Tendency for Causality in Implicit Discourse Relations

Fatemeh Torabi Asr and Vera Demberg
MMCI Cluster of Excellence, Saarland University

Discourse relations can be expressed with or without their explicit cues: "The streets are wet because] it was raining". In the latter case they are called implicit, as the inference relies only on the content of the arguments. The question here is whether language users prefer causal relations to other types of relations in the absence of cues. We elaborate on some related hypotheses such forth in the literature via studying the proportion of implicit occurrence of relations in a large body of natural text, i.e., the Penn Discourse Tree Bank [3]. Sanders [4] proposed a causality-by-default hypothesis: "readers start out assuming the relation between consecutive sentences to be causal". Ac- cording to the Uniform Information Density hypothesis [1], humans tend to spread the intended information evenly across a text. At the level of discourse connectives, this would mean that presence of these cues is necessary when the relation is unexpected, but that a connective may be implicit if the relation is predictable. Putting together these two hypotheses, we would predict that causal relations are generally expected and their markers might be left implicit much more often than that of other kinds of relationships. Murray [2] also pro- posed a continuity hypothesis put that readers expect subsequent sentences in a text to be causally congruent and continuous, and that is why adversative connectives are more essential discourse cues than causal ones. Similarly, causal connectives that imply a non-linearity by presenting a consequence before its effect are more important cues than those, which keep the forward temporal transition. Therefore, we would predict that causal relations are more often left implicit than adversative relations. Also, forward causal relations (where the reason comes first) should be more frequently expressed without connectives than backward causals. We found that causal relations are the most frequent implicit relations among 16 types. In comparison with the total implicitness calculated for all relations (0.46), pragmatic cause and cause respectively obtained 0.86 and 0.65 implicitness. However, two other types of relations, namely restatement and instantiation beat causal types (0.95 and 0.82). We observed a significantly bigger implicitness ratio for forward vs. backward causality (0.69 vs. 0.62). Furthermore, each of these causal relations showed much bigger implicitness ratio than their negative counterparts (0.19 and 0.07). All of these measurements were significant at $p\text{-value} < 0.001$. Our findings partially confirm the causality-by-default hypothesis, but suggest that other types of relations exist that tend to appear with no discourse connective. Our observation about different types of causal relations is in line with the continuity hypothesis, which invites a study of temporal relations in the same corpus.

References:

[1] Frank and Jaeger (2008). Speaking rationally: uniform information density as an optimal strategy for language production. In Cogsci
[2] Murray (1997). Connectives and narrative text: The role of continuity. Memory & cognition.
[3] Prasad, Dinesh, Lee, Miltsakaki, Robaldo, Joshi and Webber (2008). The Penn Discourse Treebank 2.0.
[4] Sanders (2005). Coherence, causality and cognitive complexity in discourse. SEM.