the explicit marking of the presence of a relation. In this paper, I question the existence of true implicit relations, those where there is no marking that indicates the presence of a relation. I present a number of open questions that can be used as starting points for course assignments and theses in this area.

1. Introduction

One of the fundamental issues in the study of discourse is the phenomenon of coherence. People perceive discourse as being coherent (or incoherent), but often fail to explain the source of that coherence. In discourse studies, coherence is often described as the way in which a discourse is sewn together, with pieces relating to other pieces. Mann and Thompson (1988) defined it as the absence of non-sequiturs, i.e., a coherent text is one where all the parts form the whole: “for every part of a coherent text, there is some function, some plausible reason for its presence, evident to readers, and furthermore, there is no sense that some parts are somehow missing” (Mann & Taboada, 2007). Renkema (2004: 103) indicates that coherence refers to “the connections which can be made by the reader or listener based on knowledge outside the discourse.” Those connections are often captured in the form of coherence relations.

The relations I am concerned with here are referred to as coherence relations, discourse relations, or rhetorical relations. They are paratactic (coordinate) or hypotactic (subordinate) relations that hold across two or more text spans. When building a text or any instance of discourse, just as when building a sentence, speakers choose

among a set of alternatives that relate two portions of the text. The two parts of the text that have been thus linked can then enter, as a unit, into another relation, making the process recursive throughout the text. Coherence relations have been proposed as an explanation for the construction of coherence in discourse. It is not clear how much speakers and hearers are aware of their presence, but it is uncontroversial that hearers and readers process text incrementally, adding new information to a representation of the ongoing discourse (e.g., van Dijk & Kintsch, 1983).

There are many classifications and a variety of labels for coherence relations. Renkema (2004) proposes to divide them into additive and causal relations. Additive relations relate two spans of text (clauses, sentence, or larger portions) hypotactically, i.e., through a form of coordination. For example, in (1), the relationship between spans 2 and 3 is one of Contrast. Span 1 is provided as additional context.¹

- (1) [1] Billy Bob Thornton is such a versatile actor. [2] He played character-driven supporting roles [3] and now he's lead in a black comedy.
(Epinions)

Causal relations, on the other hand, are paratactic (or subordinating), relating a main and a subordinate part (sometimes referred to as nucleus and satellite), and express cause, reason, purpose, condition or concession. Example (2) shows a Purpose relation, with span 2 as the satellite, or less important part.

- (2) [1] Indeed, Mike Myers recycled his entire CV of SNL characters [2] to create a Cat in the Hat that is unworthy of his name.
(Epinions)

Relations hold at all levels in a text, from the clause up. Typically, the clause is considered the minimal unit of analysis. In (2), the relation is between two clauses; in (1), between two sentences. In that example, there is a further relation between unit 1 and the unit made up of 2 and 3, which we could label as Evidence: Units 2–3 provide evidence for the claim in 1. A graphical representation of Example (1) is shown in Figure 1. Straight lines, such as those joining units 2 and 3, represent nuclei, and an arrow represents a satellite (the origin) attaching to a nucleus (the end point of the arrow). We can see that in this fashion, texts can be built recursively, with smaller units becoming part of larger units.

Throughout this chapter, I will be following Rhetorical Structure Theory, and its assumptions about discourse structure and coherence relations. Space precludes a more extensive discussion of the theory itself. More detail can be found in the original paper on RST (Mann & Thompson, 1988), a recent overview (Taboada & Mann, 2006a, 2006b), the RST website (Mann & Taboada, 2007), or in *Introduction to*

1. For a list of example sources, please see the Appendix.

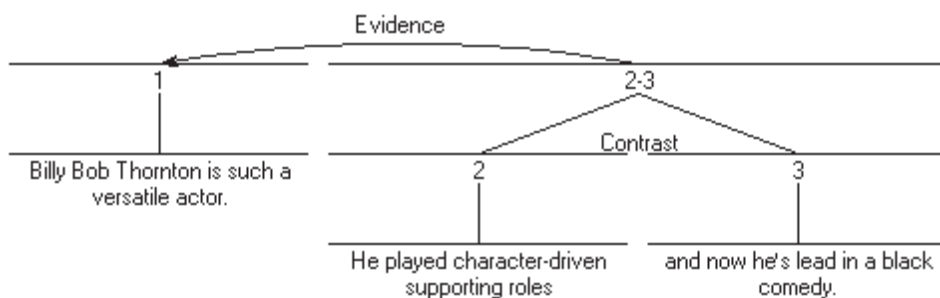


Figure 1. Structure of Example (1)

Discourse Studies (Renkema, 2004, Section 6.4). The types of relations I will be using will be hopefully self-explanatory from their labels (Evidence, Condition, Elaboration, etc.), but full definitions are available on the RST website.

2. The research challenge

One of the issues in the study of coherence relations is how to recognize them, both from the point of view of the analyst, and from the point of view of the hearer or reader. Some relations have cues that help identify them, such as conjunctions (*if, although, but*). These Renkema calls explicit relations (2004, Section 6.5). Many relations, however, do not contain clear indicators, and thus are implicit. The research challenge I present leads to questioning whether there are any true implicit relations.

The literature on coherence relations abounds with studies on signaling via cue phrases, also called discourse markers or discourse particles. For instance, in example (3), an invented example, one could easily interpret a causal relation between (3a) and (3b): The reason why Tom quit was that he was tired of the long hours. This can be made explicit through the marker *because*, as in (4). A different marker can, however, indicate that there is no strong causal relation between the two sentences, as in (5). Long hours may have been a factor in Tom's decision to quit, but (5) seems to imply that there was another reason for the decision.

- (3) a. Tom quit his job.
- b. He was tired of the long hours.
- (4) Tom quit his job because he was tired of the long hours.
- (5) Tom quit his job. He was tired of the long hours, anyway.

There are many other signaling mechanisms besides discourse markers: morphological, syntactic, semantic, and pragmatic. Morphologically, tense, for instance, helps mark temporal relations, guiding the reader in the interpretation of progressions or

flashbacks in time. One syntactic mechanism is sentence mood (indicative, imperative, interrogative). Another mechanism is embedding, such as the use of a relative clause for Elaboration. Semantically, verb meaning can point to certain relations: *cause*, *trigger*, *provoke*, or *effect* can all indicate a causal relation. Pragmatically, phenomena such as implicature establish relations between propositions that are not explicitly present in the text, but are constructed in the minds of the speakers.

Few studies have considered all possible signals for discourse relations, and thus the perception is that signaling is low. Studies that consider only discourse markers usually show that over 50% of the relations are unsignaled. The analyses on the RST website (Mann & Taboada, 2007), a very diverse collection comprising 187 units, have about 72% of the relations unsignaled (by a discourse marker). In a study of two different corpora, I found that relations were signaled about 31% of the time in conversation, and 44% in newspaper articles (Taboada, 2006), again with reference mostly to discourse markers, although a few other signals are discussed in that paper (mood, finiteness and punctuation).

Many other signaling mechanisms remain understudied. One proposal that I have put forth (Taboada, 2006) is that expectations about how texts (and possibly conversations) proceed provide enough information to interpret higher-level relations. For instance, a reader may recognize the last few sentences of a text as a summary of the whole text because he or she is familiar with the general structure of texts, and knows that a conclusion or summary typically appears at the end.

My proposal here goes a bit further: It may be the case that all relations are indeed signaled, that is, that they are all explicit. The challenge lies in finding what the particular signal is in each case. If people truly interpret different types of relations with relative ease, they must be using signals to guide that interpretation. This leads to two different problems: Establishing that relations are cognitively represented in the minds of hearers and readers; and, if indeed relations are cognitively plausible, discovering the cues used to interpret them.

In the rest of this paper, I review some of the exiting literature on relation signaling, and propose some research topics in this area. Section 3 presents examples of the different types of markers or signals for relations, and Section 4 a research method to explore different types of relation signaling, with a more detailed proposal in Section 5. Finally, Sections 6 and 7 present practical applications of the research, and further pedagogical resources.

3. Examples

Examples of signaling via discourse markers are widely discussed in the literature, and a summary of those studies can be found in Taboada and Mann (2006b). I am

using the term “discourse marker” in a loose sense, to refer to any conjunction, adverb, adverbial phrase or other type of phrase that frequently links two or more units of discourse. Examples are *if, because, then, plus, on the other hand, in summary* or *it follows that*. Discourse markers and their functions are well-studied, within a coherence relations framework or not. Examples of relations signaled by discourse markers are frequently found in research on coherence relations, and some of the examples discussed here have appeared in print elsewhere (Taboada, 2004, 2006; Taboada & Mann, 2006b).

In example (6) below, in Spanish, the situation in (6b) is presented as a Result of the constraints in (6a): given that the speaker is busy at certain times, the result is that Friday at twelve is the best time to meet. This is marked with *por lo tanto*, “as a consequence” or “so.”

- (6) a. Ah yo tengo clases el viernes de las diez a las once y de las dos a las tres,
 b. por lo tanto si (es) tú me podrías ver a las doce, ese mismo día me parecería muy bien.
 Uh I have classes on Friday from ten to eleven and from two to three, / so if (it's) you could see me at twelve, that same day I would like that.
 (ISL Corpus)

The next example, from the RST website (Mann & Taboada, 2007), shows a condition relation between spans (7b) and (7c), signaled by the prepositional phrase “in the event that.”

- (7) [Copyright notice]
 a. This notice must not be removed from the software,
 b. and in the event that the software is divided,
 c. it should be attached to every part.
 (RST Website)

Lexical items may also be used to indicate a relation, such as the verb *cause* in a causal relation, or *concede*, as in example (8) below, which in this case marks a Concession relation.

- (8) [S] Some entrepreneurs say the red tape they most love to hate is red tape they would also hate to lose. [N] They concede that much of the government meddling that torments them is essential to the public good, and even to their own businesses.
 (RST Discourse Treebank)

Sentence mood can indicate Solutionhood, as in example (9), where the relation between A's turn and B's entire turn is one of Solutionhood (there are further relations in B's turn), signaled by the interrogative mood in the satellite.

- (9) A: Qué dices. Que tienes entre las [pause] diez y las dos libre?
 B: Sí. Tengo clase de dos a cinco. Así que entre las diez y las dos, o a la una y media, estoy libre.
- A: What did you say. That you are free between ten and two?
 B: Yes. I have a class from two to five. So that between ten and two, or at one thirty, I'm free.
 (ISL Corpus)

Verb finiteness is sometimes the only indicator of a relation, as shown in example (10), from the LDC (Linguistic Data Consortium) corpus of Wall Street Journal articles (Carlson et al., 2002). The Circumstance relationship between spans 1 and 2–5 is signaled by the non-finite form of the verb *insisting*.

- (10) [1] *Insisting* that they are protected by the Voting Rights Act, [2] a group of whites brought a federal suit in 1987 [3] to demand that the city abandon at-large voting for the nine-member City Council [4] and create nine electoral districts, [5] including four safe white districts.
 (RST Discourse Treebank)

In Example (11), there is an Evaluation relation between segment 1 and 2. The author characterizes the narrator for the novel “The Wedding” as a character removed from the main protagonist, Noah, and therefore making the connection between narrator and protagonist quite indirect. The main indicator of this Evaluation relation is the semantic content of the word “indirect,” an adjective conveying subjective content.

- (11) [1] The first-person narrator of “The Wedding” is the son-in-law (Wilson) of Noah’s daughter Jane. [2] ?????? Talk about indirect.
 (Epinions)

In written text, punctuation is also used as a signal of certain relations. In example (12), there are two embedded Elaboration relations. The first one is signaled by a parenthesis, and elaborates on the first part of the news item. The second elaboration is signaled by a dash, and it elaborates on the title of the article quoted.²

- (12) [N1] QUANTUM CHEMICAL Corp.’s plant in Morris, Ill., is expected to resume production in early 1990. The year was misstated in Friday’s editions. [S1] ([N2] See: “Dividend News: Payout Stalled at Quantum Chemical Corp. — [S2] Firm Posts Quarterly Loss, Plans a Stock Dividend to Take Place of Cash — WSJ Oct. 27, 1989).
 (RST Discourse Treebank)

2. For a tree representation of this example, see Taboada (2006: 585).

Example (13) provides another instance of embedded relations, in this case Elaborations. In the first relation, the satellite starts with “Recently, the boards...” and continues to the end of the paragraph, which is longer than displayed in the example here. The only possible signal that an Elaboration relation is present is the adverb *also* before the main verb *voted* in this satellite. The second Elaboration relation has that “Recently, the boards...” sentence plus the next sentence as nucleus. The satellite starts with “The transaction...” and continues for a while. This second satellite has no adverb, punctuation mark, or any other device that indicates an elaboration on what has gone before. Knowledge of the newspaper genre leads us to think that an article, unless other cues are present, proceeds in a series of elaborations.

- (13) [N1] American Pioneer Inc. said it agreed in principle to sell its American Pioneer Life Insurance Co. Subsidiary to Harcourt Brace Jovanovich Inc.’s HBJ Insurance Cos. for \$27 million. American Pioneer, parent of American Pioneer Savings Bank, said the sale will add capital and reduce the level of investments in subsidiaries for the thrift holding company. [S1] [N2] Recently, the boards of both the parent company and the thrift also voted to suspend dividends on preferred shares of both companies and convert all preferred into common shares. The company said the move was necessary to meet capital requirements. [S2] The transaction is subject to execution of a definitive purchase agreement and approval by various regulatory agencies, including the insurance departments of the states of Florida and Indiana, the company said. [...]
(RST Discourse Treebank)

The next example shows the use of word order in German to signal a Contrast relation. Units 2 and 3 are in a Contrast relation with each other, with the entire 2–3 span being Evidence for the claim presented in 1, that the deployment of the Army resulted in a sort of personal conflict for some members of the German Green party. The two spans that are in contrast, 2 and 3, have subject-verb inversion,³ which is likely an indication of the contrast, conveyed in the English translation by the conjunction *while*. In addition, there is a “then versus now” contrast in the *nun* — *noch* pair.

- (14) [1] Es ist offensichtlich, dass dieser Bundeswehr-Einsatz bei vielen Grünen an die Wurzeln ihrer Identität rührt. [2] Konnte der Kosovo-Krieg noch mit dem Kampf für die Menschenrechte in Europa begründet werden, [3] geht es nun um die militärische Solidarität mit den USA in einem weit entfernten Dritte-Welt-Land.

It is obvious that this deployment of the German Army touches, for many Greens, on the roots of their identity. While the Kosovo War could still be

3. Strictly speaking, the choice is only for span 2; span 3 would be ungrammatically with subject-verb word order.

justified by the fight for human rights in Europe, it is now about military solidarity with the U.S. in a far away Third World country.
(Potsdam Commentary Corpus, maz-ernsfall-9725)

4. Research method

The problem that I raise in this paper can be explored using two different methods: corpus analysis and psycholinguistic experimentation. In this section, I provide general information on how to approach the research using these methodologies.

First of all, corpus linguistics is a well-known area of research for which general textbooks exist (e.g., Baker, 2006; Biber et al., 1998; Teubert & Cermakova, 2007). I assume that readers understand the need to carry out quantitative corpus research to find general linguistic trends. I will concentrate here on the particular issues around annotating and studying coherence relations in discourse.

In order to find out how relations are marked, one first needs to find the relations. This involves performing a full analysis of texts and annotating the relations. One can simply take pencil to paper and mark relations as they are found. More sophisticated methods involve the use of computer tools to mark the relations. In particular, within Rhetorical Structure Theory, the RSTTool⁴ provides a graphic interface to annotate relations, indicating nucleus and satellite status and providing a label, which can be chosen from the standard RST set, or from a purpose-built one. The next step is to find particular cues for each of the relations. Unfortunately, this cannot be done automatically with RSTTool, and thus other tools are needed. A simple database could do the trick.

Obviously, the task of finding relation signals would be much easier if the relations had been identified for us already. This is precisely what existing corpora provide. The most frequently cited corpus is the RST Discourse Treebank (Carlson et al., 2002), a collection of Wall Street Journal articles annotated according to a version of RST. The corpus contains 385 articles published in the Wall Street Journal, a subset of the large Penn Treebank (Marcus et al., 1999). A portion of the articles was also annotated following a graph structure (Wolf et al., 2005). For German, Stede and colleagues have created a corpus of German newspaper editorials annotated according to RST (Stede, 2004).

Whether one annotates a corpus, or uses an existing one, an important issue is the lack of standards across annotations. Unfortunately, the field has not produced a consensus on what the annotation standard should be. Different researchers use

4. Downloadable from <http://www.wagsoft.com/RSTTool>. Mick O'Donnell is owed thanks and praise for creating the tool and for making it freely available to the research community.

varying inventories of relations, and even different definitions of minimal discourse units. Rather than accepting this as an unfortunate problem, researchers could take this issue up as a challenge. An important contribution would be the creation of a standard for annotation and the conversion of existing corpora to that new standard. Flexibility can be built into the proposal. For example, relation sets can be more or less detailed, as long as mappings between relation sets are always available (e.g., if a researcher wishes to distinguish Cause and Result relations, that is fine, as long as the new distinction can be easily mapped onto a more general Causal relation).

Reliability of the analysis is of foremost importance in any annotation effort. Even if the annotations are performed by a single person, they should be transparent and easy to reproduce. Better yet, a test comparison with another annotator with similar experience would lend validity to the annotation scheme. Reliability measures are well studied and widely applied, especially within computational linguistics (Carletta, 1996; Di Eugenio & Glass, 2004; Passonneau, 2004). Studies that specifically address reliability of rhetorical structure annotation are den Ouden et al. (1998), Marcu et al. (1999) and Wolf and Gibson (2005).

Thus far, I have provided general information on how to find or create corpora and make sure that the annotations are reliable. The next task in such a project is the one that leads to its main focus: to find signals for discourse relations. My assumption is that, if an annotator reliably found a connection between two pieces of discourse, there must be something in those pieces that led the annotator to link them. As I mentioned in Section 3, discourse markers are the first line of action, because they tend to be the most obvious signals that a link exists. Syntactic information provides cues in a variety of cases: reported speech and certain verbs indicate an Attribution relation (Redeker & Egg, 2006); relative clauses indicate Elaboration; interrogative mood signals Solutionhood; and non-finite clauses Circumstance (Taboada & Mann, 2006b). Lexical or cohesive chains may be indicators of an Elaboration relation. Punctuation and layout are also indicators of relations. Finally, genre-related structures probably play a role, in terms of the types of relations that are typically found in a particular text type, such as Preparation and Background in newspaper articles, and also in what relations are most frequent in each part of a text or conversation (Taboada, 2004).

The second method for finding markers for relations involves experimentation. Again, the assumption is that readers are able to recognize relations. If so, they must be using cues in the text. Previous research has shown a positive effect in comprehension when discourse markers are present (e.g., Degand & Sanders, 2002).

In all cases, one should bear in mind that there is probably no one-to-one mapping between a relation and any given marking. It remains to be seen how much ambiguity is for most relations, and whether it is always resolved, or whether some instances of relations remain underspecified in the mental representation of the hearer or reader.

Beyond the signals already discussed in this section, psycholinguistic experimentation could be explored to discover other signals that we have not yet considered. For instance, if subjects shown two segments, one word at a time, make a connection between segment 1 and segment 2 at a particular point, we could infer something about the signals present up to that point.

Psycholinguistic experiments will also lend validity to findings from corpus studies. Following classical experiments, subjects can be shown the same text, but presented as an example of different genres. If the types of relations subjects find in the texts are different, then we can say that genre is a signal. Another avenue already partly explored is the use of eye-tracking devices to pinpoint the place where readers establish a connection (Mak & Sanders, 2008).

I have presented the research methodology assuming the study of a single language (e.g., English). A cross-linguistic study, albeit more complex, would shed light on different mechanisms to signal a relation. Fabricius-Hansen (2005), for instance, discusses the interesting absence of certain German connectives in English translation. If similar relations can be hypothesized in the English translations, then the relation must be signaled through other means. It would be valuable to undertake a similar study and, in general, other cross-linguistic work, whether on parallel corpora or not.

5. Recent research

Previous reports of relation signaling, including my own work, have given figures of up to 70% of relations without signals. I have already pointed out that most of this research considered only discourse markers. In fact, some lines of research study coherence relations only if they are signaled by a marker. In Section 3, I also discussed work that takes into account other sources of signaling.

Let's start with my own data. For my book (Taboada, 2004), I analyzed 30 conversations in English and 30 in Spanish, annotating them with rhetorical relations. On page 149, I present a table with each relation and the number of times a discourse marker was used to signal it. The summary is that relations were marked 30.86% of the time in English, and 45.88% in Spanish. My very narrow definition of discourse markers (p. 148) includes discourse connectives (*and, but, or*) and subordinating conjunctions (*because, so, if*). I excluded some elements that are elsewhere considered discourse markers, such as sentence adverbials (*also, on the other hand*). An extension of this work could involve studying those markers.

In Taboada (2006), I extended the types of markers considered, addressing punctuation, finiteness and order of segments. I also suggested that genre and other factors may play a role in the interpretation of relations. Current research aims at extending

that study with all other cues mentioned in the previous section: punctuation, layout, syntactic cues, semantic relations, lexical and grammatical cohesion, and genre.

In this chapter, I have mostly focused on relation signaling for written discourse, including punctuation and paragraph breaks. Obviously, spoken language brings new types of signaling into play. Intonation, pauses and gesture are all potential indicators of relations. Cassell et al. (2001) found a correlation between posture changes and the boundaries of topics, with new discourse segments more likely to be accompanied by posture shifts. It is possible that posture also plays a role in signaling particular types of relations, not just a change in topic. Den Ouden (2004) found that pause duration and pitch were strong indicators of the RST structure of read-aloud texts.

6. Research proposal

The following is a proposal for a PhD dissertation that can be, of course, amended, shortened or expanded to accommodate different interests. The proposal combines corpus and psycholinguistic experimentation, but can be modified to emphasize one or the other. Other research topics can also be found on the RST website, under “Research Topics” (Mann & Taboada, 2007).

The first step in the project involves corpus research. This could be an existing corpus that has already been annotated, to which the PhD student can add further signals, or a newly created corpus. The goal is to annotate all the relations found in the corpus first, and then look for signals.

In previous sections I have discussed some of the signals studied in the literature, including lexical cues, syntactic structure, punctuation and layout. However, there is not, to my knowledge, any work that has explored, in one study, all the signals mentioned in the literature. This could be a valuable research initiative.

A slightly different version of the corpus research could involve searching patterns in corpora. For instance, consider the possible continuations to the first (invented) sentence in (1). The relation with one of the following alternatives is one of Sequence, indicated by the adverb *then*. However, the adverb placement is different in each of the three versions. Presumably, in (1c), the hearer or reader needs to wait longer to establish the connection. One possible project involves investigating the frequency of such patterns in large corpora. If one pattern is more frequent, it may be because it correlates with ease of processing.

- (1) Sonia came home at five.
 - a. Then we went out.
 - b. We then went out.
 - c. We went out then.

The next step in the project would then be to establish whether the corpus frequencies are replicated in psycholinguistic studies. Do subjects take longer to process (1c) than (1a)? Research along these lines, tying in corpus findings with experimental work, would be a valuable contribution to the field.

Another line of psycholinguistic-based research involves investigating order of acquisition of relations, and types of signaling, both in first and second language learners. It would be interesting to see if relations that are more unambiguously signaled are acquired earlier. There is some research in this area (Degand & Sanders, 2002; Pelsmaekers et al., 1998; Spooren, 1997), but further studies including more types of signaling are necessary.

It is important to point out, before closing this proposal outline, that many issues need to be solved before annotation or analysis can take place. Deciding on the number and type of relations, establishing a reliable annotation process, and even deciding on the representation of the relations (trees vs. graphs, for instance), are all contentious issues in the literature on discourse relations. For an overview, see Taboada and Mann (2006a; 2006b).

7. Practical relevance

Coherence is the basis of all communication, and coherence relations are a very important aspect of the perception of coherence in discourse. Understanding how coherence relations are signaled and processed has applications in many fields.

The corpus study proposed here will contribute to determining what the most frequent signals for coherence relations are. That knowledge can be put to use in writing and communication. Using the most typical signal for a given relation will ensure that there is no ambiguous interpretation of the relation.

One proposal within the clear language movement is to write in short sentences. Renkema (2004: 184) illustrates the trade-off between shorter sentences and signaling with the following example.

- (16) a. If given a chance before another fire comes, the tree will heal its own wounds by growing new bark over the burned part.
 b. If given a chance before another fire comes, the tree will heal its own wounds. It will grow new bark over the burned part.

The text in (16b) has been distributed into two sentences, conforming to a general guideline to write in short, concise sentences. The problem is that the Means relation that *by* is signaling in (16a) is now implicit and potentially more difficult to identify. The research proposed here could add to manuals on clear writing by emphasizing the

need for explicit marking of relations, if that is indeed what both the corpus analysis and the experiments allow us to conclude.

The benefits from research on coherence relations and signaling extend to many other fields that rely on effective communication, from textbook writing to advertising and marketing and document design in general. Clear and unambiguous communication should also be the goal of any organization, including governments. This research could produce guidelines on how to clearly signal the coherence relations writers have in mind. Kamalski (2007), for instance, provides a very interesting study on the effects of coherence marking on the perception and persuasiveness of texts. Her study focused on discourse markers exclusively. Extensions to other types of markers and signals would further our understanding of the effect of signaling on readers' appraisal of texts.

Assignment

A challenge in the annotation of discourse relations involves subjective judgments about how to segment the text, which parts are more important (the nucleus-satellite distinction), and what relation to assign in any particular instance. The following is a text that has been analyzed in a number of ways. Try to annotate it using a standard set of coherence relations, such as the one available from the RST website (Mann & Taboada, 2007). After you have done so, examine the text again for signals of the relations that you have proposed. Take into account all signals, including discourse markers, other lexical, syntactic, semantic and typographic signals.

Mother Teresa often gives people unexpected advice. When a group of Americans, many in the teaching profession, visited her in Calcutta, they asked her for some advice to take home to their families. "Smile at your wives," she told them. "Smile at your husbands." Thinking that perhaps the counsel was simplistic, coming from an unmarried person, one of them asked, "Are you married?" "Yes," she replied, to their surprise, "and I find it hard sometimes to smile at Jesus. He can be very demanding."

(Extracted from Reader's Digest, January 1986, p. 117)

Appendix: Sources of examples

Examples marked "Epinions" are from a corpus of on-line reviews of movies, books, music and consumer products. Examples are reproduced verbatim, including any typos and grammatical errors. For more information on the corpus, see Taboada et al. (2006).

Examples marked "RST Discourse Treebank" are from the Linguistic Data Consortium corpus of Wall Street Journal articles annotated with RST relations (Carlson et al., 2002).

Examples with the label “ISL Corpus” are part of the Interactive Systems Laboratories Corpus of appointment scheduling conversations, and are described in Taboada (2004).

Examples labeled “RST Website” are from the Rhetorical Structure Theory site: <http://www.sfu.ca/rst> (Mann and Taboada, 2007).

The “Potsdam Commentary Corpus” examples are extracted from the corpus collected and annotated by Manfred Stede and colleagues. They are editorials from *Märkische Allgemeine Zeitung*, a German regional daily newspaper (Stede, 2004). Many thanks to Manfred Stede for permission to reproduce the examples.