

## A discourse connector's distribution determines its interpretation

Fatemeh Torabi Asr & Vera Demberg (Saarland University)

fatemeh@coli.uni-saarland.de

Many connectives, such as *but* and *although*, can be used to mark very similar sets of relations, see Table 1. Fraser 1999 proposes that each connective has a core meaning and that a more specific discourse relation will be inferred from the content of the involved clauses. This implies that connectives which can mark the same relations have the same core meaning, and that alternating between two such connectors should not change the meaning of the discourse. A fully distributional account (Asr & Demberg 2013), on the other hand, describes the information content of a connective based on its usage patterns. This means that a connective may even have different meanings in different sentence positions (i.e. when used sentence-initially vs. between its arguments). This study shows how the fine-grained differences in the distribution of *but* vs. *although* vs. *sentence-initial although* affect text coherence. We created stories consisting of three sentences (see below) and normed them such that the first two sentences were equally acceptable in all conditions. The design was fully counter-balanced.

(1) **Context:** Amy's friends encouraged her to try tanning because her skin was so pale.

(2a/b/c\*) She thought of going to the beach, *but/although* her friends recommended a salon tan for her skin.

(3a/b) **Consistent with contrast/concession in 2:** She went to a nearby salon/beach to get a tan.

\*the *although*-initial condition 2c is created simply by reversing the order of the clauses in the *although*-mid condition.

The third sentence (story completion) is designed to be consistent either with a contrast reading or a concession reading of the discourse relation in the second sentence. The distributional account would predict that the acceptability of the story as a whole depends on how frequently the utilized connective occurs with the relation that is confirmed by the third sentence. According to the distributions we extracted from Penn Discourse Treebank (Table 1), *but* is most frequently used in contrast relations. Conversely, *although* is more likely to mark concession relations overall, yet, in a mid-args arrangement it is slightly more frequent in contrast relations.

<b>Pattern: <i>arg1 but arg2</i></b> contrast 48%, concession 15%, other 37%
<b>Pattern: <i>arg1 although arg2</i></b> contrast 39%, concession 37%, other 24%
<b>Pattern: <i>Although arg2 arg1</i></b> contrast 31%, concession 55%, other 14%

Table 1: *but* vs. *although* in PDTB relations

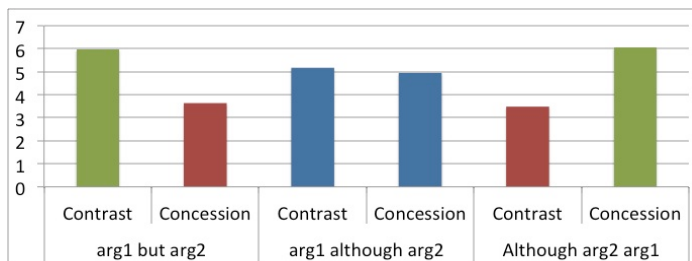


Fig. 1: Coherence scores by Amazon Mechanical Turk subjects

48 English native speakers rated 24 stories for coherence (Fig. 1). We found a significant interaction between connective type and the discourse relation type, which was disambiguated by the third sentence. While people scored contrast-based completions higher in case of *but*, they strongly preferred the completions consistent with a concession relation in the sentence-initial *although* conditions ( $p < 0.001$ ). More interestingly, the mid-sentence *although* conditions were scored to be equally coherent, i.e., no significant difference between completions of either type, which is consistent with distributions observed in the PDTB corpus. In addition to confirming the relevance of the distributional account, these results reveal that the contribution of the connectives to the meaning of a story goes beyond the interpretation within the boundaries of its argument: it can also affect the reader's expectation of the broader context, by affecting the information structure, e.g., changing the Question Under the Discussion (Roberts 1996) and modulating a possibly present implicature (Grice 1975). This study is followed up by an eye-tracking experiment.