Book Reviews

Coherence in Natural Language: Data Stuctures and Applications

Florian Wolf and Edward Gibson
(Massachusetts Institute of Technology)

Reviewed by
Alistair Knott
University of Otago

In his blurb on the back cover, Mark Liberman calls this book "the biggest step forward [in research on discourse structure] since Aristotle." Given this eminent recommendation, I read the book with great interest and some anticipation.

Wolf and Gibson's book contributes to the growing body of work on computational models of text structure. Following a tradition originating with Mann and Thompson (1988) and Hobbs (1985), Wolf and Gibson assume a "relational" account of coherence: the coherence of a text is attributed to the presence of relations holding at various levels of hierarchy between its constituent elements. Traditionally, relations hold between pairs of propositions; canonical examples are CAUSE, TEMPORAL SEQUENCE, and ELABORATION.

Work on relational models of text has pursued two related questions. One question concerns how to characterize the set of special relations whose presence can confer coherence on a text. How many such relations are there, and how should each relation be defined? A second question concerns where in a text we expect these special relations to be found. Answering this question involves using relations as building blocks in a theory of text structure. Wolf and Gibson's book addresses this latter question: their main aim is to propose and argue for a particular theory about the structure of relations in coherent text.

Broadly speaking, there are two alternative theories about the arrangement of relations in coherent text. Both are expressed using the terminology of graph theory: the sequence of atomic discourse segments in a text is interpreted as a set of nodes in a graph, ordered by adjacency, and relations introduce arcs (and sometimes additional nodes) into this graph. One theory proposes that the relations in a coherent text must form a tree structure. In this model, each relation holds between two adjacent nodes, and creates a new node representing a unit of text spanning both nodes which can itself be linked to other units by other relations; see Figure 1 for an example. The other theory is less restrictive, requiring only that the relations in a coherent text form a connected graph. In this model, arcs directly denote relations; see Figure 2 for an example. Note that in the graph model, relations can hold between discontiguous text segments and participate in crossing dependencies between segments, and that a single text segment can be involved in several relations.

Wolf and Gibson's central proposal is that a tree structure is too restrictive to represent the important relations in coherent text, and that a looser graph structure is preferable. In fact, their model allows some hierarchical structure; however, this is introduced to represent topic continuity between adjacent segments rather than text
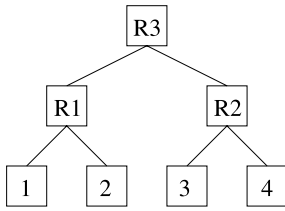
Figure 1
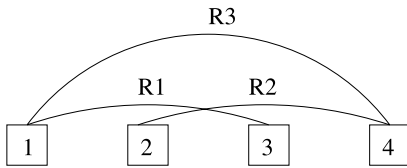Example of a tree-based model of text structure.

Figure 2
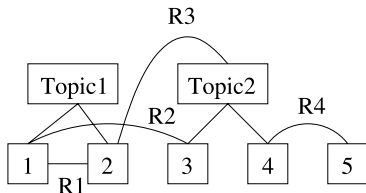Example of a graph-based model of text structure.

Figure 3
Example of Wolf and Gibson's model of text structure.

units created by regular discourse relations. An example of a Wolf-and-Gibson-style text structure is given in Figure 3.

Wolf and Gibson's arguments for their model of text structure derive mainly from an exercise in manual annotation. Two annotators hand-analyzed a corpus of 135 texts in the newspaper/newswire genre. In the first part of Chapter 2, Wolf and Gibson describe their model of text structure in detail, and give an admirably detailed protocol for manual analysis of texts, with copious examples. The remainder of the book outlines a number of separate arguments for their conception of text structure. I will first summarize these arguments, and then assess them.

Wolf and Gibson's first argument derives directly from an analysis of the text structures which make up the corpus of hand-annotated texts. In the second part of Chapter 2, they argue that there are many phenomena captured in these analyses that would have been missed by a model imposing a strict tree structure. In particular, their analyses reveal a large number of crossed dependencies, and a large number of nodes participating in more than one relation, neither of which can be accounted for with a tree-based model.

Two further arguments concern the influence of relations on the pattern of pronouns in a text. These arguments address the issue of the "psychological reality" of the relations hypothesized in the model: the authors reason that if discourse relations are

shown to influence the way pronouns are resolved or generated, this is evidence that they reflect real psychological representations in the minds of readers and writers. The first part of Chapter 3 reports on a psycholinguistic study, in which subjects read a series of two-clause sentences. The second clause of each sentence contained a pronoun whose antecedent was unambiguously found in the first clause (e.g., *Fiona defeated Craig, and so James congratulated him*). The connective in between the clauses was manipulated; it was found that this had an influence on the time taken by subjects to read the pronoun. The second part of Chapter 3 reports on a study of the pronouns in the annotated newspaper corpus. Again it was found that pronominalization preferences are different for different discourse relations.

A final argument, given in Chapter 4, concerns the utility of Wolf and Gibson's model of text structure in a practical text-processing application: text summarization. Wolf and Gibson develop a number of algorithms for text summarization that use their graph-based text structures. They then compare the quality of the summaries generated by these algorithms with summaries produced using other techniques, including some techniques using hand-annotated tree structures (from Carlson, Marcu, and Okurowski 2003). The graph-based summarization techniques outperform the other techniques, which they take as another piece of evidence in favor of a graph-based conception of text structure.

It is worth noting that the arguments relating to pronouns presented in Chapter 3 have a rather different aim from those presented in Chapters 2 and 4. The arguments given in Chapter 3 are only tangentially relevant to the issue of the structure of relations in coherent text. They bear on the psychological reality of individual relations, rather than on the general question of how relations organize text into coherent structures. The experiments in Chapter 3 certainly provide good evidence about the relevance of individual relations to pronoun generation and interpretation. In fact there is already quite a lot of experimental work showing the effect of coherence relations on pronoun interpretation; see, for example, the references cited by Stevenson et al. (2000) for a review. Much of this work uses a sentence completion paradigm, in which subjects are asked to complete a sentence like *Ken impressed Geoff because he...*; the connective used in a sentence has a strong influence on subjects' interpretation of the pronoun. Wolf and Gibson's experiments come to a similar conclusion about the influence of coherence relations on pronominal reference, using different experimental paradigms (self-paced reading and corpus analysis), and therefore extend the earlier findings. However, they do not bear on the authors' central hypothesis about text structure.

Wolf and Gibson's other two arguments bear directly on the question of whether texts should be analyzed using tree or graph structures. The argument in Chapter 2 is the most direct: the authors analyze a large number of texts, and identify many phenomena that a tree-based account would overlook. Although there have been several other analyses in this vein in the past, these have typically involved discussion of a small set of problematic example texts. Wolf and Gibson are the first to provide a large-scale quantitative evaluation of the coverage of a tree-based theory; they enumerate the frequency of crossed dependencies and of nodes with multiple parents in their corpus, to emphasize that these phenomena are widespread.

Are these analyses convincing evidence against a tree-based model of text structure? I do not believe so. My main concern is that the analyses assume the truth of the very theory they are being used to test. Wolf and Gibson's analysts follow a detailed protocol when analyzing the texts in the corpus. This protocol *allows* analysts to create crossing dependencies and re-entrant structures. It is unsurprising that the resulting analyses display these phenomena. To take a fanciful analogy: imagine proposing a

theory that holds that each discourse segment in a text is related to each other segment. We can certainly write a protocol to tell analysts how to annotate texts in line with this theory. No doubt analysts following this protocol will also achieve excellent inter-annotator agreement. However, we obviously can't use the set of analyses they produce as empirical evidence for the theory. Rather, the theory must be assessed in relation to its predictions about independently observable phenomena in discourse.[1]

It might be argued that although Wolf and Gibson *allow* re-entrancy and crossing dependencies, they do not *oblige* an analyst to find such phenomena in a text, and therefore that the presence of these phenomena still constitutes an empirical result. However, there are several somewhat nonstandard relations in Wolf and Gibson's model that introduce relations between discontinuous spans of text, and therefore heavily bias analyses towards crossing dependencies. For instance, one relation, called SAME-SEGMENT, is used to link two portions of a single sentence separated by a sentential modifier; for example, *[The economy,] according to some analysts, [is expected to improve]*. This relation is in fact *defined* to hold across discontinuous text segments. Another relation is used to analyze cases of explicit attribution of an utterance to a speaker; for example, *["Sure I'll be polite,"][promised one driver]*. Such attribution statements frequently appear within an extended speaker utterance, which again makes relations between discontinuous segments almost inevitable. Both SAME-SEGMENT and ATTRIBUTION are contentious as discourse relations because they apply between portions of propositions rather than whole propositions. And by creating relations between discontinuous text segments, both bias analyses towards crossing dependencies.

Wolf and Gibson do in fact consider the possibility that divergences from tree structure are due to a specific subset of relations. Indeed, they perform an analysis which apparently rules out this possibility. Surprisingly, this analysis seems to indicate that SAME-SEGMENT and ATTRIBUTION appear considerably *less* frequently in crossed dependencies than in analyses generally. This is very hard to believe, given that nearly all of their own examples of crossed dependencies involve one or other of these relations, and often involve both. In conclusion, Wolf and Gibson's collection of text analyses do not provide sufficient evidence to prefer the graph-based structural theory over the tree-based theory. These analyses are perhaps better thought of as embodying a statement of their theory than as an empirical test of it.

The argument presented in Chapter 4 is considerably more compelling as empirical support for a graph-based theory of text structure. To recap, the theory is used as the basis for an automatic text summarization algorithm, the results of which are compared to those of other algorithms, including algorithms based on the alternative tree-structure model of text. In this case, the theory (supplemented by its associated summarization algorithm) makes predictions about a quite distinct empirical phenomenon: the intuitions of naive readers about the relative importance of the segments of a text. The alternative tree-structure theory (supplemented by its own summarization algorithm) also makes such predictions. We can evaluate these two sets of predictions against data gathered from actual readers. And it turns out that the graph-based model of text structure outperforms the tree-based model.

It is not completely true to say that Wolf and Gibson's analysis protocol makes no reference to the concept of segment importance. In fact their definition of directed relations is similar to the definition of "nuclearity" in Mann and Thompson's (1988)

---

1 A similar warning about over-reliance on annotations in discourse theory was made recently in a squib in this journal (Zaenen 2006).

theory of coherence relations, and this makes explicit reference to the analyst's intuitions of segment importance. However, these decisions are all local to single relations between pairs of text segments. It is only if the relations in a text are combined in an appropriate way that these individual judgments will result in the right ranking of the segment importances over a whole text. (In fact, the alternative tree-based model of text also uses Mann and Thompson's notion of nuclearity, which effectively controls for the influence of these local judgments across the two types of analysis.) So the study on text summarization appears to provide quite solid evidence in favor of a graph-based model of text structure.

In conclusion: although Mark Liberman perhaps overstates the merits of Wolf and Gibson's book in his blurb, it nonetheless represents a significant advance in discourse theory.

References

Carlson, Lynn, Daniel Marcu, and Mary Ellen Okurowski. 2003. Building a discourse-tagged corpus in the framework of Rhetorical Structure Theory. In Jan van Kuppevelt, editor, *Current Directions in Discourse and Dialogue*. Kluwer, New York, NY.

Hobbs, Jerry. 1985. On the coherence and structure of discourse. Technical Report CSLI-85-37, Center for the Study of Language and Information, Stanford University, Palo Alto, CA.

Mann, William and Sandra Thompson. 1988. Rhetorical structure theory: A theory of text organization. *Text*, 8(3):243–281.

Stevenson, Rosemary, Alistair Knott, Jon Oberlander, and Sharon McDonald. 2000. Interpreting pronouns and connectives: Interactions between focusing, thematic roles and coherence relations. *Language and Cognitive Processes*, 15(3):225–262.

Zaenen, Annie. 2006. Mark-up barking up the wrong tree. *Computational Linguistics*, 32(4):577–580.

Alistair Knott is a senior lecturer in the Department of Computer Science at Otago University in New Zealand. His computational linguistics research interests are in computational models of discourse and dialogue. Knott's address is Department of Computer Science, University of Otago, P.O. Box 56, Dunedin, New Zealand; e-mail: alik@cs.otago.ac.nz.