But vs. *Although* under the microscope

Fatemeh Torabi Asr[1] and Vera Demberg[2]
[1]Indiana University, Bloomington (IN), USA
`fatorabi@indiana.edu`
[2]Saarland University, Saarbruecken, Germany
`vera@coli.uni-saarland.de`

Abstract

Previous experimental studies on concessive connectives have only looked at their local facilitating or predictive effect on discourse relation comprehension and have often viewed them as a class of discourse markers with similar effects. We look into the effect of two connectives, *but* and *although*, for inferring contrastive vs. concessive discourse relations to complement previous experimental work on causal inferences. An offline survey on AMTurk and an online eye-tracking-while-reading experiment are conducted to show that even between these two connectives, which mark the same set of relations, interpretations are biased. The bias is consistent with the distribution of the connective across discourse relations. This suggests that an account of discourse connective meaning based on probability distributions can better account for comprehension data than a classic categorical approach, or an approach where closely related connectives only have a core meaning and the rest of the interpretation comes from the discourse arguments.

Introduction

There exists a substantial body of research on the processing of discourse connectives like *because*, *but* or *nevertheless* and their role in facilitating the linking of text segments to one another (Kintsch & Van Dijk, 1978; Just & Carpenter, 1980; Millis & Just, 1994; Murray, 1995, 1997; Sanders & Noordman, 2000; Rohde & Horton, 2014), as well as a discourse connective's effect in helping comprehenders anticipate and integrate upcoming content (Köhne & Demberg, 2013; Xiang & Kuperberg, 2014; Drenhaus et al., 2014). Most studies have however compared connectives that substantially differ in meaning, e.g., additives vs. contrastives, or causals vs. concessives. The present study contributes to the existing body of experimental research as well as theoretical framing of the meaning of connectives by investigating the effect of two connectives, *but* and *although*, that can mark the same set of relations, but differ in their distribution of how frequently they are used to mark these relations.

A first question that this study addresses is therefore what effect these connectives have on the meaning of the overall discourse. Do they affect the interpretation of the arguments themselves (instead of just signaling the relation), and if yes, how can this be accounted for? We already know that connectives like *because* may trigger causal inferences, which would not be made solely based on the meaning of the related clauses (Noordman & Vonk, 1992; Millis et al., 1995; Traxler et al., 1997).

Causal connectives like *because* have been argued to carry truth-conditional meaning, i.e., to extend to the set of statements entailed by the sentences they combine (in the case of *A because B*, this additional statement would consist of *B being the cause of A*, on top of the facts asserted by the arguments, A and B separately). Some researchers distinguish between connectives such as *because* and *before* that affect the truth-conditional state of what is said, and connectives regarded as non-truth conditional, such as *but* and *furthermore*.

Note though that non-truth-conditional connectives like *although* can trigger implications (Blakemore, 2002). For example, a sentence in the experiment run by Noordman & Vonk (1992), "Chlorine compounds make good propellants because they react with almost no other substances.", entails that *propellants must not combine with other substances*. A connective like *although* (instead of *because*) in the same sentence would imply that *propellants should combine with some substances*. [1] Our study investigates the effects of the connectives *but* and *although* on interpretation of the two arguments of a discourse relation, and discusses the resulting effects on the integration of upcoming content.

A second important point of this study is addressing the ambiguity of connectives *but* and *although*. We know from large discourse relation annotated corpora such as the PDTB (Prasad et al., 2008) that almost all discourse connectives are ambiguous (Asr & Demberg, 2012, 2013): while some connectives such as *since* are ambiguous between highly different relations (temporal vs. causal), others, like *and* can be present with almost any discourse relation. Existing proposals for describing the meaning of the discourse connectives *but* and *although* have considered sentence pairs such as (1), which seem to be fairly similar in meaning. Fraser (1998) suggests to describe each connective in terms of its *core meaning*, with further effects of inference being attributed to the discourse arguments. This approach boils down to assignment of an under-specified meaning to both connectives.

(1)   She fried the onions, **but / although** she steamed the cabbage.

---

[1] Interested readers are referred to Grice (1975) and Bach (2006) for detailed discussions on entailment vs. implication (implicature).

Alternatively, the meaning of a connective can be described based on its usage, i.e., extending distributional semantics accounts to discourse connectives, such that each connective potentially has more than a single meaning, and inferences are subject to the probability distribution over the meanings.

The goals of the present paper are

- to better understand the effect of the connectives *although* and *but* on the interpretation of a text, and

- to test whether and how the distribution of different meanings of a connective affects comprehension.

We address these questions by a corpus study on the distribution of discourse relations with these connectives in PDTB as well as an offline coherence judgment task. Other experiments have shown that online and offline effects don't necessarily agree, due to shallow processing / lazy inferences during online comprehension (Noordman & Vonk, 1992). We therefore follow up the coherence judgment task with a reading study using eye-tracking.

Discourse relations marked by *but* and *although*

A first important difference to note between the connectives *but* and *although* is of course that they differ in syntax (*although* is subordinating while *but* is a coordinating connective). Additionally, they have been suggested to be subject to a *semantic asymmetry* (Blühdorn, 2008), which means that one argument is more salient or central to the discourse than the other one.

Previous work investigating the relations marked by the connectives *but* and *although* (e.g., König, 1991; Blakemore, 2002; Hall, 2004; Iten, 2000) has used different names to refer to the relevant discourse relations; in today's most well-known annotation schemes, the relation most typically expressed by *but* is known as "contrast", "antithesis" or "negative additive", depending on the scheme, while the relation most typically marked by *although* is variously known as "concessive", "negative causal", "concessive.expectation" or "violated expectation". We will therefore briefly define the relations we are interested in here, and will then proceed to a corpus analysis to assess their frequency of occurrence. A comparison between the interpretation of sentence pair (2-a) with that of (2-b) elaborates the inferences in which we are interested.

(2) a. **Although** she desired to have something savory with her drink, she took some cake from the fridge.
     b. She took some cake from the fridge, **but** she desired to have something savory with her drink.

c. She took some cake from the fridge, **although** she desired to have something savory with her drink.

In (2-a), the first clause states that she desired something savory, which gives rise to the expectation that she'll take something savory from the fridge. The second clause however contrasts with this expectation, stating that she takes something sweet (cake). This segment can hence be classified as a `violated expectation` relation.

Definition: A `violated expectation` relation holds between two discourse segments wherever a discourse connective indicates that one of its arguments describes a situation *A* which causes *C*, while the other asserts or implies a state *C′* that contrasts with *C*. In other words, one argument of a `violated expectation` relation denotes a fact that triggers a set of potential consequences, while the other argument denies one or more of them.

In (2-b) on the other hand, a different reading is possible, where she takes the cake from the fridge, and then realizes that she wants something savory. In this case, there is no direct causal relation between the wanting something savory and taking something from the fridge. Hence, we analyze it as a `contrast` relation.

Definition: A `contrast` relation holds between two discourse segments when their arguments A and C contrast with one another in one or more respects. The difference to `violated expectation` relations is that there is no straightforward causal relationship between A and a (negated) C.

Finally, consider Example (2-c). Our hypothesis is that in this case, both interpretations are possible, i.e., the relation is ambiguous between a `violated expectation` and a `contrast` relation. Our experiment hence compares interpretations of sentences like (2-b) with a preferred `contrast` interpretation to ambiguous sentences like in (2-c).

Corpus study: PDTB

The Penn Discourse Treebank contains annotation of discourse relations in newspaper text for about one hundred connective types, including *but* and *although*. `Violated expectation` relations of the kind shown in (2-a) are annotated as COMPARISON.Concession.expectation relations in PDTB, see examples (3) and (4) below.

`Contrast` relations like the one in (2-b) are annotated as COMPARISON.Contrast relations in PDTB. They differ from Concession relations in that no (denied) direct causality is present between the arguments, see examples (5) and (6).

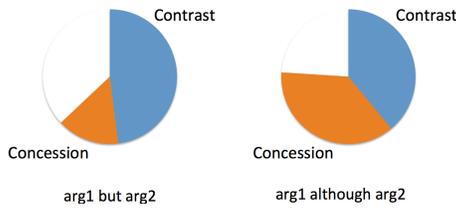(3) You might find something, but the chances are low.   (WSJ 21 54)

Figure 1: Distribution of discourse relations senses for *but* and *although* in PDTB.

(4) Oil prices haven't declined although supply has been increasing. (WSJ 02 31)

(5) He was on the board of an insurance company with financial problems, but he insists he made no secret of it. (WSJ 00 41)

(6) She didn't elaborate, although earlier US trade reports have complained of videocassette piracy in Malaysia [....] (WSJ 00 20)

We find that *but* and *although* frequently occur as markers in both COMPARISON.Contrast and COMPARISON.Concession relations. However, their distribution is different: *but* most frequently marks COMPARISON.Contrast relations while *although* most frequently marks COMPARISON.Concession relations. A closer look at the sentence initial vs. medial usage of *although* reveals that the distribution of discourse relations for sentence-initial vs. sentence-medial *although* also differs. In particular, the sentence medial use of *although* is divided half-half between Contrast and Concession (see Figure 1)[2]. Based on the corpus study, we hence predict that comprehenders will predominantly make a contrastive inference in the presence of *but*, but that both a contrastive and a violated expectation relation may be inferred when *although* in sentence-medial position is used.

Experiment 1

Our first experiment aims to investigate the difference between *but* and *although* in terms of their biases towards `contrast` and `violated expectation` inferences in an offline text comprehension setup.

Design and stimuli

We design short narrative texts like (7) embedding a discourse relation marked by *but* and *although*. Introduction and continuation are kept identical across condi-

---

[2]COMPARISON.Concession has another subtype in PDTB, called contra-expectation, which differs from the expectation subtype in terms of which of the arguments is the one that creates the expectation, and which one denies it. We found many similar instances of contra-expectation relations to Contrast, therefore, looked at the proportion of expectation vs. contra-expectations in the corpus (for *but* vs. *although*) as well and found very similar proportions to those illustrated in Figure 1.

tions. Context is changed by alternating *cake/pizza* and *savory/sweet*. The final sentence is designed to disambiguate between the two alternative discourse relations: *eating pizza* will be consistent with a `contrast` interpretation in a setting where Mary took some cake from the fridge (8a) but then realized she wants something savory. In a `violated expectation` interpretation, *eating pizza* is consistent with taking pizza from the fridge (8b), despite having originally wanted something savory.

(7) **Introduction:** Mary was feeling tired and hungry when she came home yesterday evening.

    a. She took some *cake* from the fridge, **but/although** she desired to have something *savory* with her drink.

    b. She took some *pizza* from the fridge, **but/although** she desired to have something *sweet* with her drink.

Disambiguating sentence: She had a piece of *pizza* and went to bed earlier than usual.

In this experiment, people are asked to judge the coherence of the entire story after reading it carefully (no time pressure). We expect *but* to cause a strong bias for a `contrast` interpretation. Therefore, condition (a) with *but* should prepare the reader for accepting the continuation *eating pizza*, whereas condition (b) with this connective should result in a contradiction at the end of the story, thus incoherence. On the other hand, *although* should be more ambiguous, i.e., trigger either a `contrast` or a `violated expectation` inference. We expect the average coherence rating of the two *although* conditions to be more similar if the interpretation depends on learned prior usage in text.

As a pretest of the stimuli, we included four additional conditions that exclude the final disambiguating sentence, to make sure that the coherence of the texts up to the disambiguating sentence is matched. Thus, a total of 8 conditions: 2 (connective) * 2 (context) * 2 (with/without final sentence) are constructed for 24 items.

Procedure

We recruited 48 native speakers of English (25 female and 23 male aged between 22-68) on Amazon Mechanical Turk. Eight different lists of 24 stories in equally distributed conditions mixed with filler items were published as HITs. The worker were only allowed to do a single HIT (i.e., one list) of the experiment, so that nobody would see two conditions of the same item. Subjects scored the coherence of each story on a Likert scale from 1 (incoherent) to 7 (perfectly coherent). A compensation of \$2.5 was paid per HIT.

Data treatment

In total 2400 samples (48 participant * 50 items including 26 fillers) were collected. Only 7 samples were left unanswered in total. Coherence judgment scores for the obviously incoherent and coherent filler items were checked to make sure that the participants understood the task and provided a sensible rating.

Results

Pretest (short versions): Coherence scores assigned to the stories excluding the final sentence were examined first as a pretest of local coherence. According to Anova and mixed-effect regression (with participant and item as random factors, and connective and context as fixed effects), no significant difference is observed between the local coherence of the stories containing either of the two connectives (with an mean rating of 5.14 for *although* and 5.34 for *but*). This shows that there was no difference in coherence between the usage of the connectives in our experimental items. Any difference in coherence judgments for the full texts can therefore be attributed to the final sentence which disambiguates the coherence relation to a `violated expectation` or `contrast` relation.

Main results (full stories): Table 1 presents the average ratings and standard deviations obtained for the complete versions of stories from different conditions.

Table 1: Coherence scores by context & connective

Condition	Mean score	SD
Contrast:but	**5.38**	1.71
Contrast:although	4.85	1.86
ViolExp:but	**3.31**	1.80
ViolExp:although	4.52	1.91

We fitted a linear mixed effects model with random intercepts and slopes for participant and item, using forward selection on fixed effects and backward selection on random slopes (in case the full model didn't converge). The final model (including random slopes for item and participant under connective and relation, as well as the interaction of connective and relation) shows a significant negative main effect of the connective *but* (reflecting the very low coherence judgments for the connective *but* in a `violated expectation` relation), as well as a significant interaction between connective and discourse relation. The presence of the interaction confirms that the two connectives affect interpretation of the discourse differently. Among the *but* conditions, stories with a last sentence consistent with a `violated expectation` inference were rated to be significantly less coherent than stories with a continuation consistent with a `contrast` inference. For texts including the connective *although*, completions consistent with `violated`

`expectation` inferences and `contrast` inferences were both judged to be equally coherent. This finding is in line with our hypothesis, as it directly reflects the distribution of discourse relations that were observed in the corpus.

Experiment 2

The above results show that interpretation of readers wrt. connected sentences in short stories are affected by the fine-grained inferences triggered by the specific discourse connective. The results of the coherence judgments however do not necessarily imply that these inferences are drawn also during natural reading that doesn't include a task focussing on coherence judgments. In this section, we investigate whether readers pick up on the difference between *but* and *although* during online reading, and show effects of coherence on reading times of the final sentence.

Design and stimuli

Stimuli in the eye-tracking experiment are similar to (7), except we test each item also with an alternative final sentence to achieve a fully counterbalanced design. We hence have 4 conditions in this experiment: 2 (connective) * 2 (context). Disambiguating sentence B in (8) has the same function for conditions (8b) as disambiguating sentence A has for conditions (8a).

(8) **Disambig. sent. A:** She had a piece of **pizza** and went to bed earlier than usual.
    Disambig. sent. B: She had a piece of **cake** and went to bed earlier than usual.

Items are mixed with filler stories, as well as items of two other experiments with similar length and narrative content. Every participant read 84 stories (including 12 *but* items and 12 *although* items) and answered a YES/NO comprehension question about the introduction part of the stories (not the part depending on the variable interpretation), e.g., "Was Mary at home the entire day yesterday?". The purpose of the questions was to make sure subjects read for comprehension, and to analyze the correlation between coherence of the story and the response time and correctness of the answers. Fillers had questions from all different parts of the text. For our analysis, we collected the total reading time of a story, response time to the question, and the reading time of a critical region in the disambiguating final sentence. The critical region (the word **cake/pizza**) is highlighted in example (8).

Procedure

The eye-tracking experiment was implemented within the Experiment Builder software for an EyeLink 2000 tracker, tracking at 500Hz on both eyes. All text material on the screen were shown in `Lucida Console` font (with

Table 2: Answers to the comprehension questions

Condition	Mean %correct	SD
Contrast:but	**0.85**	0.36
Contrast:although	0.83	0.38
ViolExp:but	**0.79**	0.41
ViolExp:although	0.81	0.40

Table 3: Critical region total RT

Condition	Mean RT	SD
Contrast:but	**296.28**	235.46
Contrast:although	331.90	346.72
ViolExp:but	**332.92**	266.05
ViolExp:although	328.48	263.06

same length characters), size 20 and triple line spacing. Subjects were asked to press the space key after reading a story to navigate to the question screen, and press J and F keys for YES and NO answers, respectively.

A total of 39 native English speakers were recruited for the experiment at the University of Edinburgh and received a compensation of 12 pounds for a two hour session. The eye-tracking experiment was followed by a standard memory test, to measure subjects' memory spans. People's memory span size can play a confound role in the correctness of their answers to the comprehension questions and also might affect reading patterns, e.g., when a sentence contradicts with a non-immediate but related sentence in the preceding context.

Data treatment

We had to discard data from 7 subjects because of frequent head movement, blinks or longer track losses during the experiment. The below results are thus based on a final set of 32 subjects. Standard outlier removal process was performed on the fixations before the RT analysis.

Results

Question answering correctness: Participants' question answering accuracies varied between 50% to 100% correct answers with mean and median of 82%. We observed that subjects with larger memory span size had better question answering performance ($p < 0.05$). Table 2 shows the proportion of correct answers to the comprehension questions across coherence conditions. A trend compatible with the results of the offline study can be observed. However, fitting a mixed-effect regression with all factors (connective, context and participant's memory span) as fixed effects plus participant and item as random effects only revealed a significant main effect of the memory span size. The best fit obtained through a forward model selection procedure showed only a marginal effect of the coherence condition for the subset of data including *but* ($p < 0.1$). This suggests that the coherence of the story as a whole (in terms of the congruence of the final sentence with the interpretation of the middle part) only slightly affected people's recall of the story. Total reading times of the stories and the questions did not correlate with correctness of the answers.

Interest area RT: Table 3 compares the total reading time of the critical area in the final sentence of the sto-

ries for every coherence condition. The critical area was chosen to be the only word at which the conditions differ, and which resolves the interpretation of the discourse relation expressed by *but / although*. That is, the word *pizza* vs. *cake* in (9) disambiguates whether the text segment *she desired to have something sweet with her drink* relates to the first segment by a `violated expectation` or `contrast` discourse relation.

Reading times on the critical region are consistent with corpus statistics and the offline study: *but* has a bias towards a `contrast` inference, hence the continuation consistent with this interpretation is processed faster than one confirming a `violated expectation` interpretation. For *although*, the critical word is read at a similar speed independent of whether it disambiguates in favour of a `contrast` or `violated expectation` relation.

We fitted a linear mixed effects regression model for total reading times on the critical region. While a complete model with random slopes for item and participant did not converge, the simpler model (excluding random slopes but including random intercepts) reveals an marginally significant interaction between discourse connective and discourse relation in total reading time of the critical area ($p < 0.1$). There is also a significant effect for this interaction in regressions out of this area to the preceding context ($p < 0.05$). This effect is consistent with the experimental result from the coherence judgment study. The two *but* conditions also differ in terms of the total reading time ($p < 0.1$) and regression-in ($p < 0.05$) at the area preceding the critical phrase in Arg2. Increased regression-in is also observed at the onset of Arg1 for the less coherent *but* condition ($p < 0.1$). Regressions from the final sentence to the previous context indicate that the subjects re-read the areas triggering the relational inference, after they encountered the unexpected continuation. The equivalent complete models to fit first-pass and second-pass duration measures did not converge.

Conclusion

We examined the very fine-grained inferences triggered by two discourse connectives *but* and *although*. Our findings show that these connectives trigger linguistic inferences (despite not being truth-conditional). Each connective in our experiment had a different bias for a specific semantic interpretation tested by a disambiguat-

ing statement in the final sentence, which would sound more or less coherent depending on that interpretation. Our results put into question the traditional perspective that generalizes an effect to a category of connectives, e.g., causal/additive/adversative in previous experimental studies. The probability distribution of a connective in the natural occurrences of various discourse relations (as we took from PDTB) seems to be a more accurate meaning representation for approximation of its effect on offline and online comprehension. This finding motivates future research on comprehension and production of discourse connectives as probability sensitive processes and the interaction between the two. Relevant open questions are why and how different distributional profiles emerge in a given language for closely related discourse connectives, and how such diversities can be explained from a cognitive perspective across languages.

References

Asr, F. T., & Demberg, V. (2012). Measuring the strength of linguistic cues for discourse relations. In *Proceedings of the Workshop on the Advances in Discourse Analysis and its Computational Aspects.*

Asr, F. T., & Demberg, V. (2013). On the information conveyed by discourse markers. In *Proceedings of the Fourth Annual Workshop on Cognitive Modeling and Computational Linguistics* (pp. 84–93).

Bach, K. (2006). The top 10 misconceptions about implicature. *Drawing the boundaries of meaning: Neo-Gricean studies in pragmatics and semantics in honor of Laurence R. Horn*, 21–30.

Blakemore, D. (2002). *Relevance and linguistic meaning: The semantics and pragmatics of discourse markers*. Cambridge University Press.

Blühdorn, H. (2008). Subordination and coordination in syntax, semantics and discourse. *Subordination versus Coordination in Sentence and Text. A Cross-Linguistic Perspective, Amsterdam*, 59–85.

Drenhaus, H., Demberg, V., Köhne, J., & Delogu, F. (2014). Incremental and predictive discourse processing based on causal and concessive discourse markers: ERP studies on German and English. In *Proceedings of the annual conference of the cognitive science society.*

Fraser, B. (1998). Contrastive discourse markers in english. *Pragmatics and Beyound New Series*, 301–326.

Grice, H. P. (1975). Logic and conversation. *Reprinted in Studies in the Way of Words 1985*, 2240.

Hall, A. (2004). The meaning of but: A procedural re-analysis. *UCL Working Papers in Linguistics*, *16*, 199–236.

Iten, C. (2000). Although revisited. *Working Papers in Linguistics*, *12*.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: from eye fixations to comprehension. *Psychological review*, *87*(4), 329.

Kintsch, W., & Van Dijk, T. (1978). Toward a model of text comprehension and production. *Psychological review*, *85*(5), 363.

Köhne, J., & Demberg, V. (2013). The time-course of processing discourse connectives. In *Proceedings of the 35th annual meeting of the cognitive science society.*

König, E. (1991). Concessive relations as the dual of causal relations. *Semantic Universals and Universal Semantics. Groningen-Amsterdam Studies in Semantics*, *12*, 190–209.

Millis, K., Golding, J., & Barker, G. (1995). Causal connectives increase inference generation. *Discourse Processes*, *20*(1), 29–49.

Millis, K., & Just, M. (1994). The influence of connectives on sentence comprehension. *Journal of Memory and Language*.

Murray, J. (1995). Logical connectives and local coherence. *Sources of Coherence in Reading*, 107–125.

Murray, J. (1997). Connectives and narrative text: The role of continuity. *Memory and Cognition*, *25*(2), 227–236.

Noordman, L. G., & Vonk, W. (1992). Readers' knowledge and the control of inferences in reading. *Language and Cognitive Processes*, *7*(3-4), 373–391.

Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A., & Webber, B. (2008). The Penn Discourse Treebank 2.0. In *Proceedings of the 6th international conference on language resources and evaluation* (pp. 2961–2968).

Rohde, H., & Horton, W. S. (2014). Anticipatory looks reveal expectations about discourse relations. *Cognition*, *133*(3), 667–691.

Sanders, T. J., & Noordman, L. G. (2000). The role of coherence relations and their linguistic markers in text processing. *Discourse processes*, *29*(1), 37–60.

Traxler, M., Sanford, A., Aked, J., & Moxey, L. (1997). Processing causal and diagnostic statements in discourse. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(1), 88.

Xiang, M., & Kuperberg, G. (2014). Reversing expectations during discourse comprehension. *Language, Cognition and Neuroscience*, 1–25.