

Crowd-Sourcing Concurrent Relations

Anna Dickinson, Hannah Rohde, Annie Louis, Chris Clark and Bonnie Webber
University of Edinburgh

While discourse relations can be signaled explicitly with conjunctions (Ex. 1) or adverbials (Ex. 2),

- (1) “We’ve started trying just about anything to keep sales moving in the stores,” says Kim Renk, a Swank vice president. But there are limits. [wsj0280]¹
- (2) They both called it a “welcome home” gathering. Nevertheless, an ANC rally by any other name is still an ANC rally. [wsj0559]

we also find sentences (Ex. 3–5) with both forms of DRD:

- (3) If that became public knowledge, the last bit of influence she had over her bank would be gone. So instead she hardened her soul and pretended to be a banker who was working her own will. [COCA]
- (4) It’s past ten. I could go to bed but instead I crawl out the window onto my little roof with the joint behind my ear. [COCA]
- (5) Appealing to a young audience, he scraps an old reference to Ozzie and Harriet and instead quotes the Grateful Dead. [wsj_1615]

In such cases, the conjunction and adverbial can each signal a distinct discourse relation. A previous crowd-sourced study of four adverbials that can co-occur with conjunctions (Jiang, 2013) asked respondents to insert a conjunction that had either been removed from a passage (using examples like 3-5) or never appeared (as in 6-7). The study of 100 passages (each viewed by 52 respondents) showed a strong distinction between different adverbials: With after all, they had an overall preference for because, whereas with instead, they varied passage-by-passage, reliably inserting but in (6) and so in (7):

- (6) Logically, she should be dead. Instead, she feels fine, caring for her daughters and walking a pedometer-measured two miles a day. [COCA]
- (7) He suspected he shouldn’t say that. Instead he lied. [COCA]

We call these *concurrent relations* because, as Jiang’s study shows, even without an explicit conjunction, two separate senses are concurrently conveyed. Ours is a study of concurrent relations. We aim to establish which discourse adverbials co-occur with conjunctions and which conjunctions each adverbial licenses and favors. Concurrent relations are a challenge for psycholinguists interested in whether an explicit linguistic signal is only needed when other evidence is too weak (a matter of discourse efficiency). They are also a challenge for language technology, which has heretofore assumed that a discourse relation was either signalled explicitly or inferred based on other clues, but not both at the same time.

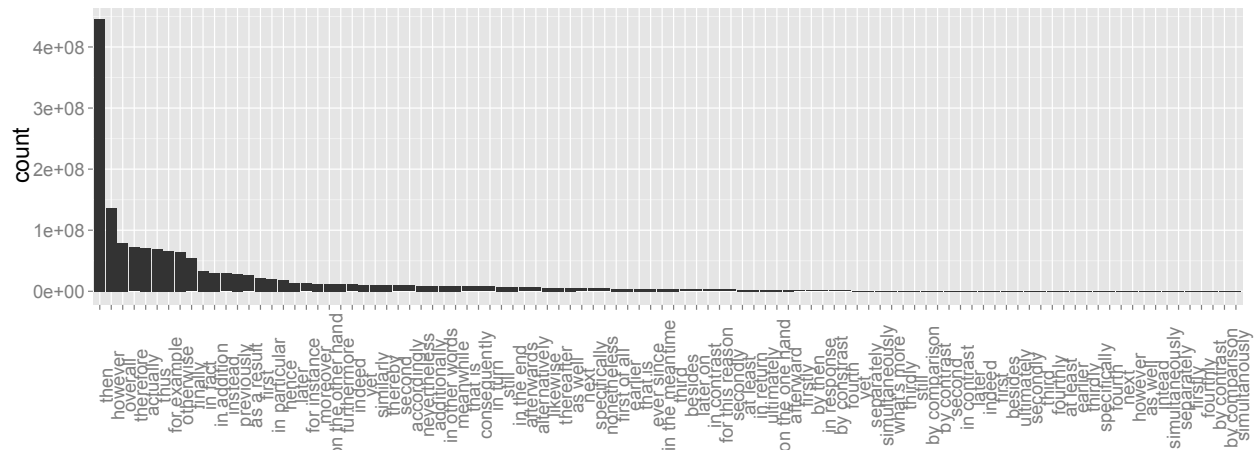


Figure 1: Adverbial ngram frequency: counts consist of instances with or without conjunction

¹References of this form are to files in the 1989 *Wall Street Journal* section of the Penn TreeBank.

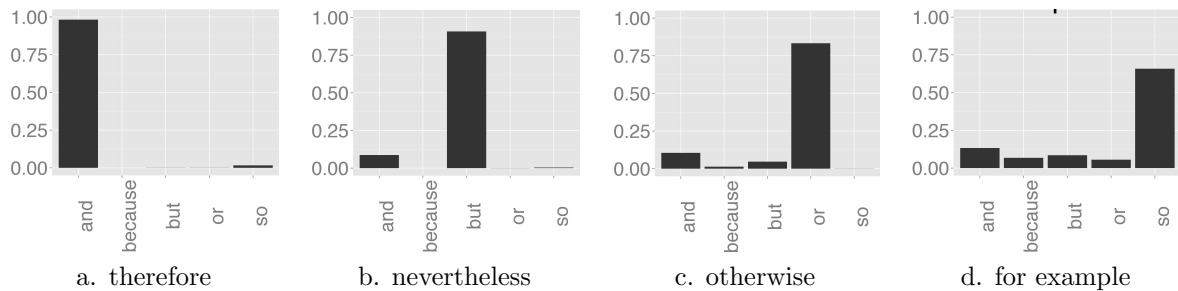


Figure 2: Sample adverbials with skewed distributions across conjunctions

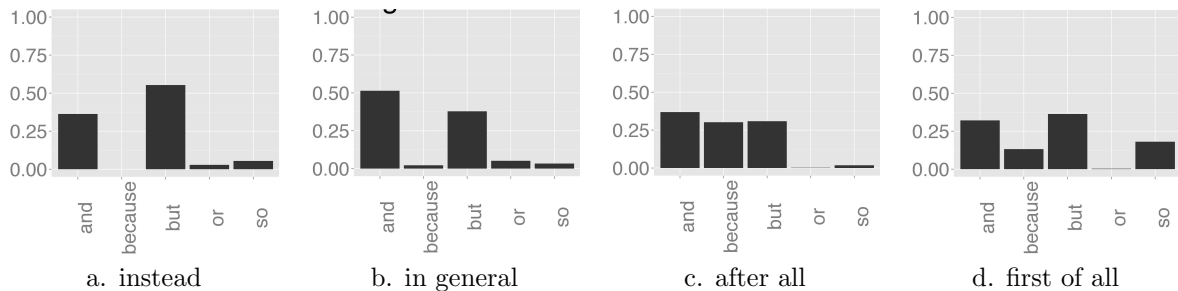


Figure 3: Sample adverbials with broader distributions across conjunctions

We have used Google N-grams to categorize discourse adverbials by frequency (high vs. low) and co-occurrence with conjunctions (skewed vs. uniform). Frequency of use is shown in Figure 1, where ‘use’ was restricted to an approximation of adverbial contexts (e.g., sentence-initial overall followed by a comma to filter out adjectival uses). While no adverbial was found to occur most frequently with a conjunction, the majority of adverbials had clear conjunction preferences (Figure 2), while some had a broader distribution (Figure 3). We speculate that adverbials with a strong conjunction bias may have an underlying semantic reason for that choice (perhaps an adverbial that signals consequence can only combine with a conjunction of that type), whereas adverbials that combine more flexibly with a set of conjunctions may permit properties of the linked propositions to determine the relevant conjunction.

Using a web interface akin to the one shown in Figure 4, the project will crowd-source judgments so as to replicate Jiang’s experiment more systematically and with broader coverage. The first dataset targets 20 adverbials, including the four tested by Jiang (*actually*, *after all*, *first of all*, *for example*, *for instance*, *however*, *in fact*, *in general*, *in other words*, *indeed*, *instead*, *nevertheless*, *nonetheless*, *on the one hand*, *on the other hand*, *otherwise*, *specifically/more specifically*, *then*, *therefore*, *thus*), appearing in 50 passages each. As in Jiang’s experiment, in half the passages, an overt conjunction was present in the original text, while in the other half, the adverbial originally appeared alone. Passages have been drawn primarily from the *New York Times Annotated Corpus* (Sandhaus, 2008) and the *Corpus of Contemporary American English*, using web text as needed for rarer conjunction-adverbial combinations.

A pre-test was conducted on the 20 targeted adverbials to pilot the interface and test whether other adverbials show patterns similar to what Jiang (2013) found for her small set. Three of the co-authors, who were naive to the status of each passage (as underlyingly implicit vs. explicit), judged a total of 760 passages, 260 implicit and 500 explicit. Their judgments confirmed Jiang’s results: For example, in the implicit after all passages, at least one judge responded with because in all but one case; likewise, in the after all passages which had originally appeared with an explicit conjunction, judges over-estimated the use of because for passages with an original and or but. In contrast, other adverbials were found to be more sensitive to the passage they appeared in: In the instead passages, for example, there was no single preferred conjunction. The example depicted in Figure 4 had an explicit so in the original text, though other senses are available (and and but are both plausible).

The pre-test pointed to several potential sources of judgment variability that we must be alert to in the crowd-sourced results. First, different readers may interpret the instructions differently. Currently, the instructions read (in part) *Your job is to make explicit the meaning that links the adjacent text spans. You must make a choice even if the insertion leads to an awkward or lengthy sentence, as long as you think the word brings out the meaning that links the adjacent text spans.* It is possible that readers may not understand that they are being encouraged to make a **sense** judgment about the author’s intended meaning, rather than a **stylistic** judgement about passage readability. The different specificity of the conjunctions may add further variability: Can we assume that a judgment to insert so is a more specific—

but not conflicting—judgment than and, such that a passage for which readers selected a mix of so and and is not necessarily an ambiguous passage? Even if yes, the pre-test has revealed that certain passages permit multiple conjunctions that are not necessarily more or less specific variants of each other. Consider example (8):

(8) You got to be nice to them ----- otherwise, they're not going to be nice to you.

While this passage had an explicit or in the original text, pre-test readers assigned a mix of or and because. And while these two conjunctions are typically associated with very different meanings, they seem to convey the same meaning here.

In summary, in contrast to models of DRD usage that assume that if one DRD is present in a passage, it signals a single relationship, our data raise three points: (i) there are many cases in which a DRD does not act alone (because afterall); (ii) a given DRD need not admit the presence of only one additional relationship (instead with but or so or because); and (iii) discourse adverbials differ in the way they combine with possible conjunctions—some demonstrating a clear preference, while others are more flexible.

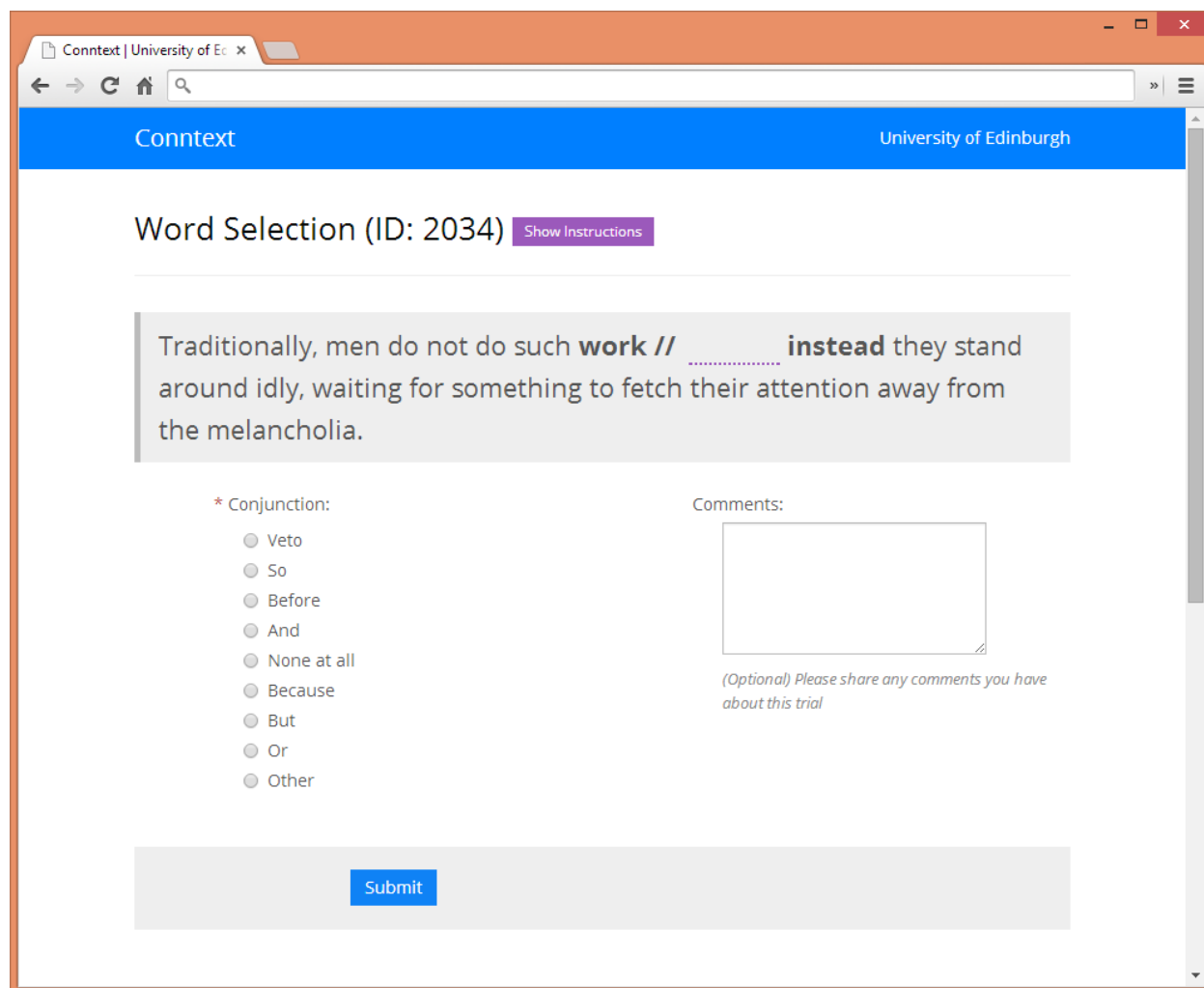


Figure 4: Experiment interface used during the pre-test, permitting either a conjunction to be selected or the example to be ‘vetoed’. The passage depicted appeared with so in the original text.

References

- Xi Jiang. *Predicting the use and interpretation of implicit and explicit discourse connectives*. MSc in English Language, Linguistics and English Language (LEL), University of Edinburgh, 2013.
- Evan Sandhaus. New York Times Annotated Corpus: Corpus overview. LDC catalogue entry LDC2008T19, 2008.