Discourse structures: trees or graphs?

When we annotate a text, we read it, derive inferences, and make explicit some of these inferences in a consistent annotation standard. The end result of our annotation process depends directly on the constraints we impose on the inferences we allow ourselves to make and the representation language we adopt for making the allowable inferences explicit.

Previous work of Mann and Thompson (1988), Marcu (2000), and Carlson et al. (2002) for example, impose strong, structural constraints on both the representations one is allowed to create; and, more subtly, the set of allowable inferences. The recent work of Wolf and Gibson puts no (or fewer) constraints on the representations one is allowed to create and the set of allowable inferences.

At prima facie, the approach proposed by Wolf and Gibson is very well motivated from an empirical linguistic perspective. On what grounds after all, should someone declare that text structures are hierarchical? Instead of taking this claim for granted, it is better to devise an experiment in which humans annotate all discourse relations in a text using a protocol that imposes no structural constraints on the representations one builds. Once this is done, one can estimate empirically the degree to which trees (or graphs) are adequate representations of discourse structures.

The devil is though in the details because the outcome of such an experiment depends heavily on the nature of the inferences one is allowed to make. Consider, for example, the following text fragment, which is used by Wolf and Gibson (ex 12, page 12) to justify the inadequacy of using trees for representing discourse:

0. There is a train at Platform A.
1. Its destination is Rome.
2. There is another train at Platform B.
3. Its destination is Zurich.

Wolf and Gibson argue that the following relations hold between the units:

\[
\begin{align*}
&0 \rightarrow 1 \text{ elab} \\
&1 \rightarrow 2 \text{ elab} \\
&0 \leftrightarrow 2 \text{ sim} \\
&1 \leftrightarrow 3 \text{ contr}
\end{align*}
\]

Note though that not all these relations are created equal. The elab and sim relations can be inferred in isolation, in a context that contains only the units that they relate. Texts \(<0,1>, <2,3>, \text{ and } <0,2>\) all make sense by themselves. However, the contrast relation between units 1 and 3 cannot be interpreted in isolation. Text \(<1,3>\) “Its destination is Rome. Its destination is Zurich” does not make sense by itself. A human is capable of inferring a contrast relation between 1 and 3 only in the context of the whole text. Different contexts may yield different inferences. One can easily imagine contexts in which there could be a temporal relation between these two units (in a story about one train that goes from place to place and that stops in Rome and Zurich on the way).
Consider another example from Wolf and Gibson (ex. 13, page 15):
0. Farm prices in October edged up 0.7% from September
1. as raw milk prices continued to rise,
2. the Agriculture Department said.
3. Milk sold to the nation’s dairy plants and dealers averaged $14.50 for each hundred pounds,
4. up 50 percent from September and up $1.50 from October 1988,
5. the department said.

Among the discourse relations that Wolf and Gibson annotate here is a sim relation between units 2 and 5 “since both segments 2 and 5 state the same source”. I am not at all convinced that it is a good idea to annotate the relation between 2 and 5 as a coherence relation simply because the text refers in both instances to the Agriculture Department. It appears to me that it is more adequate to interpret this as a simple instance of cohesive relation, i.e., a co-reference link. It seems to me that if we don’t do so, we end up with ludicrous consequences. Just imagine a longer text about a company with many attributions spread all across the text “its CEO said”, “the CTO reported that”, “the COO declined that”, “the company said”, “the board of directors said”, etc. And many repeated instances of the same entity “the company” stating different things. Do we want a sim relation between all these pairs of discourse segments? Unlikely.

A useful parallel can be made between discourse and syntax with respect to the consequences of imposing no constraints on the allowable inferences. Consider, for example, the sentence:

“After walking hand in hand along the river, Mary and John sat on the bank playing with their bare feet in the cold water.”

One can associate a dependency or phrase structure analysis with this sentence in the expected fashion. However, one can also argue, for example, that the lexical dependencies expressed in a dependency or phrase-structure tree are insufficient for capturing the semantics of the sentence. To really understand this sentence, one needs to figure out the sense of the word “bank”. That is done only when one establishes a link between “bank” and “river” and “bank” and “water”. Naturally, such links will create at the syntactic level all the problems Wolf and Gibson observe at the text level: crossing links, words with multiple heads, etc. Imposing no constraints on the allowable links/inferences one should annotate at the syntactic level leads one to the same “problems” Wolf and Gibson discuss at the discourse level.

To summarize: I believe that the empirical results reported by Wolf and Gibson are a direct consequence of a lax, under-specified annotation protocol vis-à-vis the allowable inferences one is permitted to make when producing discourse relation annotations. I completely agree with Wolf and Gibson that discourse trees are a somewhat poor representation of text phenomena. In fact, Marcu et al. (1999) also argues that in some instances trees are not enough. Much more happens at the text level and we should certainly try to be inclusive and conceive of text theories that effectively explain both coherence and cohesive phenomena. However, I don’t think we should rush and throw
the baby with the bathwater simply because we annotate some cohesive links as coherence relations or exuberantly make explicit an unreasonable number of inferences.

Looking at the half the glass that is full, I would say that I find extremely encouraging to learn that by removing only 12.5% of the links in Wolf and Gibson’s corpus one can obtain tree structures. Given that Wolf and Gibson imposed no constraints on the set of allowable inferences it is surprising to learn that only 12.5% of the links had to be removed. This suggests that although trees are an approximation of discourse phenomena, they are a good approximation.

I believe the work of Wolf and Gibson is important because it forces us to think deeply about these issues, reconsider the range of inferences we would like to operate with, and challenges our understanding of the relation between coherence and cohesion. In previous work, Marcu (2000) has proposed a compositionality criterion to restrict the range of allowable inferences. According to that principle, a discourse relation holds between two large spans only if that relation also holds between the most important units in the spans. The important units are defined recursively as the units of the nuclei children. The compositionality principle of Marcu enables one to restrict the set of allowable inferences in the context of deriving discourse trees. Carlson et al. (2001) discuss additional structural criteria for restricting the range of allowable inferences. These principles are clearly not sufficient when deriving graph structures. Additional constraints will need to come into play. Without such constraints, one can easily get into a position where any segment is linked to any segment. Many papers in psycholinguistics show that humans have a strong tendency to interpret texts as coherent. When asked whether text fragments such as “George Bush is the president of the United States. Salmon is Michael Jackson’s favorite fish”, people have a strong tendency to invent explanations/scenarios as to why these sentences are juxtaposed and declare the text coherent. An unconstrained annotation protocol should try to calibrate for this.

For convenience, I include below hierarchical, non-crossing analyses of all the graphs in Wolf and Gibson’s paper.
Figure numbers correspond to those in Wolf and Gibson.

Figure 1

Figure 4: (note that the relation between 3 and 1 is implicit in the representation above by virtue of the compositionality criterion discussed by Marcu (2000).

Figure 6: This representation does not make explicit the cohesive relation between 4 and 1.

Figure 8