

## **Representing discourse coherence: A corpus-based analysis**

Florian Wolf, Edward Gibson  
{fwolf | egibson}@mit.edu

Send correspondence to:

Florian Wolf

Massachusetts Institute of Technology

Department of Brain and Cognitive Sciences

MIT NE20-448

3 Cambridge Center

Cambridge, MA 02139, USA

Phone: ++1 617 452 2474

Email: [fwolf@mit.edu](mailto:fwolf@mit.edu)

Revised manuscript submitted for publication, June 14, 2004; comments welcome

**Abstract**

This paper aims to present a set of discourse structure relations that are easy to code, and to develop criteria for an appropriate data structure for representing these relations. Discourse structure here refers to informational relations that hold between sentences in a discourse. The set of discourse relations introduced here is based on Hobbs (1985). We present evidence that trees are not a descriptively adequate data structure for representing discourse structure: In coherence structures of naturally occurring texts, we found many different kinds of crossed dependencies, as well as many nodes with multiple parents. The claims are supported by statistical results from a database of 135 texts from the Wall Street Journal and the AP Newswire that were hand-annotated with coherence relations, based on the annotation schema presented in this paper.

## 1 Introduction

An important component of natural language discourse understanding and production is having a representation of discourse structure. A coherently structured discourse here is assumed to be a collection of sentences that are in some relation to each other. This paper aims to present a set of discourse structure relations that are easy to code, and to develop criteria for an appropriate data structure for representing these relations.

There have been two kinds of approaches to defining and representing discourse structure and coherence relations. These approaches differ with respect to what kinds of discourse structure they are intended to represent. Some accounts aim to represent the intentional-level structure of a discourse; in these accounts, coherence relations reflect how the role played by one discourse segment with respect to the interlocutors' intentions relates to the role played by another segment (e.g. Grosz & Sidner (1986)). Other accounts aim to represent the informational structure of a discourse; in these accounts, coherence relations reflect how the meaning conveyed by one discourse segment relates to the meaning conveyed by another discourse segment (e.g. Hobbs (1985); Marcu (2000); Webber et al. (1999)). Furthermore, accounts of discourse structure vary greatly with respect to how many discourse relations they assume, ranging from two (Grosz & Sidner (1986)) to over 400 different coherence relations, reported in Hovy & Maier (1995). However, Hovy & Maier (1995) argue that, at least for informational-level accounts, taxonomies with more relations represent subtypes of taxonomies with fewer relations. This means that different informational-level-based taxonomies can be compatible with each other; they differ with respect to how detailed or fine-grained they represent informational structures of texts. Going beyond the question of how different informational-level accounts can be compatible with each other, Moser & Moore (1996) discuss the compatibility of rhetorical structure theory (RST; Mann & Thompson (1988)) to the theory of Grosz & Sidner (1986). However, notice that Moser & Moore (1996) focus on the question of how compatible the claims are that Mann & Thompson (1988) and Grosz & Sidner (1986) make about intentional-level discourse structure.

In this paper, we aim to develop an easy-to-code representation of informational relations that hold between sentences or other non-overlapping segments in a discourse monologue. We describe an account with a small number of relations in order to achieve

more generalizable representations of discourse structures; however, the number is not so small that informational structures that we are interested in are obscured. The goal of the research presented is not to encode intentional relations in texts. We consider annotating intentional relations too difficult to implement in practice at this time. Notice that we do not claim that intentional-level structure of discourse is not relevant to a full account of discourse coherence; it just is not the focus of this paper.

The next section will describe in detail the set of coherence relations we use, which are mostly based on Hobbs (1985). We try to make as few a priori theoretical assumptions about representational data structures as possible. These assumptions will be outlined in the next section. Importantly, however, we do not assume a tree data structure to represent discourse coherence structures. In fact, a major result of this paper is that trees do not seem adequate to represent discourse structures.

This paper is organized as follows. Section 2 describes the procedure we used to collect a database of 135 texts annotated with coherence relations. Section 3 describes in detail the descriptive inadequacy of tree structures for representing discourse coherence, and Section 4 provides statistical evidence from our database that supports these claims. Section 5 contains some concluding remarks.

## **2 Collecting a database of texts annotated with coherence relations**

This section describes (1) how we define discourse segments, (2) which coherence relations we used to connect the discourse segments, and (3) how the annotation procedure worked.

### **2.1 Discourse segments**

There is agreement that discourse segments should be non-overlapping spans of text. However, there is disagreement in the literature about how to define discourse segments (cf. the discussion in Marcu (2000)). While some argue that discourse segments should be prosodic units (Hirschberg & Nakatani (1996)), others argue for intentional units (Grosz & Sidner (1986)), phrasal units (Lascarides & Asher (1993); Longacre (1983); Webber et al. (1999)), or sentences (Hobbs (1985)).

For our database, we mostly adopted a sentence unit-based definition of discourse segments; our method of segmenting discourse is thus very similar to Hobbs (1985). We

chose this method of segmenting discourse because it was easy to use. However, we also assume that contentful coordinating and subordinating conjunctions (cf. Table 1) can delimit discourse segments.

<i>cause-effect</i>	because
<i>violated expectation</i>	although
<i>condition</i>	if...then
<i>similarity</i>	and; (and) similarly
<i>contrast</i>	by contrast
<i>temporal sequence</i>	and then
<i>attribution</i>	according to...
<i>example</i>	for example
<i>elaboration</i>	also, furthermore, in addition
<i>generalization</i>	in general

**Table 1. Contentful conjunctions used to determine coherence relations.**

Notice that we did not classify “and” as delimiting discourse segments if it was used to conjoin nouns in a conjoined noun phrase, like “dairy plants and dealers” in (1) (example from wsj\_0306; Wall Street Journal 1989 corpus; Harman & Liberman (1993)), or if it was used to conjoin verbs in a conjoined verb phrase, like “snowed and rained” in (2) (constructed example):

- (1) Example from wsj\_0306; Wall Street Journal 1989 corpus; Harman & Liberman (1993):

Milk sold to the nation's dairy plants and dealers averaged \$14.50 for each hundred pounds.

- (2) It snowed and rained all day long.

We classified full-stops, semicolons, and commas as delimiting discourse segments. However, in cases like (3) (constructed example), where they conjoin a complex noun phrase, commas were not classified as delimiting discourse segments.

- (3) John bought bananas, apples, and strawberries.

We furthermore treat attributions (“John said that...”) as discourse segments. This is empirically motivated. The texts used here are taken from news corpora, and there, attributions can be important carriers of coherence structures. For instance, consider a case where some Source A and some Source B both comment on some Event X. It should be possible to distinguish between a situation where Source A and Source B make basically the same statement about Event X, and a situation where Source A and Source B make contrasting comments about Event X. Notice, however, that we treated cases like (4) as one discourse segment and not as two separate ones (“...cited” and “transaction costs...”). We only separated attributions if the attributed material was a complementizer phrase, a sentence, or a group of sentences. This is not the case in (4) – the attributed material is a complex NP (“transaction costs from its 1988 recapitalization”).

- (4) Example from wsj\_0667; Wall Street Journal 1989 corpus; Harman & Liberman (1993):

The restaurant operator cited transaction costs from its 1988 recapitalization.

## 2.2 Discourse segment groupings

Adjacent discourse segments could be grouped together. For example, discourse segments were grouped if they all stated something that could be attributed to the same source (cf. Section 2.3 for a definition of *attribution* coherence relations). Furthermore, discourse segments were grouped if they were topically related. For example, if a text discusses inventions in information technology, there could be groups of a few discourse segments each talking about inventions by specific companies. There might also be subgroups, consisting of several discourse segments each talking about specific inventions at specific companies. Thus, marking groups can determine a partially hierarchical structure for the text.

Other examples of discourse segment groupings included cases where several discourse segments described an event or a group of events that all occurred before

another event or another group of events described by another (group of) discourse segments. In that case, what is described by a group of discourse segments is in a *temporal sequence* relation with what is described by another (group of) discourse segments (cf. Section 2.3 for a definition of *temporal sequence* coherence relations). Notice furthermore that in cases where one topic requires one grouping and a following topic requires a grouping that is different from the first grouping, both groupings were annotated.

Unlike approaches like the TextTiling algorithm (Hearst (1997)), we allowed partially overlapping groups of discourse segments. The idea behind that was to allow groupings of discourse segments where a transition discourse segment belongs to the previous as well as the following group. However, this option was not used by the annotators (i.e. in our database of 135 hand-annotated texts, there were no instances of partially overlapping discourse segment groups).

### 2.3 Coherence relations

As pointed out in Section 1, we aim to develop a representation of informational relations between discourse segments. Notice a difference between schema-based approaches (McKeown (1985)) and coherence relations like we use them: whereas schemas are instantiated from information contained in a knowledge base, coherence relations like we use them do not make (direct) reference to a knowledge base.

There are a number of different informational coherence relations, in their basic definitions dating back to Aristotle (cf. Hobbs (1985); Hobbs et al. (1993); Kehler (2002)). The coherence relations we used are mostly based on Hobbs (1985); below we will describe each coherence relation we use and note any differences between ours and Hobbs (1985)'s set of coherence relations (cf. Table 2 for an overview of how our set of coherence relations relates to the set of coherence relations in Hobbs (1985)).

The kinds of coherence relations we used include *cause-effect* relations, as in the constructed example (5), where (5a) states the cause for the effect that is stated in (5b).

#### (5) Cause-Effect

- (5a) There was bad weather at the airport
- (5b) and so our flight got delayed.

Our *cause-effect* relation subsumes the *cause* as well as the *explanation* relation in Hobbs (1985). A *cause* relation holds if a discourse segment stating a cause occurs before a discourse segment stating an effect; an *explanation* relation holds if a discourse segment stating an effect occurs before a discourse segment stating a cause. We encoded this difference by adding a direction to the *cause-effect* relation. In a graph, this can be represented by a directed arc going from cause to effect.

Another kind of causal relation is *condition*. Hobbs (1985) does not distinguish *condition* relations from either *cause* or *explanation* relations. However, we felt that it might be important to distinguish between causal relation describing an actual causal event (*cause-effect*, cf. above) on the one hand, and a causal relation describing a possible causal event (*condition*, cf. below) on the other hand. In the constructed example (6), (6b) states an event that will only take place if the event described by (6a) also takes place.

(6) Condition

- (6a) If the new software works,
- (6b) everyone should be happy.

In a third type of causal relation, the *violated expectation* relation (also *violated expectation* in Hobbs (1985)), a causal relation between two discourse segments that normally would be present is absent. In (7) (constructed example), (7a) normally would be a cause for everyone being happy. This expectation is violated by what is stated by (7b).

(7) Violated Expectation

- (7a) The new software works great,
- (7b) but nobody was happy.



Other possible coherence relations include *similarity* (*parallel* in Hobbs (1985)) or *contrast* relations (also *contrast* in Hobbs (1985)), such as between (8a) and (8b) and (9a) and (9b) respectively (all constructed examples).

(8) Similarity

- (8a) There is a train on Platform A.
- (8b) There is another train on Platform B.

(9) Contrast

- (9a) John supported Schwarzenegger during the campaign
- (9b) but Susan opposed him.

Discourse segments might also *elaborate* (also *elaboration* in Hobbs (1985)) on other sentences, as in (10) (constructed example), where (10b) elaborates on (10a).

(10) Elaboration

- (10a) A probe to Mars was launched from the Ukraine this week.
- (10b) The European-built “Mars Express” is scheduled to reach Mars by late December.

Discourse segments can provide examples for what is stated by another discourse segment. In (11) (constructed example), (11b) states an *example* (*exemplification* in Hobbs (1985)) for what is stated in (11a).

(11) Example

- (11a) There have been many previous missions to Mars.
- (11b) A famous example is the Pathfinder mission.

Hobbs (1985) also includes an *evaluation* relation, as in (12) (example from Hobbs (1985)), where (12b) states an evaluation of what is stated in (12a). We decided to call such relations *elaborations*, since we found it too difficult in practice to reliably

distinguish *elaborations* from *evaluations* (according to our annotation scheme, what is stated in (12b) elaborates on what is stated in (12a)).

(12) Evaluation (example from Hobbs (1985))

(12a) (A story.)

(12b) It was funny at the time.

Unlike Hobbs (1985), we also did not have a separate *background* relation as in (13) (example modified from Hobbs (1985)), where what is stated in (13a) provides the background for what is stated in (13b). Similarly to the *evaluation* relation, we found the *background* relation too difficult to reliably distinguish from *elaboration* relations (according to our annotation scheme, what is stated in (13a) elaborates on what is stated in (13b)).

(13) Background (example modified from Hobbs (1985))

(13a) T is the pointer to the root of a binary tree

(13b) Initialize T.

In a *generalization* relation, as in (14) (constructed example), one discourse segment states a generalization, here (14b), for what is stated by another discourse segment, here (14a).

(14) Generalization

(14a) Two missions to Mars in 1999 failed.

(14b) There are many missions to Mars that have failed.

Furthermore, discourse segments can be in an *attribution* relation, as in (15) (constructed example), where (15a) states the source of what is stated in (15b) (cf. Bergler (1991) for a more detailed semantic analysis of *attribution* relations). Hobbs (1985) does not include an *attribution* relation. However, we decided to include *attribution* as a relation because,

as pointed out in Section 2.1, the texts we annotated are taken from news corpora. There, attributions can be important carriers of coherence structures.

(15) Attribution

- (15a) John said that
- (15b) the weather would be nice tomorrow.

In a *temporal sequence* relation, as in (16) (constructed example), one discourse segment, here (16a), states an event that takes place before another event stated by the other discourse segment, here (16b). In contrast to *cause-effect* relations, there is no causal relation between the events described by the two discourse segments. The *temporal sequence* relation is equivalent to the *occasion* relation in Hobbs (1985).

(16) Temporal Sequence

- (16a) First, John went grocery shopping.
- (16b) Then he disappeared in a liquor store.

The *same* relation, illustrated by (17) (constructed example), is an epiphenomenon of assuming contiguous distinct elements of text (Hobbs (1985) does not include a *same* relation). A *same* relation holds if a subject NP is separated from its predicate by an intervening discourse segment. For example, in (17), (17a) is the subject NP of a predicate in (17c), and so there is a *same* relation between (17a) and (17c); (17a) is the first and (17c) is the second segment of what is actually one single discourse segment, separated by the intervening discourse segment (17b), which is in an *attribution* relation with (17a) (and therefore also (17c), since (17a) and (17c) are actually one single discourse segment).

(17) Same

- (17a) The economy,
- (17b) according to some analysts,
- (17c) is expected to improve by early next year.

Table 2 provides an overview of how our set of coherence relations relates to the set of coherence relations in Hobbs (1985).

<b>Hobbs (1985)</b>	<b>Current annotation scheme</b>
occasion	temporal sequence
cause	cause-effect: cause stated first, then effect; directionality indicated by directed arcs in a coherence graph
explanation	cause-effect: effect stated first, then cause; directionality indicated by directed arcs in a coherence graph
–	condition
evaluation	elaboration
background	elaboration
exemplification: example stated first, then general case; directionality indicated by directed arcs in a coherence graph	example
exemplification: general case stated first, then example; directionality indicated by directed arcs in a coherence graph	generalization
elaboration	elaboration
parallel	similarity
contrast	contrast
violated expectation	violated expectation
–	attribution
–	same

**Table 2. Correspondence between the set of coherence relations in Hobbs (1985) and our set of coherence relations.**

We distinguish between asymmetrical or directed relations on the one hand and symmetrical or undirected relations on the other hand (Mann & Thompson (1988); Marcu (2000)). *Cause-effect*, *condition*, *violated expectation*, *elaboration*, *example*, *generalization*, and *attribution* are asymmetrical or directed relations, whereas *similarity*, *contrast*, and *same* are symmetrical or undirected relations. In asymmetrical or directed relations, the directions of relations are as follows:

- *cause-effect*: from the discourse segment stating the cause to the discourse segment stating the effect
- *condition*: from the discourse segment stating the condition to the discourse segment stating the consequence
- *violated expectation*: from the discourse segment stating the cause to the discourse segment describing the absent effect
- *elaboration*: from the elaborating discourse segment to the elaborated discourse segment
- *example*: from the discourse segment stating the example to the discourse segment stating the exemplified
- *generalization*: from the discourse segment stating the special case to the discourse segment stating the general case
- *attribution*: from the discourse segment stating the source to the attributed discourse segment
- *temporal sequence*: from the discourse segment stating the event that happened first to the discourse segment stating the event that happened second

#### 2.4 Coding procedure

In order to code the coherence relations of a text, we used a procedure consisting of three steps. In the first step, a text is segmented into discourse segments (cf. Section 2.1).

In the second step, adjacent discourse segments that are topically related are grouped together. The criteria for this step are described in Section 2.2.

In the third step, coherence relations (cf. Section 2.3) are determined between discourse segments and groups of discourse segments. Each previously unconnected (group of) discourse segment(s) is tested to see if it connects to any of the (groups of) discourse segments that have already been connected to the already existing representation of discourse structure.

In order to help determine the coherence relation between (groups of) discourse segments, the (groups of) discourse segments under consideration are connected with one of the contentful conjunctions shown in Table 1 (cf. Hobbs (1985); Kehler (2002)). If

using a contentful conjunction to connect (groups of) discourse segments resulted in an acceptable passage, this was used as evidence that the coherence relation corresponding to the contentful conjunction holds between the (groups of) discourse segments under consideration.

**2.5 Annotators**

The annotators for the database were MIT undergraduate students who worked in our lab as research students. For training, the annotators received a manual that describes the background of the project, discourse segmentation, coherence relations and how to recognize them, and how to use the annotation tools that we developed in our lab (Wolf et al. (2003)). The first author of this paper provided training for the annotators. Training consisted of explaining the background of the project and the annotation method, and of annotating example texts (these texts are not included in our database). Training took about 8-10 hours in total, distributed over 5 days of a week. After that training, annotators worked independently.

**2.6 Statistics on annotated database**

In order to evaluate hypotheses about appropriate data structures for representing coherence structures, we have collected a database of texts where the relations between discourse segments are labeled with the coherence relations described above. Table 3 shows statistics for a database of 135 texts from the Wall Street Journal 1987-1989 and the AP Newswire 1989 (both from Harman & Liberman (1993)) that have been annotated with coherence relations.

number of words		number of discourse segments	
mean	545	mean	61
minimum	161	minimum	6
maximum	1409	maximum	143
median	529	median	60

**Table 3. Database statistics for 135 texts from AP Newswire 1989 (105 texts) and Wall Street Journal 1989 (30 texts).**

Steps Two (discourse segment grouping) and Three (coherence relation annotation) of the coding procedure were performed independently by two annotators. For Step One (discourse segmentation), a pilot study on 10 texts showed that agreement on this step, *number of common segments / (number of common segments + number of differing segments)*, was never below 90%. Therefore, all 135 texts were segmented by two annotators together, resulting in segmentations that both annotators could agree on.

In order to determine inter-annotator agreement for Step Two of the coding procedure for the database of annotated texts, we calculated kappa statistics (Carletta (1996)). We used the following procedure to construct a confusion matrix: first, all groups marked by either annotator were extracted. Annotator 1 had marked 2616 groups, and Annotator 2 had marked 3021 groups in the whole database. The groups marked by the annotators consisted of 536 different discourse segment group types (for example, groupings of discourse segments 1 and 2 were marked 31 times by both annotators; groupings of discourse segments 1, 2, and 3 were marked 6 times by both annotators). Therefore, the confusion matrix had 536 rows and columns. For all annotations of the 135 texts, the agreement was 84.49%, per chance agreement was 1.61%, and kappa was 84.24. Annotator agreement did not differ as a function of text length, arc length, or kind of coherence relation (all  $\chi^2$ s < 1).

We also calculated kappa statistics to determine inter-annotator agreement for Step Three of the coding procedure for the database of annotated texts<sup>1</sup>. For all annotations of the 135 texts, the agreement was 87.61%, per chance agreement was 24.66%, and kappa was 83.55%. Annotator agreement did not differ as a function of text length ( $\chi^2 = 1.27$ ;  $p < 0.75$ ), arc length ( $\chi^2 < 1$ ), or kind of coherence relation ( $\chi^2 < 1$ ). Table 4 shows the confusion matrix for the database of 135 annotated texts that was used to compute the kappa statistics. Table 4 shows, for example, that much of the inter-annotator disagreement seems to be driven by disagreement over how to annotate *elaboration* relations (in the whole database, Annotator<sub>1</sub> marked 260 *elaboration*

---

<sup>1</sup> Notice that inter-annotator agreement for Step Three was influenced by inter-annotator agreement for Step Two. For example, one annotator might mark a group of discourse segments 2 and 3, whereas the second annotator might not mark that group of discourse segments. If the first annotator then marks e.g. a *cause-effect* coherence relation between discourse segment 4 and the group of discourse segments 2 and 3, whereas the second annotator marks a *cause-effect* coherence relation between discourse segment 4 and discourse segment 3, this would count as a disagreement. Thus, our measure of inter-annotator agreement for Step Three is conservative.

relations where Annotator<sub>2</sub> marked no relation; Annotator<sub>2</sub> marked 467 *elaboration* relations where Annotator<sub>1</sub> marked no relation).

Annotator <sub>1</sub>	Annotator <sub>2</sub>												sum	percent
	<i>contr</i>	<i>expv</i>	<i>ce</i>	<i>none</i>	<i>gen</i>	<i>cond</i>	<i>examp</i>	<i>temp</i>	<i>attr</i>	<i>elab</i>	<i>same</i>	<i>sim</i>		
<i>contr</i>	383	11	0	34	0	0	0	2	0	0	0	0	430	4.47
<i>expv</i>	4	113	0	7	0	0	0	0	0	0	0	0	124	1.29
<i>ce</i>	0	0	446	14	0	0	0	0	0	5	0	0	465	4.83
<i>none</i>	66	24	42	0	0	2	27	16	6	467	1	64	715	7.43
<i>gen</i>	0	0	0	1	21	0	0	0	0	1	0	0	23	0.24
<i>cond</i>	0	0	0	2	0	127	0	1	0	1	0	0	131	1.36
<i>examp</i>	0	0	1	18	0	0	219	0	0	3	0	0	241	2.51
<i>temp</i>	1	1	2	7	0	0	0	214	0	1	0	0	226	2.35
<i>attr</i>	0	0	0	5	0	0	0	0	1387	0	0	0	1392	14.47
<i>elab</i>	0	0	17	260	0	3	0	3	0	3913	1	0	4197	43.63
<i>same</i>	0	0	2	5	0	0	0	1	0	0	530	1	539	5.60
<i>sim</i>	7	0	3	43	0	0	0	6	0	0	3	1074	1136	11.81
<b>sum</b>	<b>461</b>	<b>149</b>	<b>513</b>	<b>396</b>	<b>21</b>	<b>132</b>	<b>246</b>	<b>243</b>	<b>1393</b>	<b>4391</b>	<b>535</b>	<b>1139</b>		
<b>percent</b>	<b>4.79</b>	<b>1.55</b>	<b>5.30</b>	<b>4.12</b>	<b>0.20</b>	<b>1.37</b>	<b>2.56</b>	<b>2.53</b>	<b>14.50</b>	<b>45.60</b>	<b>5.56</b>	<b>11.80</b>		

**Table 4.** Confusion matrix of annotations for the database of 135 annotated texts (*contr* = *contrast*; *expv* = *violated expectation*; *ce* = *cause-effect*; *none* = *no coherence relation*; *gen* = *generalization*; *cond* = *condition*; *examp* = *example*; *ts* = *temporal sequence*; *attr* = *attribution*; *elab* = *elaboration*; *sim* = *similarity*).

### 3 Data structures for representing coherence relations

In order to represent the coherence relations between discourse segments in a text, most accounts of discourse coherence assume tree structures (Britton (1994); Carlson et al. (2002); Corston-Oliver (1998); Lascarides & Asher (1993)<sup>2</sup>; Longacre (1983); Grosz & Sidner (1986); Mann & Thompson (1988); Marcu (2000); Polanyi & Scha (1984); Polanyi (1996); Polanyi et al. (2004); van Dijk & Kintsch (1983); Walker (1998)). Other accounts assume less constrained graphs (e.g. Bergler (1992); Birnbaum (1982); Danlos (2004); Hobbs (1985); McKeown (1985); Reichman (1985); Zukerman & McConachy (1995); for dialogue structure: Penstein Rose et al. (1995)).

<sup>2</sup> Although Lascarides & Asher (1993) do not explicitly assume tree structures, they argue that when building a discourse structure, the right frontier of an already existing discourse structure is the only possible attachment point for a new incoming discourse segment (cf. also Polanyi (1996); Polanyi & Scha (1984); Webber et al. (1999)). This constraint on building discourse structures leads to a tree structure.



Some proponents of tree structures assume that trees are easier to formalize and to derive than less constrained graphs (Marcu (2000); Webber et al. (2003)). We demonstrate that in fact many coherence structures in naturally occurring texts cannot be adequately represented by trees. Therefore we argue for less constrained graphs as an appropriate data structure for representing coherence, where nodes represent discourse segments and labeled directed arcs represent the coherence relations that hold between these discourse segments.

Some proponents of more general graphs argue that trees cannot account for a full discourse structure that represents informational, intentional, and attentional discourse relations. For example, Moore & Pollack (1992) point out that rhetorical structure theory (Mann & Thompson (1988)) has both informational and intentional coherence relations, but then forces annotators to decide on only one coherence relation between any two discourse segments. Moore & Pollack (1992) argue that often there is an informational as well as an intentional coherence relation between two discourse segments, which then presents a problem for RST, since only one of the relations can be annotated. Instead, Moore & Pollack (1992) propose allowing more than one coherence relation between two discourse segments, which violates the tree constraint of not having nodes with multiple parents.

Reichman (1985) argues that tree-based story grammars are not enough to account for discourse structure. Instead, she argues that in order to account for intentional structure of discourse, more general data structures are needed. We argue that the same is true for the informational structure of discourse.

Moore & Pollack (1992), Moser & Moore (1996), and Reichman (1985) argue that trees are insufficient for representing informational, intentional, and attentional discourse structure. Notice, however, that the focus of our work is on informational coherence relations, not on intentional relations. That does not mean that we think that attentional or intentional structure should not be part of a full account of discourse structure. Rather, we would like to argue that while the above accounts argue against trees for representing informational, intentional, and attentional discourse structure together, we argue that trees are not even descriptively adequate to describe just informational discourse structure by itself.

Some accounts of informational discourse structure do not assume tree structures, e.g. Bergler (1992) and Hobbs (1985) for monologue and Penstein Rose et al. (1995) for dialogue structure. However, none of these accounts provides systematic empirical support for using more general graphs rather than trees. Providing a systematic empirical study of whether trees are descriptively adequate for representing discourse coherence is the goal of this paper.

There are also accounts of informational discourse structure that argue for trees as a “backbone” for discourse structure but allow certain violations of tree constraints (crossed dependencies or nodes with multiple parents). Examples of such accounts include Webber et al. (1999) and Knott (1996). Similar to our approach, Webber et al. (1999) investigated informational coherence relations. The kinds of coherence relations they use are basically the same that we use (cf. also Hobbs (1985)). However, they argue for a tree structure as a “backbone” for discourse structure, but have also addressed violations of tree structure constraints. In order to accommodate violations of tree structure constraints (in particular crossed dependencies), Webber et al. (1999) argue for a distinction between “structural” discourse relations on the one hand and “non-structural” or “anaphoric” discourse relations on the other hand. “Structural” discourse relations are represented within a Lexicalized Tree-Adjoining Grammar framework, and the resultant “structural” discourse structure is represented by a tree. However, more recently, Webber et al. (2003) have argued that “structural” discourse structure should allow nodes with multiple parents, but no crossed dependencies. It is unclear, however, why Webber et al. (2003) allow one kind of tree constraint violation (nodes with multiple parents) but not another (crossed dependencies).

Notice that there seems to be a problem with the definition of “structural” vs. “non-structural” discourse structure in Webber et al. (1999): according to Webber et al. (1999), non-structural discourse relations are licensed by anaphoric relations and can be involved in crossed dependencies. However, Webber et al. (1999) also argue that one criterion for non-structural coherence relations is that they can cross (non-)structural coherence relations. Since this definition of non-structural appears to be circular, it is necessary to find an independent way to validate the difference between structural and non-structural coherence relations. Knott (1996) might provide a way to empirically

formalize the claims in Webber et al. (1999), or at least claims that seem to be very similar to Webber et al. (1999): Knott (1996) argues that *elaboration* relations are more permissive than other types of coherence relations (e.g. *cause-effect*, *parallel*, *contrast*). As a consequence, Knott (1996) argues, *elaboration* relations would better be described in terms of, according to him, less constrained focus structures (cf. Grosz & Sidner (1986)) than in terms of, also according to Knott (1996), more constrained rhetorical relations (cf. Hobbs (1985); Mann & Thompson (1988)). This hypothesis makes testable empirical claims: *elaboration* relations should in some way pattern differently from other coherence relations. We will come back to this issue in Sections 4.1 and 4.2.

In this paper we present evidence against trees as a data structure for representing discourse coherence. Notice though that the evidence does not support the claim that discourse structures are completely arbitrary. The goal of our research program is to first determine which constraints on discourse structure are empirically viable. To us, the work we present here seems to be the crucial first step in order to avoid arbitrary constraints on inferences for building discourse structures. In other words, the point we wish to make here is that while there might be other constraints on possible discourse annotations, tree structure constraints do not seem to be the right kinds of constraints. This appears to be a crucial difference between approaches like Marcu (2000)'s, Knott (1996)'s, or Webber et al. (2003)'s on the one hand, and our approach on the other hand. The goal of the former approaches seems to be to first specify a set of constraints on possible discourse annotations, and then to annotate texts with these constraints in mind.

The following two sections will illustrate problems with trees as a representation of discourse coherence structures. Section 3.1 will show that the discourse structures of naturally occurring texts contain crossed dependencies, which cannot be represented in trees. Another problem for trees, in addition to crossed dependencies, is that many nodes in coherence graphs of naturally occurring texts have multiple parents. This is shown in Section 3.2.

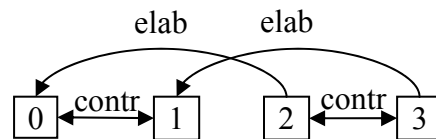
### 3.1 Crossed dependencies

Consider the text passage in (18) (modified from SAT practicing materials; not annotated by Carlson et al. (2002), so we will not present an RST-tree-based annotation). Figure 1

shows the coherence graph for (18). Notice that the arrowheads of the arcs represent directionality for asymmetrical relations (*elaboration*) and bidirectionality for symmetrical relations (*similarity, contrast*).

(18) Example text (modified from SAT practicing materials)

0. Schools tried to teach students history of science.
1. At the same time they tried to teach them how to think logically and inductively.
2. Some success has been reached in the first of these aims.
3. However, none at all has been reached in the second.



**Figure 1. Coherence graph for (18). Abbreviations used: *contr* = *contrast*; *elab* = *elaboration*.**

The coherence structure for (18) can be derived as follows: there is a *contrast* relation between discourse segments 0 and 1; discourse segments 0 and 1 describe teaching different things to students. There is another *contrast* relation between discourse segments 2 and 3; discourse segments 2 and 3 describe varying degrees of success (some vs. none). Discourse segment 2 provides more details (the degree of success) about the teaching described in discourse segment 0, so there is an *elaboration* relation between discourse segments 2 and 0. Furthermore, in another *elaboration* relation, discourse segment 3 provides more details (the degree of success) about the teaching described in discourse segment 1. In the resultant coherence structure for (18), there is a crossed dependency between  $\{2, 0\}$  and  $\{3, 1\}$ .

The coherence structure of (18) contains crossed dependencies. In order to be able to represent such a structure in a tree without violating validity assumptions about tree structures (Diestel (2000)), one might consider augmenting a tree either with feature propagation (Shieber (1986)) or with a coindexation mechanism (Chomsky (1973)).

There is a problem for both feature propagation and coindexation mechanisms: Both the tree structure itself as well as the features and coindexations represent the same kind of information (coherence relations). It is unclear how a dividing line could be drawn between tree structures and their augmentation. That is, it is unclear how one could decide which part of a text coherence structure should be represented by the tree structure and which part should be represented by the augmentation. Other areas of linguistics have faced this issue as well. Researchers investigating data structures for representing intra-sentential structure, for instance, generally fall into two groups. One group tries to formulate principles that allow representing some aspects of structure in the tree itself and other aspects in some augmentation formalism (e.g. Chomsky (1973); Marcus et al. (1994)). Another group argues that it is more parsimonious to assume a unified dependency-based representation that drops the tree constraints of allowing no crossed dependencies (e.g. Brants et al. (2002); Skut et al. (1997); König & Lezius (2000)). Our approach falls into the latter group. As we will point out, there does not seem to be a well-defined set of constraints on crossed dependencies in discourse structures. Without such constraints, it does not seem viable to represent discourse structures as augmented tree structures.

An important question is how many different kinds of crossed dependencies occur in naturally occurring discourse. If there are only a very limited number of different structures with crossed dependencies in natural texts, one could make special provisions to account for these structures and otherwise assume tree structures. (18), for instance, has a list-like structure. It is possible that list-like examples are exceptional in natural texts. However, there are many other naturally occurring non-list-like structures that contain crossed dependencies. As an example of a non-list-like structure with a crossed dependency (between {3, 1} and {2, 0-1}), consider the constructed example (19):

(19) Example text (constructed)

0. Susan wanted to buy some tomatoes
1. and she also tried to find some basil
2. because her recipe asked for these ingredients.
3. The basil would probably be quite expensive at this time of the year.

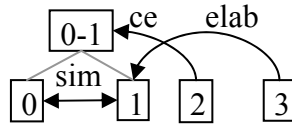


Figure 2. Coherence graph for (19). Abbreviations used: *sim* = *similarity*; *ce* = *cause-effect*; *elab* = *elaboration*.

The coherence structure for (19) can be derived as follows: there is a *parallel* relation between 0 and 1; 0 and 1 both describe shopping for grocery items. There is a *cause-effect* relation between 2 and 0-1; 2 describes the cause for the shopping described by 0 and 1. Furthermore, there is an *elaboration* relation between 3 and 1; 3 provides details about the basil in 1. Figure 2 shows the coherence graph for (19).

(20) from the AP Newswire 1989 corpus is an example with a similar structure ((20) was not annotated by Carlson et al. (2002), so we will not present an RST-tree-based annotation):

(20) Example text (from ap890109-0012; AP Newswire 1989 corpus; Harman & Liberman (1993))

0. The flight Sunday took off from Heathrow Airport at 7:52pm
1. and its engine caught fire 10 minutes later,
2. the Department of Transport said.
3. The pilot told the control tower he had the engine fire under control.

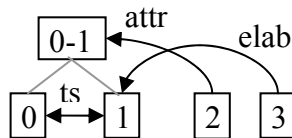


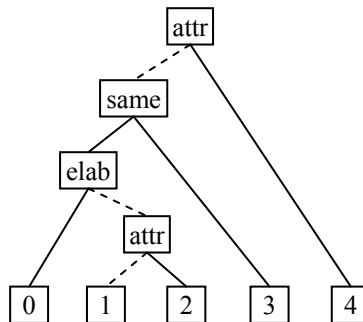
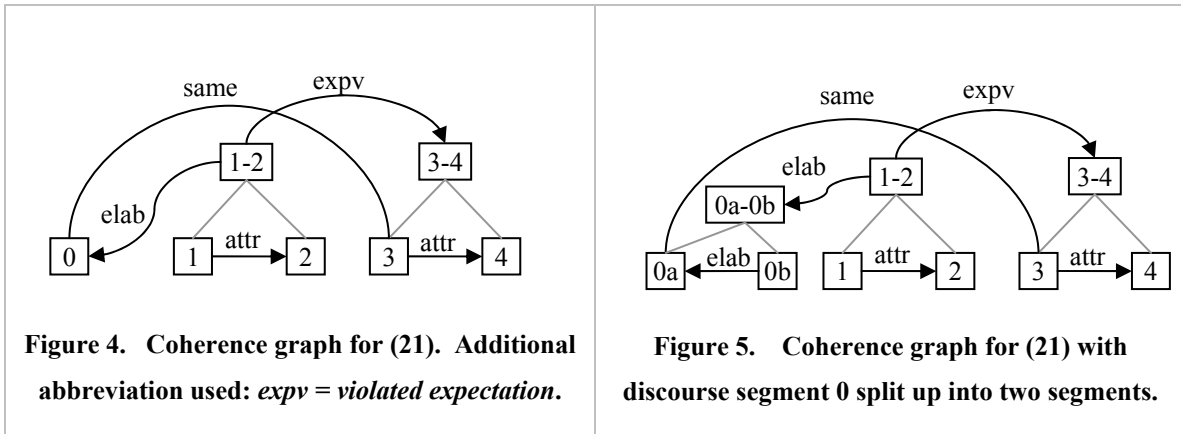
Figure 3. Coherence graph for (20). Abbreviations used: *ts* = *temporal sequence*; *attr* = *attribution*; *elab* = *elaboration*.

The coherence structure for (20) can be derived as follows: 1 and 0 are in a *temporal sequence* relation; 0 describes the takeoff that happens before the engine fire described by

1 occurs. 2 and 0-1 are in an *attribution* relation; 2 mentions the source of what is said in 0-1. 3 and 1 are in an elaboration relation; 3 provides more detail about the engine fire in 1. The resulting coherence structure, shown in Figure 3, contains a crossed dependency between {3, 1} and {2, 0-1}.

(21) Example text (from wsj\_0655; Wall Street Journal 1989 corpus; Harman & Liberman (1993))

- 0. [ Mr. Baker's assistant for inter-American affairs, ]<sub>0a</sub> [ Bernard Aronson, ]<sub>0b</sub>
- 1. while maintaining
- 2. that the Sandinistas had also broken the cease-fire,
- 3. acknowledged:
- 4. "It's never very clear who starts what."



**Figure 6. Tree-based RST-annotation for (21) from Carlson et al. (2002). Abbreviations used: attr = attribution; elab = elaboration; attr = same. Dashed lines represent the start of asymmetric coherence relations; continuous lines represent the end of asymmetric coherence relations; symmetric coherence relations have two continuous lines (cf. Section 2.3).**

Consider (21) from the Wall Street Journal 1989 corpus (Harman & Liberman (1993)). For (21) we provide annotations based on our annotation scheme, as well as the annotation that Carlson et al. (2002) provide for the same passage. The annotations based on our annotation scheme are presented with the discourse segmentation based on the segmentation guidelines in Carlson et al. (2002) (Figure 4), and based on our segmentation guidelines from Section 2.1 (Figure 5). Figure 6 shows a tree-based RST annotation for (21) from Carlson et al. (2002). The only difference between Carlson et al. (2002) and our approach with respect to how (21) is segmented is that Carlson et al. (2002) assume discourse segment 0 to be one single segment. By contrast, based on our segmentation guidelines, discourse segment 0 would be segmented into two segments (because of the comma that does not separate a complex NP or VP), 0a and 0b, as indicated by the angle brackets below<sup>3</sup>:

(22) [ Mr. Baker's assistant for inter-American affairs, ]<sub>0a</sub> [ Bernard Aronson, ]<sub>0b</sub>

We then derived the coherence structure for (21) as follows: if, following our discourse segmentation guidelines, discourse segment 0 is segmented into 0a and 0b, 0a and 0b are in an *elaboration* relation: 0b provides additional detail (a name) about what is stated in 0a (Mr. Baker's assistant). Furthermore, 0 (or 0a) and 3 are in a *same* relation: the subject NP in 0 ("Mr. Baker's assistant...") is separated from its predicate in 3 ("acknowledged") by intervening discourse segments 1 and 2 (and 0b in our discourse segmentation). 1 and 2 are in an *attribution* relation: 1 states the source of what is stated in 2 (the source in 1 is the elided "Mr. Baker"). The group of discourse segments 1 and 2 is in an *elaboration* relation with discourse segment 0 (or the group of discourse segments 0a and 0b in our discourse segmentation): 1 and 2 state additional detail (a statement about a political process) about what is stated in 0, or 0a and 0b (Mr. Baker's assistant). 3 (and by virtue of the *same* relation also 0 or 0a) and 4 are in an *attribution*

---

<sup>3</sup>Based on our segmentation guidelines, the complementizer "that" in discourse segment 2 would be part of discourse segment 1 instead (cf. (15) in Section 2.3). However, since that would not make a difference in terms of the resulting discourse structure, we do not provide alternative analyses with "that" part of discourse segment 1 instead of discourse segment 2.



relation: 3 states the source (Mr. Baker's assistant) of what is stated in 4. Furthermore, there is a *violated expectation* relation between the group of discourse segments 1 and 2 and the group of discourse segments 3 and 4: although Mr. Baker's assistant acknowledged cease fire violations by one side (discourse segments 1 and 2), he acknowledges that it is in fact difficult to clearly blame one side for cease fire violations (discourse segments 3 and 4).

The resulting coherence structure, shown in Figure 5 (discourse segmentation from Carlson et al. (2002)) and Figure 6 (our discourse segmentation), contains a crossed dependency: the *same* relation between discourse segment 0 and discourse segment 3 crosses the *violated expectation* relation between the group of discourse segments 1 and 2 and the group of discourse segments 3 and 4.

Figure 6 represents a tree-based RST annotation for (21) from Carlson et al. (2002); in Figure 6, dashed lines represent the start of asymmetric coherence relations, and continuous lines mean the end of asymmetric coherence relations; symmetric coherence relations have two continuous lines (cf. Section 2.3 for the distinction between symmetric and asymmetric coherence relations, and for the directions of asymmetric coherence relations). We do not have a description of how Carlson et al. (2002) derived a tree-based RST structure for (21); however, it seems that the tree structure shown in Figure 6 could be derived as follows: 1 states the source of what is stated in 2, so 1 and 2 are in an *attribution* relation. 1 and 2 state additional detail (a statement about a political process) about what is stated in 0 (Mr. Baker's assistant), so the group of 1 and 2 *elaborate* on 0. Then, according to Carlson et al. (2002), the group of discourse segments 0 to 2 are in a *same* relation with 3. Notice that this is different from our annotation, where a *same* relation holds only between 0 and 3. Finally, according to Carlson et al. (2002), the group of discourse segments 0 to 3 are in an *attribution* relation with 4: 0 to 3 state the source of what is stated in 4. Notice that in our annotation, the *attribution* relation holds only between 3 and 4. Notice furthermore that the *violated expectation* relation between the group of discourse segments 1 and 2 and the group of discourse segments 3 and 4 is absent from the structure shown in Figure 6. It is not clear how that relation could be annotated without violating the tree constraint of not allowing crossed dependencies.

### 3.2 Nodes with multiple parents

In addition to including crossed dependencies, many coherence structures of natural texts include nodes with multiple parents. Such nodes cannot be represented in tree structures. Consider (23) from the AP Newswire 1989 (Harman & Liberman (1993); this text was not annotated by Carlson et al. (2002), so we will not present an RST-tree-based annotation). The coherence structure for (23) can be derived as follows: 1 states the source of what is stated in 0 and in 3, so there are *attribution* relations between 1 and 0 and 1 and 3 respectively. 2 and 1 are in an *elaboration* relation; 2 provides additional detail (the name) about the BMW driver in 1. 3 and 0 are in a *condition* relation; 3 states the BMW driver’s condition for being polite, stated in 0<sup>4</sup>; the *condition* relation is also indicated by the phrase “as long as”. In the resultant coherence structure, node 1 has two parents – one *attribution* and one *condition* ingoing arc (cf. Figure 7).

(23) Example text (from ap890103-0014; AP Newswire 1989 corpus; Harman & Liberman (1993))

0. “Sure I’ll be polite,”
1. promised one BMW driver
2. who gave his name only as Rudolf.
3. “As long as the trucks and the timid stay out of the left lane.”

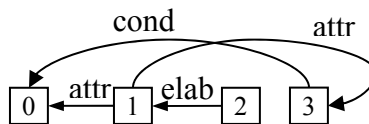


Figure 7. Coherence graph for (23). Additional abbreviation used: *cond* = *condition*.

As another example of a discourse structure that contains nodes with multiple parents, consider the structure of (24) from the AP Newswire 1989 corpus (Harman & Liberman (1993)). Since Carlson et al. (2002) also annotated that text, we also present their annotation and contrast it with ours. Our annotations are shown in Figure 8 (discourse

<sup>4</sup> A cultural reference: In Germany it is only lawful to pass on the left side when driving on a highway. Thus, Rudolf is essentially saying that he will be polite as long as “the trucks and the timid” do not keep him from passing other cars.

segmentation from Carlson et al. (2002)) and Figure 9 (our discourse segmentation). The only difference between our annotation and Carlson et al. (2002)'s is that we do not assume two separate discourse segments for 0 and 1; 0 and 1 are one discourse segment in our annotation (represented by the node "0+1" in Figure 9). Notice also that in (21), discourse segment 2, "that" is not in a separate discourse segment; it is unclear why in (24) "that" is in a separate discourse segment (discourse segment 1) and not part of discourse segment 2<sup>5</sup>.

The discourse structure for (24) can be derived as follows: According to our discourse segmentation guidelines (cf. Section 2.1), 0 and 1 should be one single discourse segment; this could be annotated by a *same* relation between 0 and 1 (cf. Figure 8) or by merging discourse segments 0 and 1 into one single discourse segment (cf. Figure 9). Discourse segment 0 or 0+1 state the source (the administration) of what is stated in discourse segments 2 and 3, so there is an *attribution* relation between 0 or 0+1 and the group of discourse segments 2-3. Furthermore, there is a *condition* relation between 2 and 3; 2 states the condition for what is stated in 3 (the *condition* relation is also signaled by the cue phrase "if" in 2). There is also an *attribution* relation between 4 and the group of discourse segments 0-3; 4 states the source of what is stated in 0-3. 4 also states the source of what is stated in 5, so there is another *attribution* relation between 4 and 5. Finally, there is an *evaluation-s* relation between 5 and the group of discourse segments 2-3; 2-3 state what is evaluated by 5 – the Contra supporters should call for military aid, and if the February election is voided (group of discourse segments 2-3), the Contra supporters might win (discourse segment 5). Notice that in our annotation scheme, the *evaluation-s* relation would be a *condition* relation (the election being voided and the Contras calling for military aid would be the condition for them to win). In the resultant coherence structure for (24), node 2-3 has multiple parents or ingoing arcs – one *attribution* ingoing arc and one *evaluation-s* ingoing arc (cf. Figure 8 and Figure 9).

---

<sup>5</sup>Based on our segmentation guidelines, the complementizer "that" in discourse segment 2 would be part of discourse segment 1 instead (cf. (15) in Section 2.3). However, since that would not make a difference in terms of the resulting discourse structure, we do not provide alternative analyses with "that" part of discourse segment 1 instead of discourse segment 2.

(24) Example text (from wsj\_0655; Wall Street Journal 1989 corpus; Harman & Liberman (1993))

0. "The administration should now state
1. that
2. if the February election is voided by the Sandinistas
3. they should call for military aid,"
4. said former Assistant Secretary of State Elliott Abrams.
5. "In these circumstances, I think they'd win."

[ "they" in 3 and 5 = "Contra supporters"; this is clear from the whole text wsj\_0655 ]

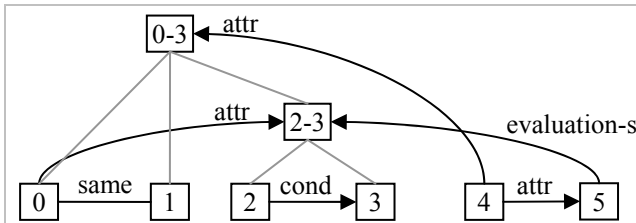


Figure 8. Coherence graph for (24). Additional coherence relation used (from Carlson et al. (2002)): *evaluation-s* = the situation presented in the satellite assesses the situation presented in the nucleus.

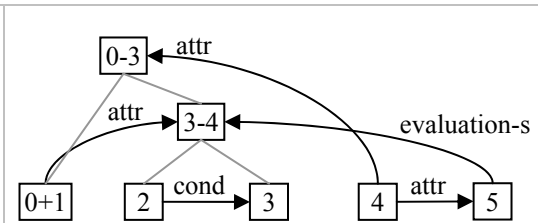


Figure 9. Coherence graph for (24) with discourse segments 0 and 1 merged into one single discourse segment.

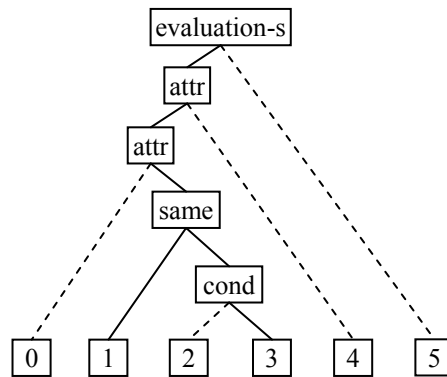


Figure 10. Tree-based RST annotation for (24) from Carlson et al. (2002). Dashed lines represent the start of asymmetric coherence relations; continuous lines represent the end of asymmetric coherence relations; symmetric coherence relations have two continuous lines (cf. Section 2.3).

As for (21), we do not have a description available to us of how Carlson et al. (2002) derived their tree-based RST annotation (Figure 10). However, the following might be a possible post-hoc explanation of how the tree-based RST annotation in Figure 10 was derived: 2 and 3 are in a *condition* relation (cf. the description of our annotation above). In *attribution* relations, Carlson et al. (2002) group the complementizer “that” with the attributed statement rather than with the discourse segment that states the source. Therefore there is a *same* relation between 1 and discourse segment 2. However, in order to keep a tree structure, the *same* relation has to be between 1 and the group of discourse segments 2-3, rather than just between 1 and 2 (with the *same* relation just between 1 and 2, 2 would have multiple parents: one *same* relation, and one *condition* relation). Furthermore, there is an *attribution* relation between 0 and the group of discourse segments 1-3 (cf. the description of our annotation above; the reason that the grouping of discourse segments is 1-3 and not 2-3 as in our annotation is that Carlson et al. (2002) group “that” in 1 with the attributed statements, and not with the source statements as in our annotation, cf. Section 2.1). There is another *attribution* relation between 4 and the group of discourse segments 0-3 (cf. the description of our annotation above). Finally, there is an *evaluation-s* relation between 5 and the group of discourse segments 0-4. Notice, however, that the *evaluation-s* relation seems to hold rather between 5 and the group of discourse segments 2-3; what is being evaluated is a chance for the Contras to win a military conflict under certain circumstances. However, annotating a coherence relation between 5 and the group of discourse segments 2-3 could not be accommodated in a tree structure. Notice also that Carlson et al. (2002)’s tree-based RST structure for (24) does not represent the *attribution* relation between 4 and 5. Figure 10 does not represent the fact that what is stated in 5 can be attributed to a source stated in 4, just like what is stated in 0-3 can be attributed to a source stated in 4. It is unclear why the *attribution* relation between 0-3 should be represented, but not the attribution relation between 4 and 5.

#### 4 Statistics

We performed a number of statistical analyses on our annotated database to test our hypotheses. Each set of statistics was calculated for both annotators separately.

However, since the statistics for both annotators were never different from each other (as confirmed by significant  $R^2$ s  $> 0.9$  or by  $\chi^2$ s  $< 1$ ), we only report the statistics for one annotator in the following sections.

An important question is how frequent the phenomena discussed in the previous sections are. The more frequent they are, the more urgent the need for a data structure that can adequately represent them. The following sections report statistical results on crossed dependencies (Section 4.1) and nodes with multiple parents (Section 4.2).

#### 4.1 Crossed dependencies

The following sections report counts on crossed dependencies in the annotated database of 135 texts (cf. Section 1). Section 4.1.1 reports results on the frequency of crossed dependencies, Section 4.1.2 reports results concerning the question of what types of coherence relations tend to be involved in crossed dependencies, and Section 4.1.3 reports results on the arc lengths of coherence relations involved in crossed dependencies. Section 4.1.4 provides a short summary of the statistical results on crossed dependencies.

##### 4.1.1 Frequency of crossed dependencies

In order to track the frequency of crossed dependencies for the coherence structure graph of each text, we counted the minimum number of arcs that would have to be deleted in order to make the coherence structure graph free of crossed dependencies. The example graph in Figure 11 illustrates this process. This graph contains the following crossed dependencies:  $\{0, 1\}$  crosses with  $\{1, 3\}$ ,  $\{2, 4\}$  with  $\{1, 3\}$ , and  $\{4, 6\}$  with  $\{5, 7\}$ . By deleting  $\{1, 3\}$ , two crossed dependencies can be eliminated: the crossing of  $\{0, 1\}$  and  $\{1, 3\}$ , and the crossing of  $\{2, 4\}$  with  $\{1, 3\}$ . By deleting either  $\{4, 6\}$  or  $\{5, 7\}$  the remaining crossed dependency between  $\{4, 6\}$  and  $\{5, 7\}$  can be eliminated. Therefore two edges would have to be deleted from the graph in Figure 11 in order to make it free of crossed dependencies.

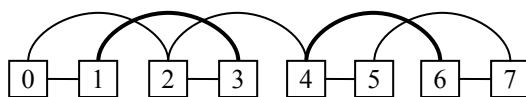


Figure 11. Example graph with crossed dependencies.

Table 5 shows the results of the counts. On average for the 135 annotated texts, 12.5% of arcs in a coherence graph have to be deleted in order to make the graph free of crossed dependencies. Seven texts out of 135 had no crossed dependencies. The mean number of arcs for the coherence graphs of these texts was 36.9 (minimum: 8; maximum: 69; median: 35). The mean number of arcs for the other 128 coherence graphs (those with crossed dependencies) was 125.7 (minimum: 20; maximum: 293; median: 115.5). Thus, the graphs with no crossed dependencies have significantly fewer arcs than those graphs that have crossed dependencies ( $\chi^2(1) = 15330.35$  (Yates' correction for continuity applied);  $p < 10^{-6}$ ). This is a likely explanation for why these seven texts had no crossed dependencies.

mean	12.5
minimum	0
maximum	44.4
median	10.9

**Table 5. Percentages of arcs to be deleted in order to eliminate crossed dependencies.**

More generally, linear regressions show a correlation between the number of arcs in a coherence graph and the number of crossed dependencies. The more arcs a graph has, the higher the number of crossed dependencies ( $R^2 = 0.39$ ;  $p < 10^{-4}$ ; cf. Figure 12). The same linear correlation holds between text length and number of crossed dependencies – the longer a text, the more crossed dependencies are in its coherence structure graph (for text length in discourse segments:  $R^2 = .29$ ,  $p < 10^{-4}$ ; for text length in words:  $R^2 = .24$ ,  $p < 10^{-4}$ ).

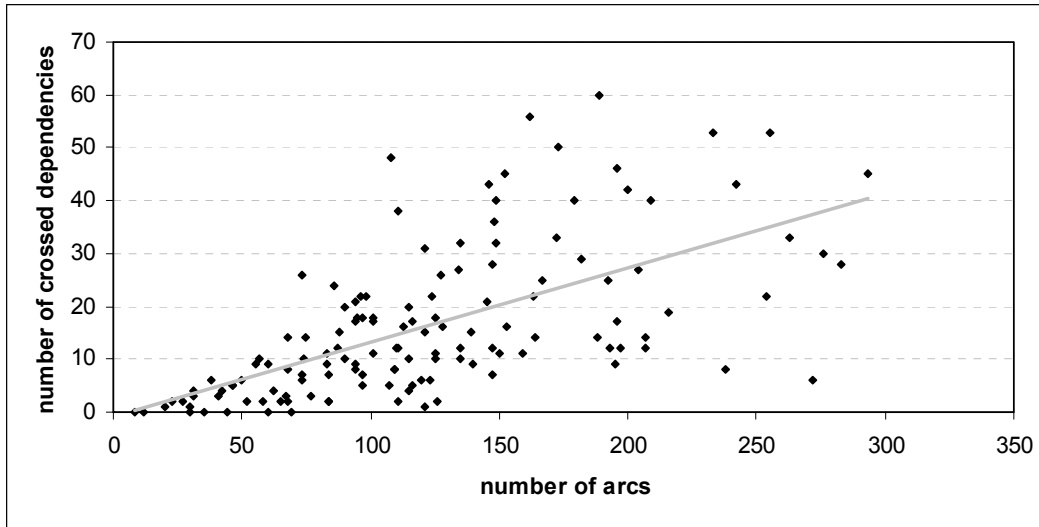


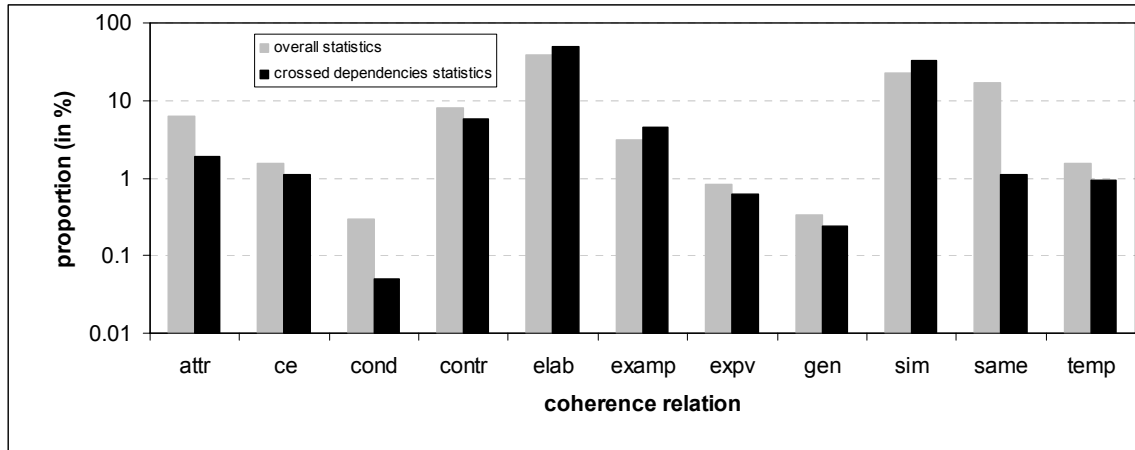
Figure 12. Correlation between number of arcs and number of crossed dependencies.

#### 4.1.2 Types of coherence relations involved in crossed dependencies

In addition to the question of how frequent crossed dependencies are, another question is whether there are certain types of coherence relations that participate more or less frequently in crossed dependencies than other types of coherence relations. For an arc to participate in a crossed dependency, it means that the arc is in the set of arcs that would have to be deleted from a coherence graph in order to make that graph free of crossed dependencies (cf. the procedure outlined in the beginning of Section 4.1). In other words, the question is whether the frequency distribution over types of coherence relations is different for arcs participating in crossed dependencies compared to the overall frequency distribution over types of coherence relations in the whole database.

Figure 13 shows that the overall distribution over types of coherence relations participating in crossed dependencies is not different from the distribution over types of coherence relations overall. This is confirmed by a linear regression, which shows a significant correlation between the two distributions of percentages ( $R^2 = 0.84$ ;  $p < .0001$ ). Notice that the overall distribution includes only arcs with length greater than one, since arcs of length one could not participate in crossed dependencies.





**Figure 13. Distributions over types of coherence relations.** For each condition (“overall statistics” and “crossed dependencies statistics”), the sum over all coherence relations is 100; each bar in each condition represents a fraction of the total of 100 in that condition. The y-axis uses a  $\log_{10}$  scale.

However, there are some differences for individual coherence relations. Some types of coherence relations occur considerably less frequently in crossed dependencies than overall in the database. Table 6 shows the data from Figure 13, ranked by the factor of “proportion of overall coherence relations” by “proportion of coherence relations participating in crossed dependencies”. The proportion of *same* relations, for instance, is 15.23 times greater, and the percentage of *condition* relations is 5.59 times greater overall in the database than in crossed dependencies. We do not yet understand the reason for these differences, and plan to address this question in future research.

## REPRESENTING DISCOURSE COHERENCE

Coherence relation	Proportion of coherence relations participating in crossed dependencies (in %)	Proportion of overall coherence relations (in %)	Factor (= overall / crossed dependencies)
<i>same</i>	1.13	17.21	15.23
<i>condition</i>	0.05	0.28	5.59
<i>attribution</i>	1.93	6.31	3.27
<i>temporal sequence</i>	0.94	1.56	1.66
<i>generalization</i>	0.24	0.34	1.40
<i>contrast</i>	5.84	7.93	1.36
<i>cause-effect</i>	1.13	1.53	1.35
<i>violated expectation</i>	0.61	0.82	1.40
<i>elaboration</i>	50.52	37.97	0.71
<i>example</i>	4.43	3.15	1.34
<i>similarity</i>	33.18	22.91	0.69

**Table 6. Proportions of coherence relations.**

Another way of testing whether certain coherence relations contribute more than others to crossed dependencies is to remove coherence relations of a certain type from the database and then count the remaining number of crossed dependencies. For example, it is possible that the number of crossed dependencies is reduced once all *elaboration* relations are removed from the database. Table 7 shows that by removing all *elaboration* relations from the database of 135 annotated texts, the percentage of coherence relations involved in crossed dependencies is reduced from 12.5% to 4.96% of the remaining coherence relations. That percentage is reduced even further, to 0.84%, by removing all *elaboration* and *similarity* relations from the database. These numbers seem to be partial support for Knott (1996)'s hypothesis: Knott (1996) argued that *elaboration* relations are less constrained than other types of coherence relations (cf. the discussion of Knott (1996) in Section 3).

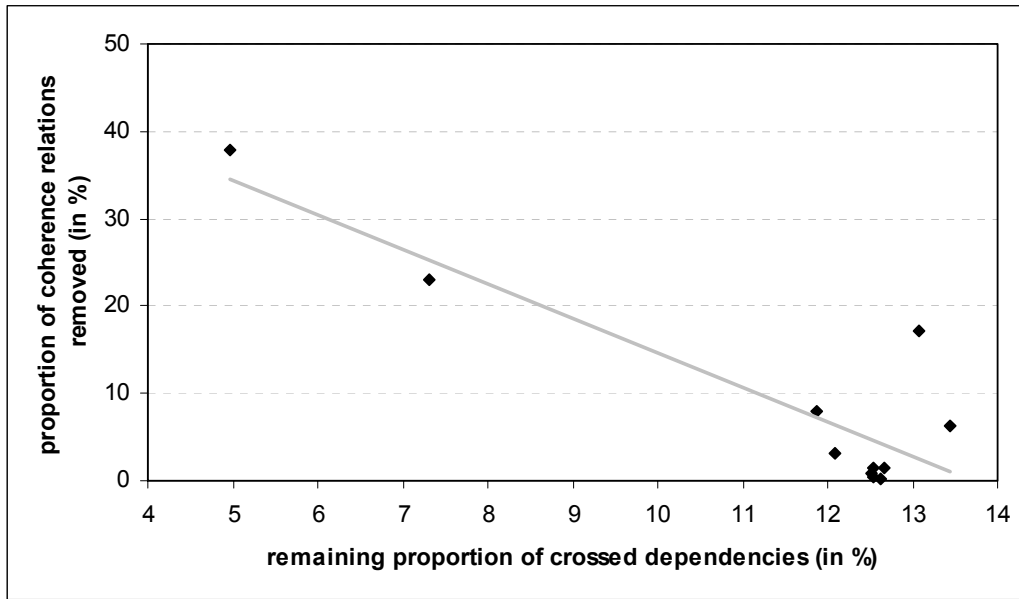
Coherence relation removed	Remaining percentage of coherence relations involved in crossed dependencies			
	mean	min	max	median
<i>Same</i>	13.08	0	44.44	11.39
<i>Condition</i>	12.63	0	45.28	10.89
<i>Attribution</i>	13.44	0	44.86	11.36
<i>temporal sequence</i>	12.53	0	44.44	10.87
<i>Generalization</i>	12.53	0	44.44	10.84
<i>Contrast</i>	11.88	0	46.15	9.86
<i>cause-effect</i>	12.67	0	49.47	11.03
<i>violated expectation</i>	12.51	0	44.44	10.87
<i>Elaboration</i>	4.96	0	47.47	1.23
<i>Example</i>	12.08	0	44.44	9.89
<i>Similarity</i>	7.32	0	24.56	7.04
<i>elaboration and similarity</i>	0.84	0	10.68	0.00

**Table 7. The effect of removing different types of coherence relations on the percentage of coherence relations involved in crossed dependencies.**

However, there is a possible alternative hypothesis to Knott (1996). In particular, *elaboration* relations are very frequent (37.97% of all coherence relations, cf. Table 6). It is possible that removing *elaboration* relations from the database only reduces the number of crossed dependencies because a large number of coherence relations, 37.97%, are removed when *elaborations* are removed. In other words, an alternative hypothesis to Knott (1996) is that the lower number of coherence relations is just due to less dense coherence graphs (i.e. the less dense coherence graphs are, the lower the chance for crossed dependencies). We tested this hypothesis by correlating the proportion of coherence relations removed with the proportion of crossed dependencies that remain after removing a certain type of coherence relation<sup>6</sup>. Figure 14 shows that the higher the proportion of removed coherence relations, the lower the proportion of coherence relations becomes that are involved in crossed dependencies. This correlation is confirmed by a linear regression ( $R^2 = 0.7697$ ;  $p < .0005$ ; after removing the *elaboration*

<sup>6</sup> Notice that the proportions of removed coherence relations do not include coherence relations of absolute arc length 1, since removing those coherence relations cannot have any influence on the number of crossed dependencies (coherence relations of absolute arc length 1 cannot be involved in crossed dependencies). Thus, the proportions of coherence relations removed in Figure 14 are from the third column of Table 6.

data point:  $R^2 = 0.4504$ ;  $p < .05$ ; these linear regressions do not include the data point *elaboration + similarity*). Thus, while removing certain types of coherence relations reduces the number of crossed dependencies, it results in a very impoverished representation of coherence structure (i.e. after removing all *elaboration* and all *similarity* relations, only 39.12% of all coherence relations would still be represented, cf. Table 6; 52.13% based on the distribution over coherence relations including those with absolute arc length 1, cf. Table 9).



**Figure 14. Correlation between removed proportion of overall coherence relations and remaining proportion of crossed dependencies. Notice that the data point for *elaboration + similarity* is not included in the graph above. Both axes represent percent values.  $R^2 = 0.7699$ ;  $p < .0005$ .**

Our present data do not distinguish between Knott (1996)’s hypothesis, i.e. that *elaboration* relations are less constrained, and a simple alternative hypothesis, i.e. that removing *elaboration* relations reduces the number of crossed dependencies just because removing *elaborations* removes a large number of coherence relations, leading to a much less dense graph with fewer chances of crossed dependencies. However, with respect to Knott (1996)’s hypothesis, notice that leaving out *elaboration* relations still leaves the proportion of remaining crossed dependencies at 4.96% (cf. Table 7). In order to further reduce the proportion of remaining crossed dependencies, removing *similarity* relations

in addition to removing *elaboration* relations has the greatest effect (cf. Table 7). But it is unclear a priori why *elaboration* and *similarity* relations should be left out but not, for example, *contrast* relations. It seems that some of the underlying mechanisms of establishing *similarity*, *contrast*, and *elaboration* relations are related: according to Kehler (2002), the first step in establishing either a *similarity* or a *contrast* coherence relation is to find similar or contrasting sets of entities in the two discourse segments under consideration; in order to establish an *elaboration* relation, two discourse segments have to describe a set of entities whose members are centered around a common event or entity (Kehler (2002)). Thus, it seems that anaphoric processes (finding similar, contrasting, or common entities) are important to establishing *similarity*, *contrast*, and *elaboration* relations. However, as Table 7 shows, removing contrast relations only has a negligible effect on the proportion of remaining crossed dependencies. At the same time, *contrast* relations are much less frequent than both *elaboration* and *similarity* relations: if the percentages in the third column of Table 7 are converted back into raw counts, the numbers are 4133 for *elaboration*, 2254 for *similarity*, and 826 for *contrast*;  $\chi^2$ -tests with Yates' correction for continuity applied show that *elaboration* is more frequent than *similarity* ( $\chi^2(1) = 552.20$ ;  $p < 10^{-6}$ ), and that *contrast* is less frequent than *elaboration* ( $\chi^2(1) = 2204.00$ ;  $p < 10^{-6}$ ) and *similarity* ( $\chi^2(1) = 661.15$ ;  $p < 10^{-6}$ ).

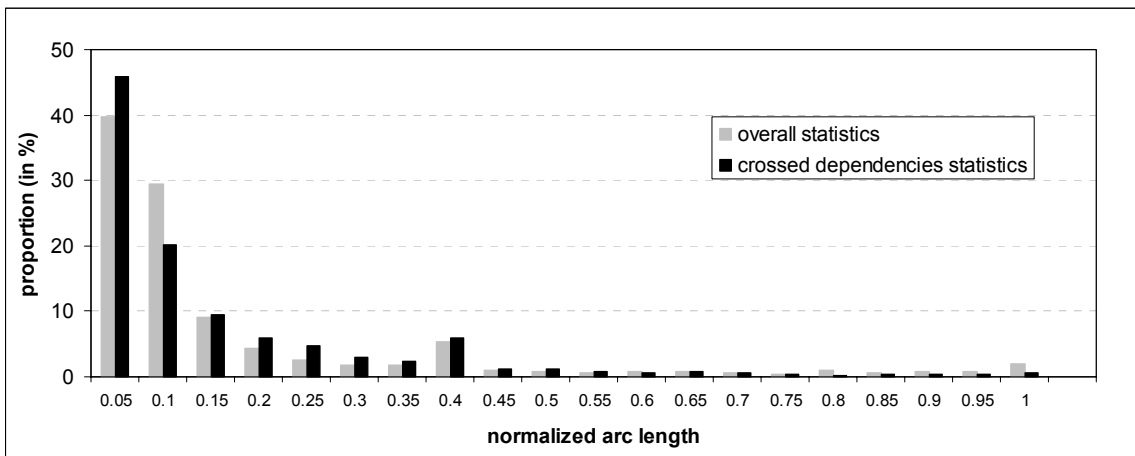
#### 4.1.3 Arc length of coherence relations involved in crossed dependencies

Another question is how great the distance typically is between discourse segments that participate in crossed dependencies, or how great the arc length is for coherence relations that participate in crossed dependencies<sup>7</sup>. It is possible, for instance, that crossed dependencies primarily involve long-distance arcs and that more local crossed dependencies are disfavored. However, Figure 15 shows that the distribution over arc lengths is practically identical for the overall database and for coherence relations participating in crossed dependencies (linear regression:  $R^2 = 0.937$ ,  $p < 10^{-4}$ ), suggesting a strong locality bias for coherence relations overall as well as for those participating in

---

<sup>7</sup> The distance between two discourse segments is measured not in terms of how many coherence links one has to follow from any discourse segment  $x$  to any discourse segment  $y$  to which discourse segment  $x$  is related via a coherence relation. Instead, distance is measured in terms of the number of intervening discourse segments. Thus, distance between nodes reflects linear distance between two discourse segments in a text. For example, the distance between a discourse segment 1 and a discourse segment 4 would be 3.

crossed dependencies<sup>8</sup>. The arc lengths are normalized in order to take into account the length of a text. Normalized arc length is calculated by dividing the absolute length of an arc by the maximum length that that arc could have, given its position in a text. For example, if there is a coherence relation between discourse segment 0 and discourse segment 3 in a text, the raw distance would be 3. If these discourse segments are part of a text that has 5 discourse segments total (i.e. 0 to 4), the normalized distance would be  $3 / 4 = 0.75$  (because 4 would be the maximum possible length of an arc that originates in discourse segment 0 or 3, given that the text has 5 discourse segments in total).



**Figure 15. Comparison of normalized arc length distributions. For each condition (“overall statistics” and “crossed dependencies statistics”), the sum over all coherence relations is 100; each bar in each condition represents a fraction of the total of 100 in that condition.**

#### 4.1.4 Summary on crossed dependencies statistics

Taken together, statistical results on crossed dependencies suggest that crossed dependencies are too frequent to be ignored by accounts of coherence. Furthermore, the results suggest that any type of coherence relation can participate in a crossed dependency. However, there are some cases where knowing the type of coherence relation that an arc represents can be informative as to how likely that arc is to participate in a crossed dependency. The statistical results reported here also suggest that crossed

<sup>8</sup> The arc length distribution for the database overall does not include arcs of (absolute) length 1, since such arcs could not participate in crossed dependencies.

dependencies occur primarily locally, as evidenced by the distribution over lengths of arcs participating in crossed dependencies.

## 4.2 Nodes with multiple parents

Section 3.2 provided examples of coherence structure graphs that contain nodes with multiple parents. In addition to crossed dependencies, nodes with multiple parents are another reason why trees are inadequate for representing natural language coherence structures. The following sections report statistical results from our database on nodes with multiple parents. Similar to the previous section on crossed dependencies, we report results on the frequency of nodes with multiple parents (Section 4.2.1), the types of coherence relations ingoing to nodes with multiple parents (Section 4.2.2), and the arc length of coherence relations ingoing to nodes with multiple parents (Section 4.2.3). Section 4.2.4 provides a short summary of the statistical results on nodes with multiple parents.

### 4.2.1 Frequency of nodes with multiple parents

We determined the frequency of nodes with multiple parents by counting the number of nodes with in-degree greater than 1. We assume nodes with in-degree greater than 1 in a graph to be the equivalent of nodes with multiple parents in a tree. The result of our count indicated that 41.22% of all nodes in the database have an in-degree greater than 1. In addition to counting the number of nodes with in-degree greater than 1, we determined the mean in-degree of the nodes in our database. Table 8 shows that the mean in-degree (=mean number of parents) of all nodes in the investigated database of 135 texts is 1.6. Similar as for coherence relations involved in crossed dependencies (cf. Section 4.1.1), a linear regression showed a significant correlation between the number of arcs in a coherence graph and the number of nodes with multiple parents (cf. Figure 16;  $R^2 = 0.7258$ ,  $p < 10^{-4}$ ; for text length in discourse segments:  $R^2 = .6999$ ,  $p < 10^{-4}$ ; for text length in words:  $R^2 = .6022$ ,  $p < 10^{-4}$ ). The proportion of nodes with in-degree greater than 1 and the mean in-degree of the nodes in our database suggest that even if a mechanism could be derived for representing crossed dependencies in (augmented) tree graphs, nodes with multiple parents present another significant problem for trees representing coherence structures.

mean	1.60
min	1
max	12
median	1

Table 8. In-degrees of nodes in the overall database.

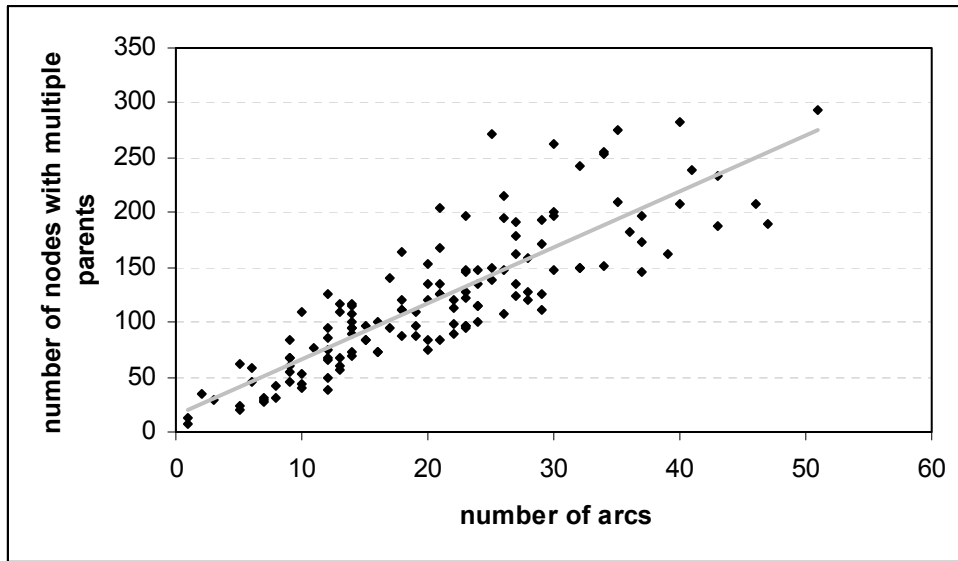


Figure 16. Correlation between number of arcs and number of nodes with multiple parents.

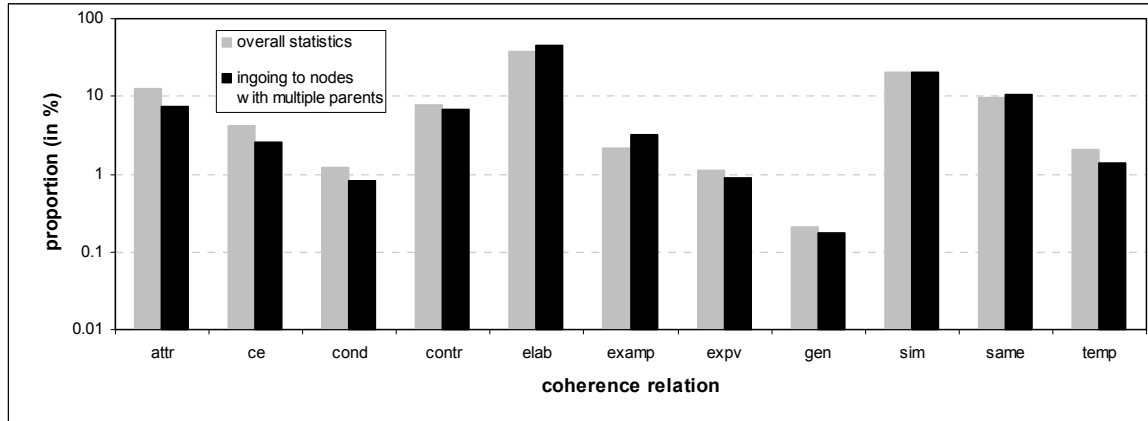
#### 4.2.2 Types of coherence relations ingoing to nodes with multiple parents

As with crossed dependencies, an important question is whether there are certain types of coherence relations that are more or less frequently ingoing to nodes with multiple parents than other types of coherence relations. In other words, the question is whether the frequency distribution over types of coherence relations is different for arcs ingoing to nodes with multiple parents compared to the overall frequency distribution over types of coherence relations in the whole database. Figure 17 shows that the overall distribution over types of coherence relations ingoing to nodes with multiple parents is not different from the distribution over types of coherence relations overall<sup>9</sup>. This is

<sup>9</sup> Notice that, unlike in Section 4.1.2, the distribution over coherence relations for all coherence relations includes arcs with length 1, since this time there was no reason to exclude them.



confirmed by a linear regression, which shows a significant correlation between the two distributions of percentages ( $R^2 = 0.967$   $p < 10^{-4}$ ).



**Figure 17. Distributions over types of coherence relations. For each condition (“overall statistics” and “ingoing to nodes with multiple parents”), the sum over all coherence relations is 100; each bar in each condition represents a fraction of the total of 100 in that condition. The y-axis uses a  $\log_{10}$  scale.**

Unlike for crossed dependencies (cf. Table 6), there are no big differences for individual coherence relations. Table 9 shows the data from Figure 17, ranked by the factor of “proportion of overall coherence relations” by “proportion of coherence relations ingoing to nodes with multiple parents”.

## REPRESENTING DISCOURSE COHERENCE

Coherence relation	Proportion of coherence relations ingoing to nodes with multiple parents (in %)	Proportion of overall coherence relations (in %)	Factor (= overall / ingoing to nodes with multiple parents)
<i>attribution</i>	7.38	12.68	1.72
<i>Cause-effect</i>	2.63	4.19	1.59
<i>temporal sequence</i>	1.38	2.11	1.53
<i>condition</i>	0.83	1.21	1.46
<i>violated expectation</i>	0.90	1.13	1.26
<i>generalization</i>	0.17	0.21	1.22
<i>contrast</i>	6.72	7.62	1.13
<i>same</i>	20.22	20.79	1.03
<i>similarity</i>	10.72	9.74	0.91
<i>elaboration</i>	45.83	38.13	0.83
<i>example</i>	3.20	2.19	0.68

**Table 9. Proportion of coherence relations.**

As for crossed dependencies, we also tested whether removing certain kinds of coherence relations reduces the mean in-degree (number of parents) and / or the proportion of nodes with in-degree greater than 1 (more than one parent). Table 10 shows that removing all *elaboration* relations from the database reduces the mean in-degree of nodes from 1.60 to 1.238, and the proportion of nodes with in-degree greater than 1 from 41.22% to 20.29%. Removing all *elaboration* as well as all *similarity* relations reduces these numbers further to 1.142 and 11.24% respectively. As Table 10 also shows, removing other types of coherence relations does not lead to as great reduction of the mean in-degree and proportion of nodes with in-degree greater than one.

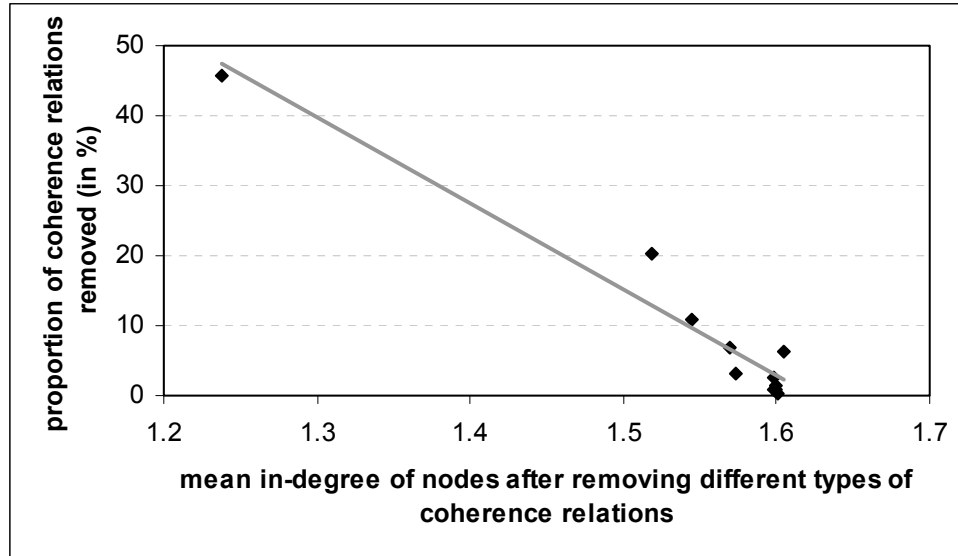
Coherence relation removed	In-degree of nodes				Proportion of nodes with in-degree > 1 (in %)
	mean	min	max	median	
<i>same</i>	1.519	1	12	1	35.85
<i>condition</i>	1.599	1	12	1	41.01
<i>attribution</i>	1.604	1	12	1	41.18
<i>temporal sequence</i>	1.599	1	12	1	41.12
<i>generalization</i>	1.6	1	12	1	41.16
<i>contrast</i>	1.569	1	12	1	39.45
<i>cause-effect</i>	1.599	1	12	1	41.14
<i>violated expectation</i>	1.598	1	12	1	40.96
<i>elaboration</i>	1.238	1	11	1	20.29
<i>example</i>	1.574	1	11	1	40.37
<i>similarity</i>	1.544	1	12	1	36.25
<i>elaboration and similarity</i>	1.142	1	11	1	11.24

**Table 10. The effect of removing different types of coherence relations on the mean in-degree of nodes and on the proportion of nodes with in-degree > 1.**

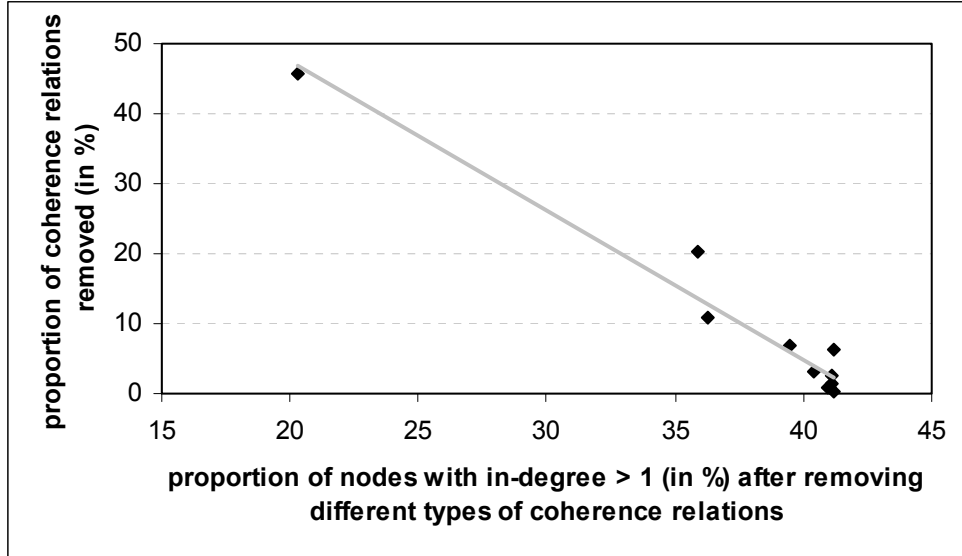
However, as with crossed dependencies (cf. Section 4.1.2), we also tested whether the reduction in nodes with multiple parents could simply be due to removing more and more coherence relations (i.e. the less dense a graph is, the smaller the chance that there are nodes with multiple parents). We correlated the proportion of coherence relations removed with the mean in-degree of the nodes after removing different types of coherence relations<sup>10</sup>. Figure 18 shows that the higher the proportion of removed coherence relations, the lower the mean in-degree of the nodes in the database becomes. This correlation is confirmed by a linear regression ( $R^2 = 0.9455$ ;  $p < 10^{-4}$ ; after removing the *elaboration* data point:  $R^2 = 0.8310$ ;  $p < .0005$ ; notice that these linear regressions do not include the data point *elaboration + similarity*). We also correlated the proportion of coherence relations removed with the proportion of nodes with in-degree greater than one after removing different types of coherence relations. Figure 19 shows that the higher the

<sup>10</sup> Notice that in the correlations in this chapter, the proportions of removed coherence relations include coherence relations of absolute arc length 1, because removing these coherence relations also has an effect on the mean in-degree of nodes and the proportion of nodes with in-degree greater than 1. Thus, the proportions of coherence relations removed in Figure 18 and in Figure 19 are from the third column of Table 9.

proportion of removed coherence relations, the lower the proportion of nodes with in-degree greater than one. This correlation is also confirmed by a linear regression ( $R^2 = 0.9574$ ,  $p < 10^{-4}$ ; after removing the *elaboration* data point:  $R^2 = 0.8146$ ,  $p < .0005$ ; notice that these correlations do not include the data point *elaboration + similarity*).



**Figure 18.** Correlation between proportion of removed coherence relations and mean in-degree of remaining nodes. Notice that the data point for *elaboration + similarity* is not included in the graph above. Both axes represent percent values.  $R^2 = 0.9455$ ;  $p < 10^{-4}$ .



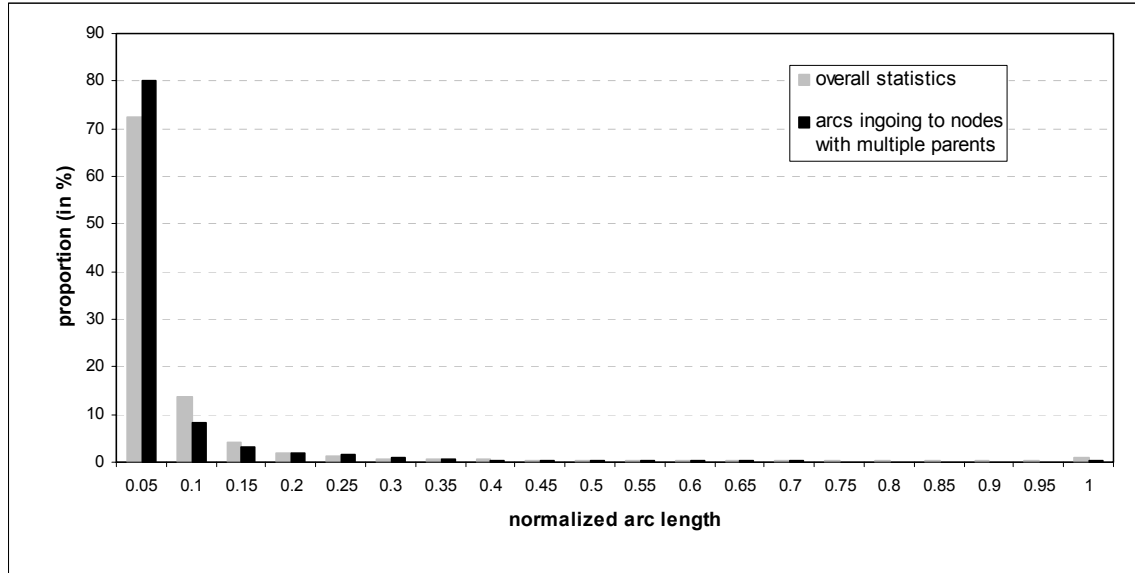
**Figure 19. Correlation between proportion of removed coherence relations and proportion of nodes with in-degree > 1. Notice that the data point for *elaboration* + *similarity* is not included in the graph above. Both axes represent percent values.  $R^2 = 0.9574$ ;  $p < 10^{-4}$ .**

Thus, while removing certain types of coherence relations (the same as for crossed dependencies, i.e. *elaboration* and *similarity*, cf. Section 4.1.2) can reduce the mean in-degree of nodes and the proportion of nodes with in-degree greater than one, the result is a very impoverished coherence structure. For example, after removing both *elaboration* and *similarity* relations, only 52.13% of all coherence relations would still be represented (cf. Table 9). In addition to having an impoverished coherence structure, notice that it is unclear a priori why removing *elaboration* and *similarity* relations but not removing other types of coherence relations should have an effect on the mean in-degree of nodes and on the proportion of nodes with in-degree greater than one (cf. the discussion in the last paragraph of Section 4.1.2).

#### 4.2.3 Arc lengths of coherence relations ingoing to nodes with multiple parents

As for crossed dependencies, we also compared arc lengths. Here, we compared the length of arcs that are ingoing to nodes with multiple parents to the overall distribution of arc length. Again, we compared normalized arc lengths (see Section 4.1 for the normalization procedure). By contrast to the comparison for crossed dependencies, we included arcs of (absolute) length 1 because such arcs can be ingoing to nodes with either single or multiple parents. Figure 20 shows that the distribution over arc lengths is

practically identical for the overall database and for arcs ingoing to nodes with multiple parents (linear regression:  $R^2 = 0.993$ ,  $p < 10^{-4}$ ), suggesting a strong locality bias for coherence relations overall as well as for those participating in crossed dependencies.



**Figure 20. Comparison of normalized arc length distributions. For each condition (“overall statistics” and “arcs ingoing to nodes with multiple parents”), the sum over all coherence relations is 100; each bar in each condition represents a fraction of the total of 100 in that condition.**

#### 4.2.4 Summary of statistical results on nodes with multiple parents

In sum, statistical results on nodes with multiple parents suggest that they are a frequent phenomenon, and that they are not limited to certain kinds of coherence relations. However, similar to crossed dependencies, removing certain kinds of coherence relations (*elaboration* and *similarity*) can reduce the mean in-degree of nodes and the proportion of nodes with in-degree greater than 1. But, also similar to crossed dependencies, our data at present do not distinguish whether this reduction in nodes with multiple parents is due to a property of the coherence relations removed (*elaboration* and *similarity*), or whether it is just that removing more and more coherence relations simply reduces the chance for nodes to have multiple parents. We plan to address this question in future research. In addition to the results on frequency of nodes with multiple parents and types of coherence

relations ingoing to nodes with multiple parents, the statistical results reported here suggest that ingoing arcs to nodes with multiple parents are primarily local.

## 5 Conclusion

The goals of this paper have been to present a set of coherence relations that are easy to code, and to illustrate the inadequacy of trees as a data structure for representing discourse coherence structures. We have developed a coding scheme with high inter-annotator reliability and used that scheme to annotate 135 texts with coherence relations. An investigation of these annotations has shown that discourse structures of naturally occurring texts contain various kinds of crossed dependencies as well as nodes with multiple parents. Both phenomena cannot be represented using trees. This implies that existing databases of coherence structures that use trees are not descriptively adequate.

Our statistical results suggest that crossed dependencies and nodes with multiple parents are not restricted phenomena that could be ignored or accommodated with a few exception rules. Furthermore, even if one could find a way of augmenting tree structures to account for crossed dependencies and nodes with multiple parents, there would have to be a mechanism for unifying the tree structure with the augmentation features. Thus, in terms of derivational complexity, trees would just shift the burden from having to derive a less constrained data structure to having to derive a unification of trees and features or coindexation.

Because trees are neither a descriptively adequate data structure for representing coherence structures nor easier to derive, we argue for less constrained graphs as a data structure for representing coherence structures. Such less constrained graphs would have the advantage of being able to adequately represent coherence structures in one single data structure (cf. Brants et al. (2002); Skut et al. (1997); König & Lezius (2000)). Furthermore, they are at least not harder to derive than (augmented) tree structures. The greater descriptive adequacy might in fact make them easier to derive. However, this is still an open issue and will have to be addressed in future research.

Another issue that should be addressed in future research is empirically viable constraints on inferences for building discourse structures. As pointed out in Section 3, we have argued against trees as a data structure for representing discourse structures;

however, that does not necessarily mean that discourse structures can be completely arbitrary. Future research should investigate questions such as whether there are structural constraints on coherence graphs (e.g. as proposed by Danlos (2004)), or whether there are systematic structural differences between the coherence graphs of texts that belong to different genres (e.g. as proposed by Bergler (1992)).

## References

- Sabine Bergler. 1991. *The semantics of collocational patterns for reporting verbs*. Paper presented at the 5th Conference of the European Chapter of the Association for Computational Linguistics, Berlin, Germany.
- Sabine Bergler. 1992. *Evidential analysis of reported speech*. Unpublished PhD thesis, Brandeis University, Waltham, MA, USA.
- Lawrence Birnbaum. 1982. *Argument molecules: A functional representation of argument structures*. Paper presented at the Third National Conference on Artificial Intelligence (AAAI-82), Pittsburgh, PA.
- Sabine Brants, Sabine Dipper, Silvia Hansen, Wolfgang Lezius, & George Smith. 2002. *The TIGER Treebank*. Paper presented at the Workshop on Treebanks and Linguistic Theories, Sozopol, Bulgaria.
- Bruce K Britton. 1994. Understanding expository text. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 641-674). Madison, USA: Academic Press.
- Jean Carletta. 1996. Assessing agreement on classification tasks: The kappa statistic. *Computational Linguistics*, 22(2), 249-254.
- Lynn Carlson, Daniel Marcu, & Mary E Okurowski. 2002. *RST Discourse Treebank*. Philadelphia, PA: Linguistic Data Consortium.
- Noam Chomsky. 1973. Conditions on transformations. In S. Anderson & P. Kiparsky (Eds.), *A Festschrift for Morris Halle* (pp. 232-286). New York: Holt, Rinehart and Winston.
- Simon Corston-Oliver. 1998. *Computing representations of the structure of written discourse*. Redmont, WA.
- Laurence Danlos. 2004. *Discourse dependency structures as DAGs*. Paper presented at the SigDIAL2004, Cambridge, MA, USA.



- Reinhard Diestel. 2000. *Graph theory*. New York: Springer Verlag.
- Barbara J Grosz, & Candace L Sidner. 1986. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3), 175-204.
- Donna Harman, & Mark Liberman. 1993. *TIPSTER complete*. Philadelphia, PA: Linguistic Data Consortium.
- Marti Hearst. 1997. TextTiling: Segmenting text into multi-paragraph subtopic passages. *Computational Linguistics*, 23(1), 33-64.
- Julia Hirschberg, & Christine H Nakatani. 1996. *A prosodic analysis of discourse segments in direction-giving monologues*. Paper presented at the 34th Annual Meeting of the Association for Computational Linguistics, Santa Cruz, CA.
- Jerry R Hobbs. 1985. *On the coherence and structure of discourse*. Stanford, CA.
- Jerry R Hobbs, Martin E Stickel, Douglas E Appelt, & Paul Martin. 1993. Interpretation as abduction. *Artificial Intelligence*, 63, 69-142.
- Eduard Hovy, & Elisabeth Maier. 1995. *Parsimonious or profligate: How many and which discourse relations?*
- Andrew Kehler. 2002. *Coherence, reference, and the theory of grammar*. Stanford, CA.
- Alistair Knott. 1996. *A data-driven methodology for motivating a set of coherence relations*. Unpublished PhD thesis, University of Edinburgh, Edinburgh.
- Esther König, & Wolfgang Lezius. 2000. *A description language for syntactically annotated corpora*. Paper presented at the Computational Linguistics Conference (COLING), Saarbrücken, Germany.
- Alex Lascarides, & Nicholas Asher. 1993. Temporal interpretation, discourse relations and common sense entailment. *Linguistics and Philosophy*, 16(5), 437-493.
- Robert E Longacre. 1983. *The grammar of discourse*. New York: Plenum Press.
- William C Mann, & Sandra A Thompson. 1988. Rhetorical structure theory: Toward a functional theory of text organization. *Text*, 8(3), 243-281.
- Daniel Marcu. 2000. *The theory and practice of discourse parsing and summarization*. Cambridge, MA: MIT Press.
- Mitchell Marcus, Grace Kim, Mary Ann Marcinkiewicz, Robert MacIntyre, Ann Bies, Mark Ferguson, et al. 1994. *The Penn Treebank: Annotating predicate argument*

- structure*. Paper presented at the ARPA Human Language Technology Workshop, San Francisco, CA.
- Kathleen R McKeown. 1985. *Text generation: Using discourse strategies and focus constraints to generate natural language text*. Cambridge, UK: Cambridge University Press.
- Johanna D Moore, & Martha E Pollack. 1992. A problem for RST: The need for multi-level discourse analysis. *Computational Linguistics*, 18(4), 537-544.
- Megan Moser, & Johanna D Moore. 1996. Toward a synthesis of two accounts of discourse structure. *Computational Linguistics*, 22(3), 409-419.
- Carolyn Penstein Rose, Barbara Di Eugenio, Lori S Levin, & Carol Van Ess-Dykema. 1995. *Discourse processing of dialogues with multiple threads*. Paper presented at the 33rd Annual Meeting of the Association for Computational Linguistics, Cambridge, MA.
- Livia Polanyi. 1996. *The linguistic structure of discourse*. Stanford, CA.
- Livia Polanyi, Chris Culy, Martin van den Berg, Gian Lorenzo Thione, & David Ahn. 2004. *A rule based approach to discourse parsing*. Paper presented at the SigDIAL 2004, Cambridge, MA, USA.
- Livia Polanyi, & Remko Scha. 1984. *A syntactic approach to discourse semantics*. Paper presented at the 10th International Conference on Computational Linguistics, Stanford, CA, USA.
- Rachel Reichman. 1985. *Getting computers to talk like you and me*. Cambridge, MA, USA: MIT Press.
- Stuart M Shieber. 1986. *An introduction to unification-based approaches to grammar*. Stanford, CA.
- Wojciech Skut, Brigitte Krenn, Thorsten Brants, & Hans Uszkoreit. 1997. *An annotation scheme for free word order languages*. Paper presented at the Fifth Conference on Applied Natural Language Processing (ANLP-97), Washington, D.C.
- Teun A van Dijk, & Walter Kintsch. 1983. *Strategies of discourse comprehension*. New York: Academic Press.

- Marilyn A Walker. 1998. Centering, anaphora resolution, and discourse structure. In E. Prince, A. Joshi & M. Walker (Eds.), *Centering Theory in discourse*. Oxford: Oxford University Press.
- Bonnie L Webber, Alistair Knott, Matthew Stone, & Aravind K Joshi. 1999. *Discourse relations: A structural and presuppositional account using lexicalized TAG*. Paper presented at the 37th Annual Meeting of the Association for Computational Linguistics (ACL-99), College Park, MD, USA.
- Bonnie L Webber, Matthew Stone, Aravind K Joshi, & Alistair Knott. 2003. Anaphora and discourse structure. *Computational Linguistics*, 29(4), 545-587.
- Florian Wolf, Edward Gibson, Amy Fisher, & Meredith Knight. 2003. *A procedure for collecting a database of texts annotated with coherence relations*. Unpublished manuscript, Cambridge, MA.
- Ingrid Zukerman, & Richard McConachy. 1995. *Generating discourse across several user modules: Maximizing belief while avoiding boredom and overload*. Paper presented at the Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI-95), Montreal, Canada.