Discourse Expectations and Implicitness of (Causal) Discourse Relations

Fatemeh Torabi Asr[1], Vera Demberg[1]

1. MMCI, Cluster of Excellence, Saarland University, Germany. fatemeh@coli.uni-saarland.de
    Keywords: Causality, Discourse relations, Discourse cues

Sentences are connected through discourse relations, which can be expressed explicitly using a discourse connector, or implicitly: "The streets are wet (because) it was raining." We here want to test existing hypotheses on discourse relations, which may suggest when and why discourse connectors can be omitted.

It has been argued that language users prefer causal relations to other types of relations (*causality-by-default hypothesis*, Sanders 2005). A second hypothesis is the *continuity hypothesis* (Murray 1997), which proposes that readers expect subsequent sentences to be causally congruent and continuous, hence predicting that causal relations are more expected than adversative ones and that causal relations that imply a non-linearity by presenting a consequence before its effect are less expected than those keeping the forward temporal transition. Finally, a recent study by Rohde and Horton (2010) indicates that implicit causality (IC) verbs (as in "Peter scolded Mary") may raise the expectation for a backward causal relationship.

Viewing these hypotheses in the context of the Uniform Information Density (UID) hypothesis (Frank and Jaeger 2008), which suggests that humans tend to spread information evenly across a text or utterance, thereby reducing or omitting redundant optional markers, we predict that (1) causal relationships are implicit more often than other discourse relationships, (2) forward causal relations should be implicit more often than backward causal relations and (3) backward causal relations that contain an IC verb are implicit more often than ones without an IC verb.

We test these hypotheses on the Penn Discourse Treebank (Prasad et al. 2008), a large body of text, which is annotated with explicit as well as implicit discourse relations. We find that causal relations are the most frequent implicit relationship in the corpus, and that the implicitness ratio for causals is significantly higher (0.65) than the general ratio of implicit relations (0.46; $p<0.001$). Overall, our findings partially confirm prediction (1), but also suggest that other commonly implicit discourse relations exist (in particular, instantiation and restatement). Regarding prediction (2), we observe a significantly larger implicitness ratio for forward causality in comparison to backward causality (0.69 vs. 0.62; $p<0.001$), thus providing support for the continuity hypothesis.

Finally, we find support for (Rohde and Horton 2010) in that sentences with an IC verb are more likely to be followed by a backward causal relationship than other verbs, and backward causal relationships are more likely to have an IC verb in their first argument (both $p<0.01$), but we do not find evidence of a higher likelihood to omit the connective when an IC verb is present, thus not confirming prediction (3).

References

Frank, A. and Jaeger, T.F. (2008). Speaking rationally: uniform information density as an optimal strategy for language production. In Cogsci

Murray, J.D. (1997). Connectives and narrative text: The role of continuity. *Memory & cognition*.

Prasad, R., Dinesh, N., Lee, A., Miltsakaki, L., Robaldo, L., Joshi, A. and Webber B. (2008). The Penn Discourse Treebank s2.0.

Rohde, H. and Horton, W. (2010). Why or what next? Eye movements reveal expectations about discourse direction. *CUNY*.

Sanders, T. (2005). Coherence, causality and cognitive complexity in discourse. *SEM*.