Making sense of discourse
On discourse segmentation and the linguistic marking of coherence relations

When people read or listen to a discourse, they will try to make sense of it. To fully understand a discourse, it is essential not only to know the meaning of each individual clause, but also to figure out how all clauses are related to each other. If all goes well, language users end up with an accurate representation of the discourse. An important aspect of building a mental representation of a discourse is inferring coherence relations between discourse segments. When inferring coherence relations, language users have to deduce whether two or more chunks of text constitute, for instance, a cause-consequence relation, a rule and an exception, alternatives, etc. Coherence relations can be explicitly marked by a connective (e.g., because, although) or a cue phrase (e.g., except for the fact that, by comparison), but this need not be the case; in many instances, coherence relations have to be established without the instructions connectives provide or, in the case of underspecified connectives, with only limited instructions.

This dissertation investigates both discourse segmentation and the linguistic marking of coherence relations. Using a combination of theoretical exploration, parallel corpus studies, and experimental methods, it provides new insights into how and why coherence relations are explicitly signaled and between which parts of a text people infer coherence relations. In addition, it contributes toward refining discourse segmentation and annotation guidelines, both of which are important methodological tools in the research of discourse coherence.