



Discourse Expectations and Implicitness of [Causal] Discourse Relations

Fatemeh Torabi Asr and Vera Demberg {fatemeh, vera}@coli.uni-saarland.de / MMCI Cluster of Excellence, Saarland University



1. Continuity and Causality

Do we expect consecutive sentences in a text to be causally related and temporally continuous?

Background:

- **Discourse relations:** Sentences in a text are related to one another via different types of discourse relations (e.g., cause-consequence: "Mary took Tim's ball. So he cried.")
- **Continuity (Murray 1997):** readers assume that events described in consecutive sentences follow a linear flow.
- **Causality (Sanders 2005):** readers try to establish causal relations between adjacent sentences by default.
- **Uniform Information Density (Frank & Jaeger 2008):** writers tend to spread information evenly across the text, thereby reducing or omitting optional discourse markers.

Our Corpus study:

- **Hypothesis 1:** explicit connectives of continuous and causal relations should be dropped more often than that of other relation types ("so" vs. "although").
- **Hypothesis 2:** In presence of other cues for a causal relation such as Implicit Causality verbs (Rohde & Horton 2010) omission of sentence connectives should happen more often.

2. Relations in the PDTB Corpus

Implicit vs. Explicit:

- **Penn Discourse Tree Bank** contains annotations of relations between adjacent sentences with or without discourse connectives. They added connectives to the implicit relations.

Example relations:

Explicit:

"The federal government suspended sales of U.S. saving bonds **because** Congress hasn't lifted the ceiling on government debt."
 – tagged as CONTINGENCY.Cause.reason

Implicit:

"The market was dragged up by the scruff of its neck by Wall Street and by market markers getting caught short. **but** No one wants stock on their books."
 – tagged as COMPARISON.Concession.contra-expectation

3. Implicitness

Method:

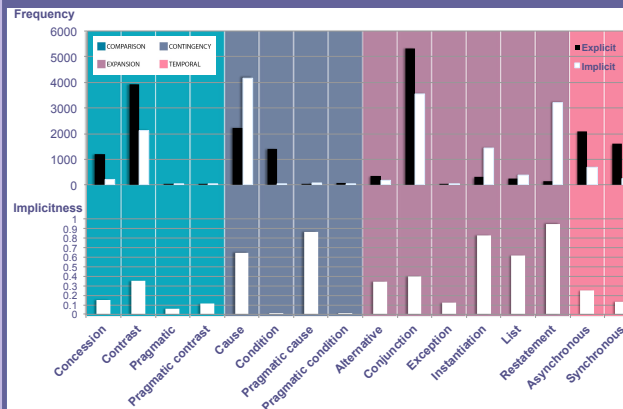
- We calculate the **implicitness** of a relation in the corpus as:

$$\text{Implicitness}(\text{relation}) = \frac{\text{implicit occurrences of the relation}}{\text{total occurrences of the relation}}$$

Intuition:

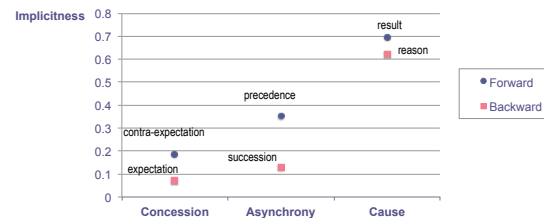
- We propose that a **bigger implicitness** denotes that a relation is **more expected** in establishing coherence during reading.

4. Results



Evidence for continuity:

- **Continuous relations** (Cause, Instantiation, Restatement, List) are more implicit than the discontinuous ones (Contrast, Concession, Exception).
- **Forward temporal flow** is more implicit than backward, in all subtypes of Cause, Concession and temporal Asynchronous relations (binomial test: p<0.001).



...Results

Evidence for causality:

- **Causal relations** are the most frequent implicit discourse relation.
- **Causal relations** exhibit a higher implicitness than that of other relations, typically compared in the literature including adversative and additive relations (binomial test: p<0.001).

Implicit Causality verbs as local cues:

- **IC verbs** are followed by causal discourse relations more often than other verbs (14% vs. 11.7%, significant at p< 0.01).
- **Unexpected result:** implicitness of the reason relations with an IC verb in their first argument is lower than that of reason relations with other verbs in the first argument (61% vs. 65% -- even after manual clean-up).

Relation	Total	Including IC in Arg1
Implicit reason	2462	164 (153 manually filtered)
Explicit reason	1324	108 (96 manually filtered)
Implicit all	15682	910
Explicit all	16147	1034

5. Conclusions

- **Continuity** in terms of temporal ordering of the events is implicit in the text, and when the relation between two statements is continuous, sentence connectives tend to be dropped.
- **Causal relations** are very probable relations to be inferred in the absence of discourse markers, however other continuous relations (according to the PDTB hierarchy) such as Restatement and Instantiation also tend to appear without specific sentence connectives.
- **IC verbs** are signals for causal reason relations, however, there was no larger rate of implicit connectors following IC verbs.

Abstract is available here: <http://www.coli.uni-saarland.de/~fatemeh>

[1] Frank and Jaeger (2008). Speaking rationally: uniform information density as an optimal strategy for language production. *Cogsci*.
 [2] Murray (1997). Connectives and narrative text: The role of continuity. *Memory & Cognition*.
 [3] Prasad, Dinesh, Leu, Mitsuko, Robaldo, Joshi and Weber (2008). The Penn Discourse Treebank 2.0.
 [4] Sanders (2005). Coherence, causality and cognitive complexity in discourse. *SEM*.
 [5] Rohde and Horton (2010). Why or what next? Eye movements reveal expectations about discourse direction. *CUNY*.
 Pictures of the tree and the apple are taken from <http://lars-berni.nl> and <http://s226.photobucket.com/profile/TheAppleGoddess>.

