Discourse Coherence: Concurrent Explicit and Implicit Relations

Hannah Rohde[1]   **Alexander Johnson**[1]   **Nathan Schneider**[2]   **Bonnie Webber**[1]
[1] University of Edinburgh   [2] Georgetown University
{Hannah.Rohde, Bonnie.Webber}@ed.ac.uk, s1668719@ed-alumni.net, nathan.schneider@georgetown.edu

Abstract

Theories of discourse coherence posit relations between discourse segments as a key feature of coherent text. Our prior work suggests that multiple discourse relations can be simultaneously operative between two segments for reasons not predicted by the literature. Here we test how this joint presence can lead participants to endorse seemingly divergent conjunctions (e.g., *but* and *so*) to express the link they see between two segments. These apparent divergences are not symptomatic of participant naïveté or bias, but arise reliably from the concurrent availability of multiple relations between segments – some available through explicit signals and some via inference. We believe that these new results can both inform future progress in theoretical work on discourse coherence and lead to higher levels of performance in discourse parsing.

1 Introduction

A question that remains unresolved in work on discourse coherence is the nature and number of relations that can hold between clauses in a coherent text (Halliday and Hasan, 1976; Stede, 2012).

Our earlier work (Rohde et al., 2015, 2016) showed that, in the presence of explicit discourse adverbials, people also infer additional discourse relations that they take to hold jointly with those associated with the adverbials. For example, in:

(1) It's too far to walk. Instead let's take the bus.

people infer a RESULT relation in the context of the adverbial *instead*, which itself signals that the bus stands in a SUBSTITUTION relation to walking. We showed this using crowdsourced conjunction-insertion experiments (Rohde et al., 2015, 2016), in which participants were asked to insert into the gap between two discourse segments, a *conjunction* that best expressed how they took the segments to be related. Rohde et al. (2017) also asked participants to select any other conjunctions that they took to convey the same sense as their "best" choice. (More details of these experiments are given in Section 3.)

All three studies showed participants selecting conjunctions whose sense differed from that of the explicit discourse adverbial. But Rohde et al. (2015, 2016) also showed participants often selecting conjunctions that signal different coherence relations than those selected by other participants. And Rohde et al. (2017) showed participants often identifying very different conjunctions as conveying the same meaning. For example, in passage (2), with the discourse adverbial *in other words*, one large fraction of participants chose to insert OR, while another large fraction inserted SO. Since the two are neither synonymous nor representative of the same relation, either the participants have come up with different analyses of the passages (Section 2) or something more surprising is at work.

(2) Unfortunately, nearly 75,000 acres of tropical forest are converted or deforested every day _____ in other words an area the size of Central Park disappears every 16 minutes. [SO∼OR]

Rohde et al. (2017) noted other cases where different pairs of conjunctions (e.g., BECAUSE and BUT, BUT and OR, and BECAUSE and OR) appear systematically across participants and across passages for particular adverbials, and speculated on what these odd pairings may reveal, but did not provide any empirical evidence for why this happens. Here we present such evidence from an experiment on three discourse adverbials (*in other words*, *otherwise*, and *instead*).

After describing related work on multiple discourse relations (Section 2) and then our experimental methodology (Section 3), we step through results for these three adverbials. As a final piece of evidence, we manipulate the presence and absence of a fourth adverbial, *after all*, in order to

demonstrate that inference of the relation(s) between segments in a passage is not always driven by the presence of such an adverbial.

2 Related Work

This is not the first work on discourse coherence to acknowledge the possibility of multiple relations holding between given discourse segments.

For example, the developers of Rhetorical Structure Theory acknowledged that even experienced RST analysts may interpret a text differently in terms of the relations they take to hold (Mann and Thompson, 1988, p. 265). But while RST allows for multiple **alternative analyses** of a text in terms of discourse relations, in practice, researchers working in the RST framework standardly produce a single analysis of a text, with a single relational labeling, selecting the analysis that is "most plausible in terms of the perceived goals of the writer" (Mann et al., 1989, pp. 34–35). If that single analysis is later mapped into a different structure to support further processing – e.g., a binary branching tree structure – the mapping does not change the chosen relational labeling.

Multiple relations may additionally hold in theories of discourse coherence that posit multiple *levels* of text analysis. For example, following Grosz and Sidner (1986), Moore and Pollack (1992) characterized text as having both an *informational structure* (relating information conveyed by discourse segments) and an *intentional structure* (relating the functions of those segments with respect to what the speaker is trying to accomplish through the text). The kinds of relations at the two levels are different, as can be seen in the following example from (Moore and Pollack, 1992, p. 540):

(3) a. George Bush supports big business.
b. He's sure to veto House Bill 1711.

At the level of intentions, (3a) aims to provide EVIDENCE for the claim in (3b), while at an informational level, (3a) serves as the CAUSE of the situation in (3b). RST would force annotators to choose only the analysis that best reflected the perceived goals of the writer.

Additionally, multiple relations can hold where there are **distinct explicit signals** for distinct discourse relations holding between a pair of segments (Cuenca and Marin, 2009; Fraser, 2013), as in:

(4) It's too far to walk. So instead let's take the bus.

where the conjunction *so* signals a RESULT relation and the adverbial *instead* signals that taking the bus stands in an SUBSTITUTION relation to walking.

Finally, a fourth way in which the previous literature has taken multiple discourse relations to hold is when a single phrase or lexico-syntactic construction **jointly signals** multiple discourse relations as holding over a text – for example, *since* as a subordinating conjunction may, in particular contexts, signal both a TEMPORAL relation and a CAUSAL relation, rather than just one or the other (Miltsakaki et al., 2005).

We are aware of only two resources that allow more than one discourse relation to be annotated between two segments – the Penn Discourse TreeBank (PDTB; Prasad et al., 2008, 2014) and, more recently, the BECauSE Corpus 2.0 (Dunietz et al., 2017). The PDTB allows multiple discourse relations of the third and fourth types noted above. It also allows them to be annotated if there is no explicit connective between a pair of segments but annotators see more than one sense relation as linking them, as in the following variant of (4):

(5) It's too far to walk. Let's take the bus.

Here a RESULT relation can be associated with an implicit token of *so* between the clauses, while a SUBSTITUTION relation can be associated with an implicit token of *instead*. The above are the main cases in which PDTB annotates multiple relations. Relevant to this paper, the PDTB does not annotate implicit conjunction relations where there is already an explicit discourse adverbial. Thus the PDTB would either ignore the implicit RESULT relation for (1) or (incorrectly) annotate *instead* in (1) as conveying both SUBSTITUTION and RESULT.

Moreover, while the PDTB has been used in training many (but not all) discourse parsers (Marcu, 2000; Lin et al., 2014; Feng and Hirst, 2012; Xue et al., 2015, 2016; Ji and Eisenstein, 2014), discourse parsing has for the most part ignored its annotations of multiple concurrent relations between clauses, except in the case of distinct explicit connectives expressing distinct relations. Instead, they have arbitrarily taken just a single relation to hold, even though the relations are simply recorded in an *a priori* canonical order. This practice is problematic because, for example, there may well be a difference in the properties of segments where two relations are jointly seen to hold, versus those segments in which only one or the other holds. This can result in unwanted noise in the data and lower the reliability of whatever is induced.

While our previous studies showed another source of multiple discourse relations holding con-

currently between discourse segments, the work reported here explains how, in the context of multiple relations, participants can take very different conjunctions to be conveying the same relation, and what can change participants' selection of a conjunction to mark the relation they infer alongside that conveyed by an explicit discourse adverbial.

3 Methodology

A locally crowdsourced conjunction-insertion task provided a proxy for labelling relations between adjacent discourse segments within a passage.

Our materials consisted of passages containing an explicit discourse adverbial, preceded by a gap, which effectively separated the passage into two segments. The passages consisted of 16 with *in other words*, 16 with *instead*, 16 with *after all*, and 48 with *otherwise*. Participants were asked to read each passage and choose the conjunction(s) that best expressed how the two segments link together. The presentation of conjunction choices varied in order for each participant, but always consisted of AND, BECAUSE, BUT, OR, SO, NONE. While the task admittedly encourages participants to select one (or more) conjunctions, our prior work has shown that participants are very willing to use NONE if no conjunction is appropriate. We therefore take their insertion of a conjunction as their endorsement of the relation signaled by that conjunction. To further control data quality, we included 6 catch trials with an expected correct conjunction like "To be _____ not to be".

Three of the explicit discourse adverbials that we chose are *anaphoric*: *in other words*, *otherwise*, and *instead* (Webber et al., 2000). Unlike conjunctions such as AND, BECAUSE, BUT, OR and SO, they are not constrained by structure as to what they establish discourse relations with. So a conjunction-insertion task can be used to assess links between the segments (see also Scholman and Demberg 2017). Our three anaphoric adverbials share a core meaning of 'otherness' via their lexical semantics and flexibility in the relations they can participate in, making them a fruitful set to compare. The fourth adverbial, *after all*, allows us to test a hypothesis that the inferred connection between clauses is not driven by the adverbial alone.

These particular adverbials were selected because they had yielded unexpected combinations of conjunction insertions in our prior work (e.g., OR/SO with *in other words*). This is in contrast to adverbials like *therefore* and *nevertheless*, for which participants' conjunction combinations could be attributed to variation in the specificity of the conjunctions (SO/AND for *therefore*, BUT/AND for *nevertheless*). For our selection of a set of conjunctions to use as proxies for relation labels, we included all the coordinating conjunctions in English, as well as the subordinating conjunction BECAUSE as EXPLANATION relations are frequent.

All participants (N=28) were monolingual native English speakers who were selected following a pre-test to measure their ability to consistently insert conjunctions that captured the underlying coherence relations in a series of passages. All gave informed consent. They each received £50 for their time. Each participant saw one of two randomly ordered lists. Passages were presented in batches of 34, one batch per day for three days.

The materials were simplified variants of naturally occurring passages. Some were also manipulated systematically, in ways aimed at altering the availability of different coherence relations. Passages are available via the "dataset" link on the paper in the ACL anthology, and predictions about them are laid out in Sections 4.1–4.4.

4 Datasets

4.1 *In other words* Dataset

Rohde et al. (2016) report an OR∼SO response split for *in other words* when participants could insert only their top choice of conjunction. Figure 1 shows SO dominating participants' choice in all cases, but OR showing up among their choices in all but one passage (leftmost vertical bar). Additionally, several passages elicited BUT as the top choice of some participants.
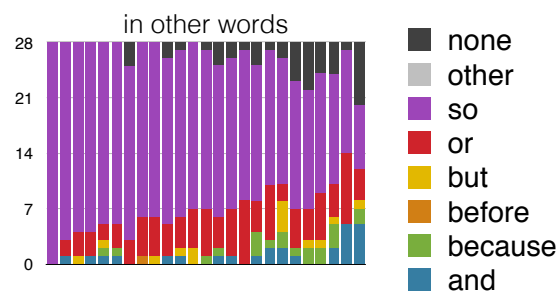
Figure 1: Stacked bar chart for conjunction insertions in passages with *in other words* (Rohde et al., 2016). Each vertical bar represents a passage with one response from each participant (N=28, no overlap with current participants).

The *in other words* passages of the current experiment tested two linked hypotheses: The first is that OR~SO response splits arise from two components of the lexical semantics of the adverbial itself: its sense of an evoked alternative and its sense of a consequence via restatement, whereby the truth of the second segment holds because it provides a reformulated restatement of the first segment's content. For passage (2), this corresponds to the deforestation of 75,000 acres of tropical forest entailing the disappearance of an area the size of Central Park every 16 minutes.

The second hypothesis is that the prevalence of and substitutability between SO and OR in (2) depends on the immediately adjacency of the two segments. This was suggested by participant choices of BUT (cf. Figure 1), as well as the observation that *in other words* does not always license OR via its lexical semantics and SO via entailment, as shown in (6), where BUT has become more available. Note that none of the relations conveyed by these conjunctions (CONTRAST or CONCESSION for BUT, DISJUNCTION for OR, CONSEQUENCE for SO) are already conveyed by the adverbial itself, which for *in other words*) would be RESTATEMENT.

(6) Unfortunately, nearly 75,000 acres of tropical forest are converted or deforested every day. *I don't know where I heard that* _____ in other words an area the size of Central Park disappears every 16 minutes.

We tested these hypotheses by creating minimal pairs of 16 passages containing *in other words*. The pairs varied in the presence/absence of a metalinguistic comment intervening between the original description and its reformulation, as in (7)–(8).

(7) Typically, a cast-iron wood-burning stove is 60 percent efficient _____ in other words 40 percent of the wood ends up as ash, smoke or lost heat.

(8) Typically, a cast-iron wood-burning stove is 60 percent efficient. *How this is measured is unclear* _____ in other words 40 percent of the wood ends up as ash, smoke or lost heat.

For each passage, participants identified their preferred conjunction and then any others that they took to convey the same sense. Half the participants saw a given passage with no intervening metalinguistic comment, half with.

If our hypotheses are confirmed, it will show that manipulating the immediately preceding segment can shift participants' preference from relations associated with OR and SO (ALTERNATIVE and CONSEQUENCE) to relations of CONTRAST or CONCESSION. This would then be evidence that adjacency affects what coherence relations participants take to be available.
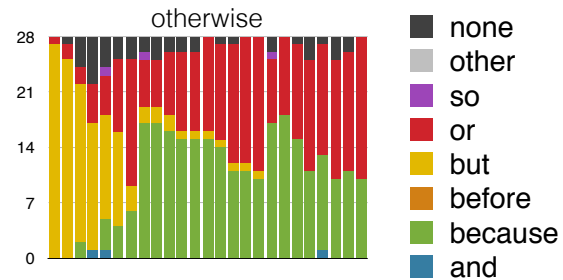
Figure 2: Stacked bar chart for participants' (N=28) conjunction insertions in *otherwise* passages (Rohde et al., 2016)

4.2 *Otherwise* Dataset

Rohde et al. (2016) report surprising response splits amongst BECAUSE~BUT~OR for *otherwise* in their conjunction-insertion data (Figure 2). Given that *otherwise* has several different functions (described below), we hypothesize that different response splits arise from the lexical semantics of *otherwise*, combined with inference as to the function of the *otherwise* clause in a given passage.

One function of *otherwise* is in ARGUMENTATION. Here, an *otherwise* clause provides a reason for a given claim, as in (9). Another function is in ENUMERATION, when the speaker first gives some preferred or more salient options, the *otherwise* clause introduces other alternative options, as in (10). A third use is in expressing an EXCEPTION to a generalization. Here, the main clause expresses a generalization, while *otherwise* clause specifies an exception (disjunctive alternative) to it, as in (11).

(9) Proper placement of the testing device is an important issue _____ otherwise the test results will be inaccurate.

(10) A baked potato, plonked on a side plate with sour cream flecked with chives, is the perfect accompaniment _____ otherwise you could serve a green salad and some good country bread.

(11) Mr. Lurie and Mr. Jarmusch actually catch a shark, a thrashing 10-footer _____ otherwise the action is light.

Results presented in (Rohde et al., 2017) for passages like (9) showed participant judgments of OR and BECAUSE, but not BUT. Passages like (10) yielded pairings of OR and BUT, but not BECAUSE. Lastly, passages like (11) yielded response splits between BUT and the less specific AND (Knott, 1996).

Note that due to overlaps in conjunction choice, some conjunctions cannot be unambiguously associated with a single use of *otherwise*: While BECAUSE may unambiguously signal that a participant has inferred ARGUMENTATION, OR might indicate inference of either ARGUMENTATION or

ENUMERATION. Thus we probe both participant choices of connectives and (via paraphrase) the use of *otherwise* that they take to hold.

We chose 16 passages for each use of *otherwise*, based on our own category judgments. For each passage, we asked participants to select the conjunction that best expressed how its two segments were related, and then any other connectives that they took to express the same thing.

A paraphrase task was then used as further evidence for the relation participants inferred in the *otherwise* passages. After completing a given session's batch of passages, participants were asked to select which of three options they took to be a valid paraphrase of the passage. Each use of *otherwise* was assigned a distinct paraphrase to link the left-hand and right-hand segments (LHS, RHS).

- ARGUMENTATION: "A reason for ⟨LHS⟩ is ⟨RHS⟩."

- EXCEPTION: "Generally ⟨RHS⟩. An exception is when ⟨LHS⟩."

- ENUMERATION: "There's more than one good option for ⟨goal⟩. They are: ⟨LHS⟩, ⟨RHS⟩."

We also allowed participants to choose a second paraphrase if they thought it appropriate.

4.3 *Instead* Dataset

Rohde et al. (2016) report a range of participant choices in conjunction-insertion passages involving *instead* (Figure 3). For passages on the left of the figure, participants uniformly chose BUT, while the passage on the far right yielded a strong preference for SO. Elsewhere, some chose BUT and some chose SO. (For the current experiment, we ignore the fact that AND can contingently substitute for either BUT or SO as a connective in text (Knott, 1996), focussing only on passages where participants explicitly choose BUT and/or SO.)

Rohde et al. (2017) report even more surprising participant responses to passages such as (12), where some participants selected both BUT and SO as equally expressing how the segments in the passage were related.

(12) There may not be a flight scheduled to Loja today _____ instead we can go to Cuenca. [BUT∼SO]

Neither the inter-participant split between BUT and SO in (Rohde et al., 2016) nor the intra-participant split between them (Rohde et al., 2017) can be explained in terms of *instead* itself, since
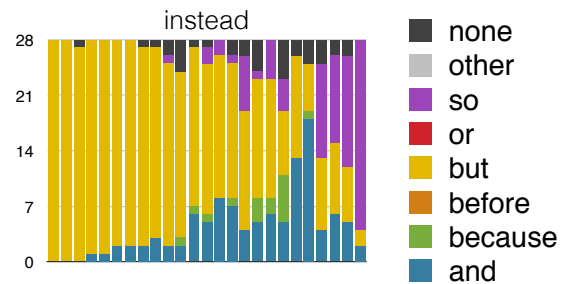
Figure 3: Stacked bar chart for participants' (N=28) conjunction insertions in *instead* passages (Rohde et al., 2016)

instead simply conveys that what follows is an alternative to an unrealised situation in the context (Prasad et al., 2008; Webber, 2013). The current experiment tests the hypothesis that this BUT∼SO split is a consequence of inference from properties of the segments themselves.

To test this hypothesis, we created 16 minimal pairs of passages containing *instead*, one of which emphasized the information structural parallelism between the clauses, as in (13a), and another variant (13b) that de-emphasized that parallelism in favor of a causal link implied by a downward-entailing construction such as *too X* (Webber, 2013). For each passage, half the participants saw the parallelism variant in the conjunction-insertion task, while half saw the causal variant.

(13) a. There was no flight scheduled to Loja yesterday _____ instead there were several to Cuenca.

   b. There were too few flights scheduled to Loja yesterday _____ instead we went to Cuenca.

4.4 *After all* Dataset

In (Rohde et al., 2017), we reported a BECAUSE∼BUT response split for passages containing *after all*. We speculated that this may be because a passage such as (14) below presents an argument in which the second segment serves as a REASON (hence, BECAUSE) for the first segment, but also serves to CONTRAST with it (hence, BUT).

(14) Yes, I suppose there's a certain element of danger in it _____ (after all) there's a certain amount of danger in living, whatever you do.

We hypothesize that the BECAUSE∼BUT split cannot be a consequence of the adverbial *after all*, which the Cambridge Dictionary indicates is "used to add information that shows that what you have just said is true".[1] If REASON and/or CONTRAST

---

[1] https://dictionary.cambridge.org/us/dictionary/english/after-all
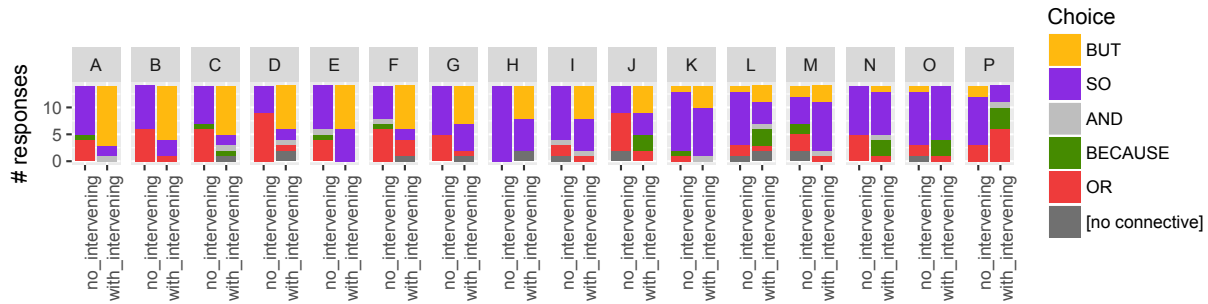
Figure 4: Distribution of participants' first choice of conjunction for passages with *in other words*. Each participant saw only one variant. Each vertical bar represents a passage with the responses from each participant, color-coded by conjunction.

are being conveyed, it can't be a consequence of *after all*. As such, this response split must depend on the reasoning that supports the inference of coherence between the two segments, separate from the adverbial itself.

We test the hypothesis that the response split is independent of the presence or absence of *after all*. Starting with 16 passages that originally contained *after all*, we created a variant of each passage without the adverbial. The conjunction insertion task was the same as with the other datasets.

5 Results

5.1 *In other words*: Inference and adjacency

Section 4.1 lays out the joint hypotheses that inferred relations in passages with *in other words* reflect two components of the lexical semantics of the adverbial (leading to the OR∼SO split) and that the presence of intervening material before *in other words* reduces the availability of those relations, favoring BUT instead.

Figure 4 shows the predicted pattern: The no-intervening-content condition primarily yields OR/SO responses (with variation across passages on the OR-vs.-SO preference) with a relative increase in BUT responses in the intervening-content condition.[2] Passage B corresponds to the pair of examples (2)/(6), and passage C reflects (7)/(8).

For the analysis here and in Section 5.3, a relevant first-choice conjunction was chosen and the binary outcome of its insertion was modeled with a mixed-effect logistic regression. Here, the insertion of OR indeed varied with the presence/absence of intervening material ($\beta = -1.569$, $p < 0.005$).

We posit that increases in BUT associated with the intervening content indicate either an interruption of the meta-linguistic tangent or an intention to signal a contrast with the negative affect of the

---

[2]For Passage P in Figure 4, participants may have linked the *in other words* clause to the intervening material itself.

tangent itself (e.g., "I don't know where…", "frustrating way of putting it", "how this is measured is unclear"). We speculate that the presence of BECAUSE in passages with intervening content may arise when that content implies that the situation is somehow surprising, which in turn merits explanation (e.g., "it's an UNUSUAL role for her", "their ability to actually work sensitively is perhaps QUESTIONABLE", "it's STRANGE to think of a planet being born"). These hypotheses will themselves need to be tested.

### 5.2 *Otherwise*: Inference from semantic features of segments

As noted in Section 4.2, passages containing *otherwise* were used to test how semantic properties of the segments themselves influenced conjunction choice. The categorization of passages by the researchers (16 ARGUMENTATION, 16 EXCEPTION, 16 ENUMERATION) predicts the conjunctions chosen by participants. In aggregate, ≈99% of responses to ARGUMENTATION passages were BECAUSE or OR or both. ≈92% of responses to EXCEPTION passages were BUT, AND, or both BUT and AND. And ≈98% of responses to ENUMERATION passages were BUT, AND, OR, or some subset thereof. For analysis, a mixed-effect logistic regression modeled the binary outcome of BUT insertion and showed significant variation across the three categories ($p < 0.001$). This measure captures the difference between pairs of categories: ARGUMENTATION permits BECAUSE and OR (hence BUT is rare) while ENUMERATION permits BUT and OR (hence BUT is present) and EXCEPTION favors BUT (hence BUT is very frequent). All pairwise comparisons yielded a main effect of category on this dependent measure (p's < 0.001).

Turning to individual passages, participant choices are shown in Figures 5–7. For ARGUMENTATION (Figure 5), the effect is uniformly strong, with all passages showing BECAUSE or OR as
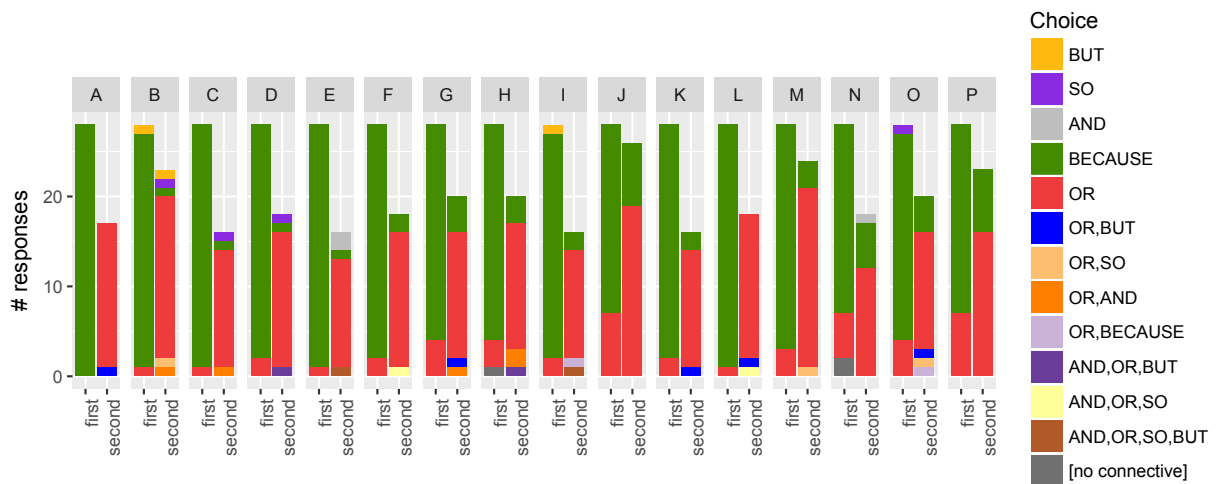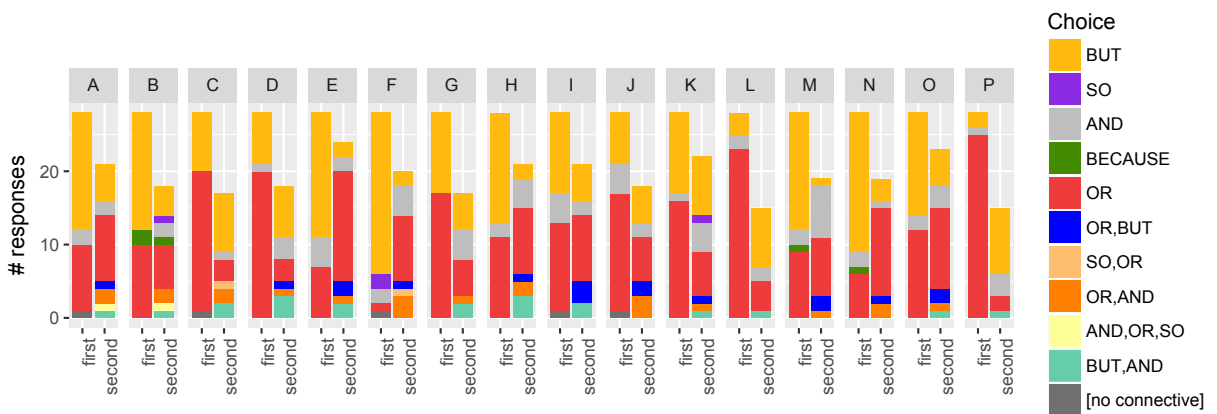
Figure 5: Distribution of first and second choice conjunctions for ARGUMENTATION *otherwise*. Labels such as "OR,BUT" are for multiple second choices. Each vertical bar represents a passage with the responses from each participant, color-coded by conjunction. Enlarged B/W versions of Figures 4–8 are available via the "notes" link on the paper in the ACL anthology.
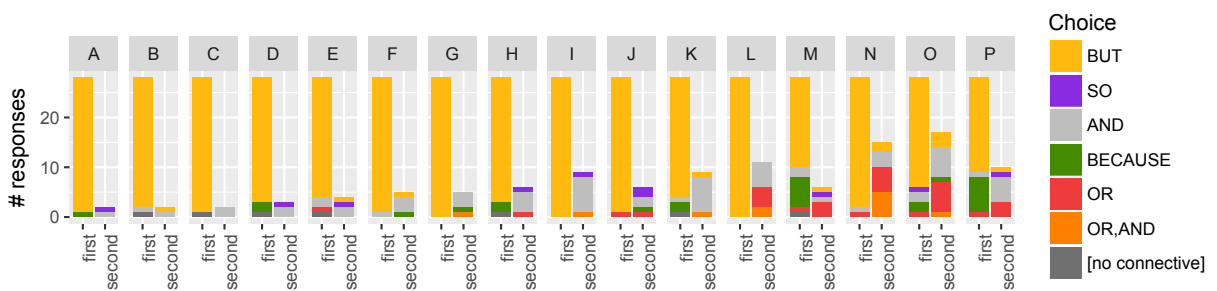
Figure 6: Distribution of first and second choice conjunctions for EXCEPTION *otherwise*. The label "OR,AND" in the legend implies both as second choices.

Figure 7: Distribution of first and second choice conjunctions for ENUMERATION *otherwise*. Labels in the legend such as "SO,OR" are for multiple second choices.
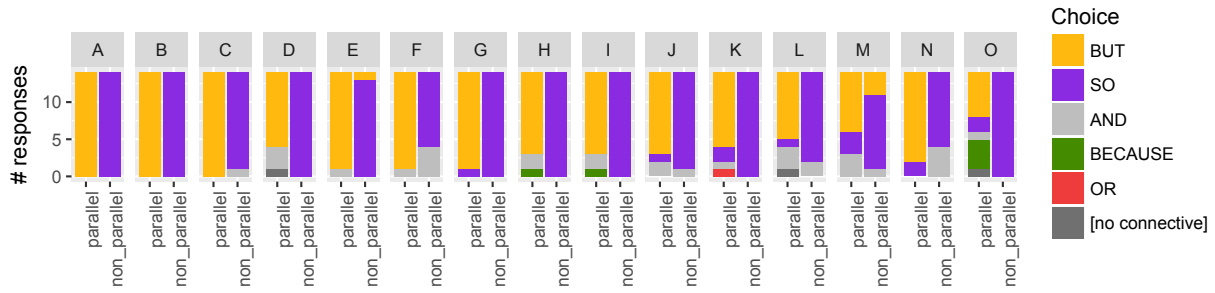
Figure 8: *Instead* passages, pairing a parallel variant and a causal variant. Each column shows the distribution of participants' first choice in the conjunction-insertion task. Each participant saw only one variant.

participants' top choice, with OR or BECAUSE chosen as equivalent (shown in the columns labelled "second"). For EXCEPTION (Figure 6), BUT is consistently the participants' top choice.

There are a few deviations from this near uniform endorsement of BUT for EXCEPTION (Figure 6, passages L–P). Any hypotheses, however, would require further experimentation to test. For example, in passage M (see (15)) and P (see (16)), participants rarely identified any conjunction as conveying the same sense as BUT. However, when their top choice was BECAUSE, they also selected OR as conveying the same sense. As noted above, BECAUSE and OR predominate with *otherwise* used in ARGUMENTATION. This raises the question of why passages M and P lead some participants to infer ARGUMENTATION and other participants, either EXCEPTION or ENUMERATION.

(15) Democrats insist that the poor should be the priority, and that tax relief should be directed at them _____ otherwise they lack a cogent vision of the needs of a new economy.

(16) He said that the proposed bill would give states more flexibility in deciding whether they wanted to use the Federal money for outright grants to municipalities or to set up loan programs _____ otherwise it left last fall's Congressional legislation unchanged.

Finally, though the pattern for ENUMERATION (Figure 7) is harder to see, combinations of BUT, OR and AND predominate as participants' top choices, with a few tokens of BECAUSE and SO, but too few to analyse as anything but noise.

The above results reflect researcher-assigned use labels. However, the confusion matrix in Table 1 shows that on the whole, participants agree with

that assignment. The column labelled Multiple is for cases where participants offered two paraphrases. For ARGUMENTATION, at least one paraphrase always corresponded to EXCEPTION, while for ENUMERATION, it did so for most of these tokens (9/14). We comment on this below.

While there was less agreement when participants offered multiple paraphrases for researcher-assigned EXCEPTION, there may be too few tokens here to draw any kind of conclusion. In any case, the results for ARGUMENTATION and ENUMERATION agree both across participants (in what paraphrase they choose when they don't choose the researcher-assigned label) and within participants (in what pairs of paraphrases they gave for the original passage).

The above results support our hypothesis that variability in participants' choice of conjunctions follows from both the lexical semantics of *otherwise* and the relation that participants infer between the segments in the passage.

5.3 *Instead*: Inference from a single manipulated property

On aggregate, participants responded very differently to the parallel and causal variants of *instead* passages (cf. Section 4.3). Figure 8 shows that in all cases, the parallel variant yielded more BUT responses, whereas the non-parallel (causal) variant yielded significantly more SO responses (main effect of (non-)parallelism: $\beta = -7.0008$, $p < 0.001$).[3]

Some of these results are very strong. For example, Passage A (17) drew all BUT responses for the

---

[3] We analyzed only 15 passages for *instead* and *after all*, due to a presentation error of the 16th for these adverbials.

| | **Participant** | | | |
Researcher	ARGUMENTATION	ENUMERATION	EXCEPTION	Multiple
ARGUMENTATION	401 (91.5%)	4	25	18
ENUMERATION	23	364 (81.4%)	46	14
EXCEPTION	21	29	393 (87.7%)	5

Table 1: Researcher labels assigned to *otherwise* passages vs. labels implied by participant paraphrases
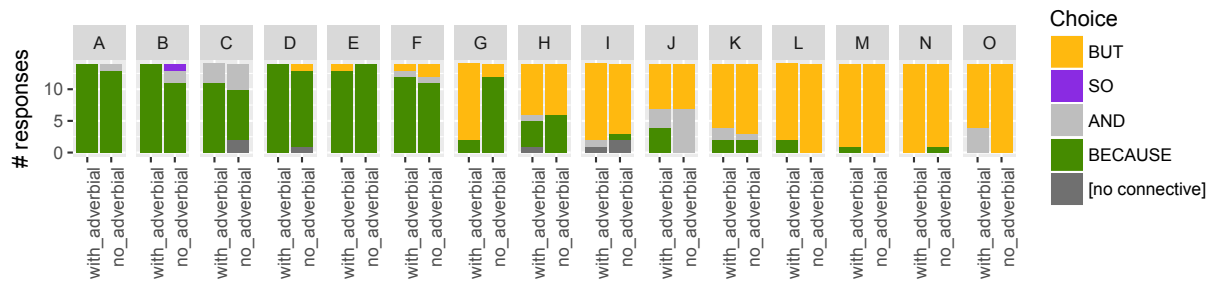
Figure 9: Distribution of first choice in conjunction selection task for passages with *after all*

parallel variant in (17a) and all SO responses for the causal variant in (17b), as did Passage B. In a few cases, however, the parallel variant drew variable responses, even while its causal variant drew strong SO responses. This is true of Passage O, with parallel and causal variants in (18a–b).

(17) a. They could have been playing football in the village green ＿＿＿ instead they played in the street.

   b. They didn't like playing football in the village green ＿＿＿ instead they played in the street.

(18) a. Smugglers nowadays don't use overland passages ＿＿＿ instead they use the seas to transport their goods.

   b. Smugglers' overland passages nowadays are too visible ＿＿＿ instead they use the seas to transport their goods.

One possible explanation is that participants varied in the role they assigned to the positive claim in the second segment of (18a) – either as a reason for the negative claim in the first segment (BECAUSE), as a contrast with that claim (BUT), or as its result (SO). Although manipulating the segment to enhance either parallelism or causality can change participant responses, it is clear that parallelism alone doesn't guarantee contrast.

5.4 *After all*: Adverb adds little to inference

Figure 9 shows participant choice of conjunction when *after all* is present and when it is absent. Their choice is largely the same for passages A–F and K–N, with and without the adverbial. As for passage O, since AND can contingently substitute for BUT (Knott, 1996), the response pattern can be considered the same as well. A by-passage correlation between the rate of BUT and BECAUSE responses across the two conditions confirms this similarity (R^2=.70, F(1,13)=30.98, p<0.001). The outlier is passage G:

(19) There was a testy moment driving over the George Washington Bridge when the toll-taker charged him $24 for his truck and trailer ＿＿＿ after all it was New York.

With *after all*, the majority of participants chose BUT as best expressing how the two segments are connected, while without it, the majority chose BECAUSE. Whatever explanation we gave here would be pure speculation. We trust that the fact that the other 14 passages demonstrate the predicted effect provides sufficient evidence that splits in participant responses are not simply a result of the presence of a discourse adverbial.

6 Conclusion

While our previous work showed that multiple discourse relations can hold between two segments – relations at the same semantic level, simultaneously available to a reader – we provided no evidence as to what influences the particular relations that are taken to be available. Our current experiments have provided some such evidence. Specifically, we have shown that participant responses to systematically manipulated passages involving discourse adverbials can be explained in terms of both the lexical semantics of discourse adverbials and properties of the passages that contain them. As the conjunctions chosen by participants convey senses that differ from those of the discourse adverbials, we also provided evidence for the simultaneous availability of multiple coherence relations that arise from both explicit signals and inference. We hope the reader is now convinced that, in both psycholinguistic research on discourse coherence and computational work on discourse parsing, one needs to identify and examine evidence for coherence involving more than one discourse relation.

Acknowledgments

References

M.J. Cuenca and M.J. Marin. 2009. Co-occurrence of discourse markers in Catalan and Spanish oral narrative. *Journal of Pragmatics*, 41(5):899–914.

Jesse Dunietz, Lori Levin, and Jaime Carbonell. 2017. The BECauSE Corpus 2.0: Annotating causality and overlapping relations. In *Proceedings of the 11th Linguistic Annotation Workshop*, pages 95–104, Valencia, Spain.

Vanessa Wei Feng and Graeme Hirst. 2012. Text-level discourse parsing with rich linguistic features. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, pages 60–68, Jeju Island, Korea.

Bruce Fraser. 2013. Combinations of contrastive discourse markers in English. *International Review of Pragmatics*, 5:318–340.

Barbara Grosz and Candace Sidner. 1986. Attention, intention and the structure of discourse. *Computational Linguistics*, 12(3):175–204.

Michael Halliday and Ruqaiya Hasan. 1976. *Cohesion in English*. Longman.

Yangfeng Ji and Jacob Eisenstein. 2014. Representation learning for text-level discourse parsing. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, volume 1, pages 13–24, Baltimore, Maryland.

Alistair Knott. 1996. *A Data-driven Methodology for Motivating a Set of Coherence Relations*. Ph.D. dissertation, Department of Artificial Intelligence, University of Edinburgh.

Ziheng Lin, Hwee Tou Ng, and Min-Yen Kan. 2014. A PDTB-styled end-to-end discourse parser. *Natural Language Engineering*, pages 151–184.

William C. Mann, Christian M. I. M. Matthiessen, and Sandra A. Thompson. 1989. Rhetorical structure theory and text analysis. Technical Report ISI/RR-89-242, USC-ISI, Marina del Rey CA.

William C. Mann and Sandra A. Thompson. 1988. Rhetorical Structure Theory: Toward a functional theory of text organization. *Text*, 8(3):243–281.

Daniel Marcu. 2000. *The theory and practice of discourse parsing and summarization*. MIT Press.

Eleni Miltsakaki, Nikhil Dinesh, Rashmi Prasad, Aravind Joshi, and Bonnie Webber. 2005. Experiments on sense annotation and sense disambiguation of discourse connectives. In *Proceedings of the Fourth Workshop on Treebanks and Linguistic Theories (TLT'05)*, Barcelona, Spain.

Johanna Moore and Martha Pollack. 1992. A problem for RST: The need for multi-level discouse analysis. *Computational Linguistics*, 18(4):537–544.

Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi, and Bonnie Webber. 2008. The Penn Discourse TreeBank 2.0. In *Proceedings of the 6th International Conference on Language Resources and Evaluation*, pages 2961–2968, Marrakech, Morocco.

Rashmi Prasad, Bonnie Webber, and Aravind Joshi. 2014. Reflections on the Penn Discourse TreeBank, comparable corpora and complementary annotation. *Computational Linguistics*, 40(4):921–950.

Hannah Rohde, Anna Dickinson, Chris Clark, Annie Louis, and Bonnie Webber. 2015. Recovering discourse relations: Varying influence of discourse adverbials. In *Proceedings of the First Workshop on Linking Computational Models of Lexical, Sentential and Discourse-level Semantics*, pages 22–31, Lisbon, Portugal.

Hannah Rohde, Anna Dickinson, Nathan Schneider, Christopher Clark, Annie Louis, and Bonnie Webber. 2016. Filling in the blanks in understanding discourse adverbials: Consistency, conflict, and context-dependence in a crowdsourced elicitation task. In *Proceedings of the Tenth Linguistic Annotation Workshop (LAW-X)*, pages 49–58, Berlin, Germany.

Hannah Rohde, Anna Dickinson, Nathan Schneider, Annie Louis, and Bonnie Webber. 2017. Exploring substitutability through discourse adverbials and multiple judgments. In *Proceedings of the 12th International Conference on Computational Semantics (IWCS)*, Montpellier, France.

Merel Scholman and Vera Demberg. 2017. Crowdsourcing discourse interpretations: On the influence of context and the reliability of a connective insertion task. In *Proceedings of the 11th Linguistic Annotation Workshop*, pages 24–33, Valencia, Spain.

Manfred Stede. 2012. *Discourse Processing*. Morgan & Claypool Publishers.

Bonnie Webber. 2013. What excludes an alternative in coherence relations? In *Proceedings of the 10th International Conference on Computational Semantics*, Potsdam, Germany.

Bonnie Webber, Aravind Joshi, and Alistair Knott. 2000. The anaphoric nature of certain discourse connectives. In *Making Sense: From Lexeme to Discourse*, Groningen, The Netherlands.

Nianwen Xue, Hwee Tou Ng, Sameer Pradhan, Rashmi Prasad, Christopher Bryant, and Attapol T. Rutherford. 2015. The CoNLL-2015 Shared Task on Shallow Discourse Parsing. In *Proceedings of the Nineteenth Conference on Computational Language Learning: Shared Task*, pages 1–16, Beijing, China.

Nianwen Xue, Hwee Tou Ng, Sameer Pradhan, Attapol Rutherford, Bonnie Webber, Chuan Wang, and Hongmin Wang. 2016. CoNLL 2016 Shared Task

on Multilingual Shallow Discourse Parsing. In *Proceedings of the 20th Conference on Computational Natural Language Learning – Shared Task*, pages 1–19, Berlin, Germany.