

Concurrent Discourse Relations

ALTA 2013 Keynote Presentation

Bonnie Webber

School of Informatics
The University of Edinburgh
Edinburgh, United Kingdom
bonnie@inf.ed.ac.uk

Abstract

The Penn Discourse Treebank (PDTB) was released to the public in 2008 and remains the 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 ones that otherwise appear implicit.

The Penn Discourse TreeBank also diverges from other discourse-annotated corpora in permitting more than one discourse relation to be annotated as holding concurrently. Annotators could indicate this by assigning multiple sense labels to an explicit connective. Or, in those cases where adjacent sentences had no explicit connective, annotators could indicate concurrent discourse relations by either annotating a single implicit connective that concurrently conveyed multiple senses or annotating multiple implicit connectives, each conveying one of the concurrent relation(s). Subsequent experiments carried out using Mechanical Turk showed that, when a discourse adverbial explicitly signalled a discourse relation, there was often a separate concurrent relation that could be associated with an implicit coordinating or subordinating conjunction.

There are different circumstances in which different sets of 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, including statistical machine translation.