

Computational Linguistics and Intellectual Technologies:
Proceedings of the International Conference “Dialogue 2016”

Moscow, June 1–4, 2016

CONCURRENT DISCOURSE RELATIONS

Bonnie Webber (bonnie@inf.ed.ac.uk)

University of Edinburgh, UK

The Penn Discourse Treebank 2.0 (PDTB2) was released to the public in 2008 and remains the world’s largest corpus of manually annotated discourse relations—both relations that are signaled explicitly (e.g., by a coordinating or subordinating conjunction, or by a discourse adverbial or other construction) and relations that otherwise appear implicit. Work is progressing on augmenting the PDTB2 in three ways: (1) by annotating many more forms of sentence-internal discourse relations; (2) by annotating implicit relations across paragraph boundaries; and (3) by identifying and annotating the presence of concurrent discourse relations. The new corpus will be called the PDTB 3.0 (or PDTB3).

The Penn Discourse TreeBank differs from other discourse-annotated corpora, not just in its size or its grounding in lexical, syntactic and/or positional evidence, but also in permitting more than one discourse relation to be annotated as holding concurrently.

In the PDTB2, annotators could indicate concurrent discourse relations by assigning multiple sense labels to an explicit connective. In the absence of an explicit connective, annotators could indicate concurrent discourse relations either by annotating a single implicit connective that they took to concurrently convey multiple senses or by annotating multiple implicit connectives, each conveying one of the concurrent relation(s). In the PDTB3, we have also had to allow for the possibility that a distinct implicit discourse relation can be inferred alongside an explicitly signalled one.

Evidence for such concurrent relations comes from linguist-generated minimal pairs, from existing corpora, and from judgments elicited through crowdsourcing experiments that we have been carrying out for the past year.

There are different circumstances in which different concurrent discourse relations are taken to hold. I will go through these, and conclude with what I take the implications of this to be for various language technologies.