The Holy Grail of discourse study is undoubtedly an account of coherence. As yet there are no definitive answers to questions such as: How do speakers/authors create coherent texts? How do hearers/readers understand sequences of sentences as coherent texts, even understanding them as the speaker/author intended (within acceptable limits)? How do two or more communicants cooperate in creating coherent dialog?

In Coherence in Natural Language: Data Structures and Applications (henceforth CNL), Wolf and Gibson (henceforth W&G) make a modest but significant contribution toward an account of discourse coherence. Theirs is a brief book, clearly and concisely written. It summarizes considerable relevant literature and reports on original research, interesting both for the results and for the methods used. CNL is directed primarily at computational linguists but it is by no means limited to this audience; others interested in discourse will benefit from reading it, if for no other reason than to get a glimpse of some ways computational linguistics is grappling with discourse.

To state the main issue W&G address, we must first make some assumptions: (1) A text is coherent because its parts (clauses) are “knit” together by “coherence relations ” such as temporal sequence, cause-effect, explanation, contrast, and so forth; some of these might be explicitly marked but others must be inferred. (2) A discourse is coherent when every part is tied by at least one coherence relation to another part or group of parts. (“No clause left behind!”). (3) Most discourses are not just chains of clauses, each related simply to the preceding clause. Rather, coherence relations structure the discourse in nontrivial ways.

W&G’s main concern is to understand the nature of the structure of discourse coherence. They state, “So far, the issue of descriptively adequate data structure for representing coherence has not been empirically evaluated” (CNL:3). It is this deficiency they address, but it is more than this: they seek a “data representation” for discourse coherence that facilitates the automation of processes that depend on coherence, such as information extraction and text summarization.
W&G state, “Most accounts [of discourse coherence] employ tree structure to represent the coherence relations between discourse segments in a text” (CNL:12). In section 2.2 they discuss this for various theories: Discourse Lexicalized Tree-Adjoining Grammar (Webber, Knot, Joshi, Stone); Segmented Discourse Representation (Lascarides, Asher); the three-level intentional structure of Grosz and Snider, and Rhetorical Structure Theory (Mann, Thompson). They also mention an approach proposed by Hobbs that does not use trees.

The problem with tree structures is that they do not handle (1) crossed dependencies and (2) nodes with multiple parents. An example of the former is seen in the following (CNL:15):

(1a) I wanted to drive through many states  
(1b) and I also want to see as many mountain ranges as possible.  
(1c) When I got to Arkansas,  
(1d) I stopped in the Ozarks.

Putting aside other relations, we note that (1c) relates to (1a) in speaking of states, and (1d) relates to (1b) in speaking of mountains. Thus the relation of (1d) to (1b) “crosses” the relation of (1c) to (1a).

An example of multiple parents is seen in the following (CNL:15):

(2a) The first to do that were the German jewelers,  
(2b) in particular Klaus Burie.  
(2c) And Morris followed very quickly after,  
(2d) using a lacquetry technique to make the brooch.

Here, (2c) is related by succession to the group (2a–2b) and by manner to (2d).

Although some theories based on tree structure exceptionally admit crossed dependencies and/or multiple parenthood, W&G argue that discourse coherence is better represented with “connected labeled chained graphs.” In these, discourse structure is a sequence of nodes, each node representing a discourse segment (corresponding roughly to a clause) or a group of segments bound by coherence (so, roughly, a coherent sequence of clauses). These are connected by labeled arcs that represent coherence relations between nodes.

W&G discuss how to define and identify segments, how these segments might be grouped, and various sets of coherence relations that have been proposed. Interestingly, these sets number from 2 (causal versus not-causal) to 400! (CNL:10); this is obviously an issue of gradience. For the purposes of their first research experiment, to which we now turn, W&G make modest modifications to a list by Hobbs and end up with just over a dozen relations.

W&G form a corpus of 135 texts from the Wall Street Journal and the AP Newswire. The first step was to code these texts for their discourse coherence structure. This was done by MIT undergraduate students. After some orientation, working independently, each identified the discourse segments, grouped discourse segments based on various unifying relations (e.g., being topically centered around the same entities or events), and determined coherence relations
between discourse segments or groups of segments. They did not—of course—always reach the same analysis, but the degree to which they agreed was impressive, and statistical analysis showed them to be significant. The results make it hard to deny the reality of coherence relations and the structures to which they give rise.

(A confession: the statistics were—for this reader—very rough going, and I cannot claim to have understood them all.)

Further statistics analysis of the annotated texts yielded an important fact: instances of both crossed dependencies and multiple parents abound. About crossed dependencies, W&G conclude, “Taken together, statistical results suggest that crossed dependencies are too frequent to be ignored by accounts of coherence relations. Furthermore, the results suggest that any type of coherence relation can participate in crossed dependency” (CNL:60). About multiple parents, W&G write, “In sum, statistical results on nodes with multiple parents suggests that they are a frequent phenomenon and that they are not limited to certain kinds of coherence relations” (CNL:65). Further, “…discourse structures of naturally occurring texts contain various kinds of crossed dependencies as well as nodes with multiple parents. Neither phenomenon can be represented using trees. This implies that existing databases of coherence structures that use trees are not descriptively adequate” (CNL:67). This strengthens W&S's suggestion that chain graphs are better for representing coherence structure.

In chapter 3, W&G discuss the relationship between coherence and pronoun resolution. They first review various approaches to pronoun resolution. Some of these seek to identify antecedents based on grammatical or semantic positions, like subject, object, or topic. Other accounts “argue that pronoun resolution strategies depend on the coherence relation between the clause containing the pronoun and the clause containing the antecedent” (CNL:69). Notable among these is the account proposed by Hobbs and pursued by Kehler. “Kehler extends Hobb's (1979) key insight that the establishment of coherence guides pronoun resolution and vice versa, noting that discourse coherence and pronoun resolution also mutually constrain each other. Thus, Kehler hypothesizes that how a pronoun is resolved may depend on the coherence relation between the clauses” (CNL:72).

To evaluate these theories, W&G did two experiments, about which we will give no details here. The first was an online self-paced reading experiment to test the different predictions of the pronoun processing accounts that had been discussed (CNL:73). The bottom line: “To summarize, the only account that makes the correct predictions for all conditions is Kehler's (2002). It predicts different preferences in pronoun resolution depending on the coherence relation between the clauses containing the pronoun and the antecedent.”

Whereas the first experiment tested online language comprehension, the second tested off-line language production preferences. About this test, W&G conclude, “In sum, Kehler's (2002) account is the only one that predicts the observed interaction between cause-effect and similarity. We also observe an overall preference for pronouns and antecedents in subject position across cause-effect and similarity, which is predicted by Centering Theory” (CNL:83). And further, “…Kehler's (2002) is the only one that predicts all observed preferences to be not a result of
operations of pronoun-specific mechanisms, but a byproduct of more general cognitive mechanisms and their interaction—establishing coherence and focusing attention” (CNL:83).

Chapter 4 is dedicated to a practical concern, that of the automatic generation of text summaries, driven by “the ever-increasing volume of text information available through the Internet” (CNL:87). W&G discuss methods of doing this, both how humans can do it and ways it has been automated. To get a benchmark for assessing automated performance, W&G first ask people to read 15 Wall Street Journal texts from the corpus and, for each sentence, rank its importance with respect to the text’s content. They then compare automated approaches.

One automated method ranks sentences based on layout, assuming that important information is located in prominent positions, e.g., the first sentence of each paragraph. Another is based on “important” words, ones that are very frequent in the particular text but not frequent in the corpus as a whole. Important sentences are identified as those in which “important” words cluster.

And then there are coherence-based approaches, based on the informational relations that hold between the sentences of a discourse. Among these, some represent the discourse structure as trees, and one such approach ranks sentences based on how high the sentence is in the (tree) structure of the discourse. Others assume a less constrained graph, like chain graph. For these, three methods of ranking sentences are discussed. “All three methods implement the intuition that sentences are more important if other sentences relate to them” (CNL:94).

W&S describe experiments and present the results: the simple layout-based algorithm “performed very poorly,” whereas “word-based and some coherence-based algorithms showed the best performance” (CNL:112). Further, “We found that coherence-based algorithms operating on chain graphs performed better than those performing on trees. Some preliminary results from our experiments suggest that this might in part be due to the fact that trees cannot represent nodes with multiple parents…” (CNL:113).

W&G end this chapter with a somewhat strange comment: “If our preliminary results […] can be supported in a larger, quantitative study, it would suggest that sentence informational content is another factor relevant to determining sentence importance, besides the relative position of a sentence in a coherence structure” (CNL:113). Hah! …sounds like a good pitch for more funding and support to really nail down what seems most obvious!

A final chapter “General Conclusions” recapitulates the various conclusions reached in the preceding chapters and suggests directions for further research.

References