"On the one hand" as a cue to anticipate upcoming discourse structure

Merel C.J. Scholman[*,a], Hannah Rohde[b], Vera Demberg[a,c]

[a] *Saarland University, Department of Language Science and Technology, Saarbrücken, Germany*
[b] *University of Edinburgh, Department of Linguistics & English Language, Edinburgh, UK*
[c] *Saarland University, Department of Computer Science, Saarbrücken, Germany*

Abstract

Research has shown that people anticipate upcoming linguistic content, but most work to date has focused on relatively short-range expectation-driven processes within the current sentence or between adjacent sentences. We use the discourse marker *On the one hand* to test whether comprehenders maintain expectations regarding upcoming content in discourse representations that span multiple sentences. Three experiments show that comprehenders anticipate more than just *On the other hand*; rather, they keep track of embedded constituents and establish non-local dependencies. Our results show that comprehenders disprefer a subsequent contrast marked with *On the other hand* when a passage has already provided intervening content that establishes an appropriate contrast with *On the one hand*. Furthermore, comprehenders maintain their expectation for an upcoming contrast across intervening material, even if the embedded constituent itself contains contrast. The results are taken to support expectation-driven models of processing in which comprehenders posit and maintain structural representations of discourse structure.

Key words: Discourse structure, anticipation, coherence relations

*Corresponding author. *E-mail address:* m.c.j.scholman@coli.uni-saarland.de

Introduction

There is a growing body of research suggesting that readers and listeners make predictions about upcoming sounds, words and syntactic structures based on information that they have encountered so far (see Kuperberg & Jaeger, 2016 for a review). However, this prior work has focused largely on anticipation of relationships within the sentence (e.g. Altmann & Kamide, 1999; Arai & Keller, 2013; Clifton, Frazier & Connine, 1984; DeLong, Urbach & Kutas, 2005; Federmeier & Kutas, 1999; Staub & Clifton, 2006; Yoshida, Dickey & Sturt, 2013) or locally between adjacent sentences (Arnold, 1998; Van Berkum, Brown, Zwitserlood, Kooijman & Hagoort, 2005; Drenhaus, Demberg, Köhne & Delogu, 2014; Kehler, Kertz, Rohde & Elman, 2008; Rohde & Horton, 2014; Xiang & Kuperberg, 2015). Across multi-sentence passages, the possible relationships to be established are much more flexible than those afforded by sentence-internal phonological, syntactic, and lexical-semantic constraints. Given this flexibility, an open question is whether and how comprehenders manage expectations regarding cross-sentence relationships.

The cross-sentence relationships we focus on here are the discourse coherence relations that supply semantic links between propositions (Hobbs, 1979; Kehler, 2002; Mann & Thompson, 1988; Sanders, Spooren & Noordman, 1992). These relations bind together individual propositions and multi-sentence sequences into a logical whole. Examples include, among many others, relations that establish contrast (e.g., relations marked with *but* or *on the other hand*).

Coherence relations are posited to hold both locally between adjacent sentences and remotely across intervening clauses, as well as to combine recursively across greater distances to yield larger discourse structures with embedded constituents and long-distance dependencies (as shown empirically in corpus annotations, e.g., Rhetorical Structure Theory Discourse Treebank: Carlson, Marcu & Okurowski, 2003; Penn Discourse Treebank: Prasad, Dinesh, Lee, Miltsakaki, Robaldo, Joshi & Webber, 2008; Discourse Graphbank: Wolf & Gibson, 2005). Very few constraints dictate the nature of the growing discourse structure such

as where or how subsequent utterances will attach (cf. Polanyi, 1988). This raises the question of whether comprehenders anticipate how upcoming coherence relations will link future sentences to the prior discourse, and whether they do so across multi-sentence passages in which the growing structure may be very open-ended.

A starting place to address this question is the presence of linguistic cues that constrain possible upcoming discourse structures. Consider (1).

(1) John is considering a job at the Edinburgh zoo.
On the one hand, he really needs the money, because he should start paying off his student loans.
On the other hand, he hates the idea of cleaning out panda cages.

The first sentence in (1) introduces an issue that is up for debate. The next two sentences present two contrasting perspectives, marked respectively with *On the one hand* (OT1H) and *On the other hand* (OTOH). These two perspectives, contrast1 and contrast2, expand on the issue in the context sentence and are linked to each other via the overt OT1H/OTOH markers. The marker OT1H signals to the comprehender to expect an upcoming contrast2, an expectation which typically must be satisfied for the passage to constitute a felicitous discourse. What is interesting about OT1H as a cue is that, though strongly constraining, it does not fully determine the nature of the next sentence. For instance, contrast2 may be signalled by some marker other than OTOH (e.g., *but* or *however*). Likewise, contrast1 itself need not be marked with OT1H, as evidenced by the acceptability of (1) without the overt OT1H marker. Furthermore, the material following contrast1 may contain content that does not support the inference that contrast2 is being conveyed, as is the case with the *Also* sentence in (2), which intervenes before the OTOH-marked contrast2.

(2) John is considering a job at the Edinburgh zoo.
On the one hand, he really needs the money, because he should start paying off his student loans.

Also, his car needs to be serviced.

On the other hand, he hates the idea of cleaning out panda cages.

To understand the passage in (2), a comprehender must build a sufficiently rich discourse structure to link the OT1H and OTOH sentences even though they are not adjacent and to establish that the *Also* sentence participates in the resulting discourse structure as part of the constituent conveying the contrast1 perspective. An intervening sentence could itself convey some type of contrast without specifically contrasting with the OT1H proposition—e.g., a local contrast with the immediately preceding embedded material like the student loans in (2). As such, OT1H/OTOH contexts allow us to test the generation and maintenance of discourse-level expectations across multiple sentences. This extends prior work on the processing of discourse markers (Canestrelli, Mak & Sanders, 2013; Drenhaus, Demberg, Köhne & Delogu, 2014; Xiang & Kuperberg, 2015; Xu, Jiang & Zhou, 2015) and coherence relations (Dery & Koenig, 2015; Kehler, Kertz, Rohde & Elman, 2008; Köhne & Demberg, 2013; Mak & Sanders, 2013; Rohde, Levy & Kehler, 2011; Rohde & Horton, 2014) by moving beyond relationships that hold between adjacent sentences and considering instead how specific manipulable cues can guide larger scale structure building during processing.

The studies presented here test whether comprehenders can use a discourse marker to anticipate an upcoming coherence relation across intervening material and, if so, how fine-grained their expectations are regarding the discourse structure. One might imagine, given the surface form of OT1H and the disambiguating power of OTOH to establish the OT1H/OTOH link, that comprehenders need only anticipate the upcoming OTOH form, which triggers the establishment of the contrast1∼contrast2 pairing. Why posit discourse structure before the relevant pieces are even available? Indeed, some processing accounts assume that comprehenders wait to hear complete propositions before initiating discourse-level processing (Garnham, Traxler, Oakhill & Gernsbacher, 1996; Stewart, Pickering & Sanford, 2000). Alternatively, under a model of processing

in which comprehenders build, maintain, and update rich discourse structures (Kuperberg, 2016), OT1H may operate as a cue to signal that the upcoming discourse will contain contrast2. In that case, contrast2 must be linked to contrast1 even if it is unmarked, non-adjacent, or if additional material intervenes which conveys another type of (e.g., locally attaching) contrast, as noted for OT1H/OTOH in theoretical work by Cristea and Webber (1997).

The paper is laid out as follows. The next section reviews related work on cue-based anticipation and the processing of discourse structure. Section then describes the design of the materials used in our experiments. The rest of the paper presents three studies. Study 1 is a story acceptability study, and functions as a norming study for our materials. Study 2 is a story continuation study, designed to test comprehenders' anticipation regarding the content of the next sentence. Study 3, an eye-tracking while reading experiment, tests whether structure-sensitive biases emerge during online processing.

Background

That comprehenders can and must represent the structure of an unfolding discourse is taken for granted in theoretical models (Asher & Lascarides, 2003; Hobbs, 1979; Kehler, 2002; Mann & Thompson, 1988; Sanders et al., 1992) and in claims about the constraints on possible discourse structure (Asher & Vieu, 2005; Jasinskaja & Karagjosova, submitted; Polanyi, 1988). However, psycholinguistic evidence for comprehenders' awareness of multi-sentence discourse representations is limited and little is known about the types of representations that comprehenders use in real-time to understand a multi-sentence discourse as it unfolds. Existing work also does not address anticipatory processing to establish whether comprehenders are in fact making predictions about how a discourse will progress when the relevant dependencies may be non-local. The studies we review below provide the context for asking how comprehenders' anticipatory processes can be cued and what structures they may build to adequately represent a discourse.

Cue-based anticipation of discourse content

Existing work points to the role connectives play in guiding comprehenders' expectations about upcoming content. For example, coreference processing is shown to vary with the coherence relation that holds between sentences, both when signalled directly by a connective or when inferred in context (Kehler, Kertz, Rohde & Elman, 2008; Koornneef & Sanders, 2013; Stevenson, Knott, Oberlander & McDonald, 2000; Wolf, Gibson & Desmet, 2004). Likewise, processing of event descriptions varies depending on the connective: Xiang and Kuperberg (2015) report that the presence of the connective *even so* in passages like (3) can reverse comprehenders' expectations quickly, thereby making an unexpected event expected, i.e. celebrating after failing a test (see also Asr & Demberg, 2015; Drenhaus et al., 2014).

(3) Elizabeth has a history exam on Monday. She took the test and aced/failed it. (Even so,) She went home and celebrated wildly.

These results provide evidence that readers are sensitive to connectives, and that they can use them to make fine-grained and rapid predictions about upcoming content. The current study extends this work by investigating whether readers are able to make similar fine-grained predictions about structure, rather than content.

Cue-based anticipation of discourse structure

Not only the content of an upcoming sentence is constrained by available cues, but also the role that the upcoming sentence will play relative to the current sentence. Drenhaus et al. (2014) investigated the processing of stories marked by the concessive connective *however* or the causal connective *therefore*, and found that the connective itself yielded an immediate effect: Late positivities were elicited at *however* in comparison with *therefore*. This is interpreted as an updating process: Encountering a concessive connective switches

the comprehenders' anticipation of a causal relation to an (until then) unexpected concessive relation (see Xu, Jiang & Zhou, 2015, for similar results in Chinese). These results suggest that comprehenders are aware of the probability of different coherence relations, updating their expectations about the operative coherence relation when they encounter a connective that signals that an unexpected relation holds.

Relations within sentences can also be cued by paired markers, demonstrated by Staub and Clifton (2006) with the markers *either/or*. They find evidence for comprehenders' prediction of a coordination structure within sentences. In (4), the material after *or* can attach at two possible points: at the sentence-level or at the noun phrase (e.g., *Either Linda bought the red car or the green one.*).

(4) (Either) Linda bought the red car or her husband leased the green one.

Staub and Clifton found that participants read the *or*-clause faster when *either* was present, meaning that *either* cued them to expect an upcoming coordinated structure. In addition, there was evidence that readers misanalyzed the sentence coordination as noun-phrase coordination when *either* was absent, leading to longer reading times at the end of the sentence coordination condition (in comparison with a noun-phrase-only coordination condition: *The workers painted (either) the house or the barn over the summer*). The word *either* thus enabled readers to predict upcoming sentence coordination, which facilitated the processing of that structure when they encountered it.

Besides connectives, a number of other cues have been shown to bias comprehenders' expectations about the operative coherence relation. For example, implicit causality verbs have been shown to elicit expectations about who will be referred to next (see, for example, Ehrlich, 1980; Kehler et al., 2008; Koornneef & Van Berkum, 2006; Pyykkönen & Järvikivi, 2009; Stevenson et al., 2000; Wolf et al., 2004) and what relation will hold between the current and the upcoming sentence (Kehler et al., 2008). Using a novel eyetracking paradigm, Rohde and Horton (2014) find that readers anticipate the upcoming coherence relation immediately after the offset of an implicit causality verb. This indicates that cues

for coherence anticipation are integrated and deployed rapidly.

However, the demonstrated anticipatory effects just described are local—either within one sentence or between one sentence and an adjacent sentence. The new studies we report here target non-local dependencies to test how the anticipation of coherence relations scales to multi-sentence discourse. *On the one hand* is notable (Cristea & Webber, 1997) in establishing an expectation for a subsequent sentence that will express a relevant contrast, an expectation that crucially need not be satisfied immediately.

Structural sensitivity

The analysis of discourse structure—its categories and constraints—largely grew out of computational linguistic traditions (going back to researchers such as Hobbs, 1985), not the psycholinguistics literature. What has emerged are a number of models of discourse coherence (Asher & Lascarides, 2003; Hobbs, 1979; Kehler, 2002; Mann & Thompson, 1988; Prasad et al., 2008; Sanders et al., 1992), with some shared properties but notably few structural constraints. What all these theories share is that discourse structure is built up recursively via the combination of discourse units. These units combine via a number of different semantic links, drawn from an inventory of possible coherence relations (the size of which is debated; see Hovy & Maier, 1995). The recursive combination of discourse units yields a discourse structure. Different frameworks make different assumptions about such a structure; for example, certain theories represent the discourse structure as a hierarchical parse tree (cf. Asher & Lascarides, 2003; Polanyi, 1988), whereas others allow for more flexible structures (cf. Carlson et al., 2003; Wolf & Gibson, 2005). The focus of the current paper, however, is on whether readers are sensitive to non-local structural dependencies, rather than the ultimate parse structure for the entire discourse or the inventory of relations that are required for full coverage.

Even though theories of discourse coherence do make specific predictions about where a subsequent utterance can attach in the existing structure, such constraints have only recently been subject to experimental testing. The Right

Frontier Constraint states that, given a representation for a set of clauses in a passage, a subsequent clause can only create a new branch at an open node on the right edge of the discourse structure; other nodes are unavailable because they represent closed positions (Polanyi, 1988; see also Asher, 1993; Asher & Lascarides, 2003). The existence of this right frontier, depicted in Figure 1, was originally based on observations about the use of a pronoun in a subsequent clause and the pronoun's sensitivity to the status of certain constituents as open/closed. The observation was that pronouns find their antecedents more easily in open positions than closed positions.

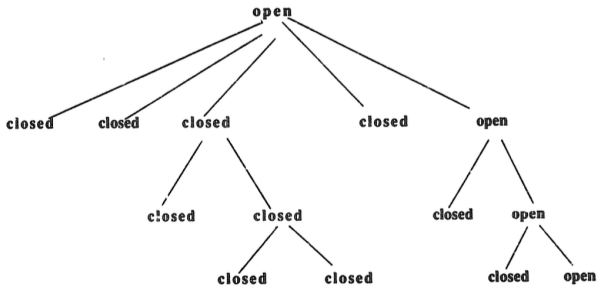

Figure 1: A sample discourse structure containing open positions at the right frontier (from Polanyi (1988), p.613)

In keeping with the Right Frontier Constraint, comprehenders indeed favor an antecedent at an open position, even when the linear distance to that antecedent was greater (Holler & Irmen, 2007). In more recent work, Kim (2015) uses presupposition to test for structural sensitivity and finds evidence that comprehenders favor interpretations that preserves discourse-level constituency, even when such a preference violates strict linear locality.

Lastly, comprehenders' show evidence of awareness of discourse structure in the way they answer questions about a structurally ambiguous passage like (5), which is compatible with a subordinating interpretation (whereby reading and watching are elaborations of what happened in the history class) or a coordinating interpretation (with three separate events).

(5) I sat in on a history class. I read about housing prices. And I watched

a cool documentary.

Tyler (2014) found that prosody influenced the structure that participants inferred: Interpretations of subordinating structures were reduced when the first sentence was uttered with a final pitch rise.

However, these prior studies on structural sensitivity do not establish how fine-grained comprehenders' discourse representations can be, whether they can be constructed in real time, and whether comprehenders use cues to anticipate their construction. The new studies presented here aim to address all three of these points using the markers "on the one hand" and "on the other hand".

Occurrence of OT1H and OTOH in natural text

In order to better understand the degree to which OT1H is predictive of OTOH, we investigated the occurrence of the markers in natural text. We extracted all fragments containing one or both markers from the ukWaC corpus, a 2 billion-word corpus of English UK webpages (Baroni, Bernardini, Ferraresi & Zanchetta, 2009). For every occurrence of OT1H, the following three paragraphs were searched for OTOH. Similarly, for every occurrence of OTOH, the preceding three paragraphs were searched for OT1H. This amounted to 60,749 instances containing one or both markers. The passages reveal several features of OT1H and OTOH that show why they are suitable for our experimental aims.

The appearance of OT1H or OTOH does not wholly determine the presence of the other: only 18% of passages contain both OT1H and OTOH, and in 3% of the data OT1H occurs without OTOH. In the latter set, OTOH was replaced by other connectives and cue phrases (most commonly *but*, *at the same time* and *while*), suggesting that natural text often requires readers to link contrast2 to contrast1 even without the OT1H/OTOH pairing.

In 7% of the OT1H-marked data, other sentences intervene between the OT1H-sentence and the OTOH-sentence. Importantly, this intervening material can itself contain discourse relations (see Appendix A for sample data). Contrastive markers also occurred between OT1H and OTOH. For example,

but occurred between the pair in more than 500 passages. *But* is therefore an ambiguous marker when it follows OT1H: it can mark contrast2, as well as other contrastive relations that are presumably embedded within contrast1. The structure requires comprehenders to process an embedded discourse relation, while also maintaining a prediction for contrast2.

This corpus investigation hence attests to the frequent presence in naturally occurring contexts of the types of complex discourse structures that we will target in the subsequent experiments presented in the next sections.

Experimental design and predictions

We use contexts like (6), varying the type of additional material intervening before OTOH and the presence/absence of OTOH-marked contrast2.

(6) **Intro:** Joseph got a job offer from the Edinburgh Zoo and he's pondering whether he should take it.

OT1H: *On the one hand*, he needs the money that this job will pay,
Cause: *because* he should start paying off his student loans this year.

(i) **Global contrast:** But he could keep looking for a nicer, better-paying job.

(ii) **Local contrast:** But the loans could be deferred for a few more months.

(iii) **No contrast:** Also, his car needs to be serviced by the end of the month.

OTOH: *On the other hand*, he hates the idea of cleaning out panda cages and lion dens every day.

The first sentence in (6) introduces a situation in which a decision for or against a certain action is considered. The second sentence presents contrast1, which is an argument in favor of one option, and an explanation marked by

because. What follows is one of three different types of intervening sentences. The global contrast condition is shown in (i) with a sentence marked by *but*, whose content plausibly contrasts with the content of the OT1H-sentence and satisfies the contrast1∼contrast2 pairing (take the job vs. keep looking for a job). The local contrast condition is shown in (ii) with a sentence marked by *but*, whose content most plausibly contrasts with information embedded in the subordinating *because*-clause that directly precedes it and does not fully satisfy the contrast1∼contrast2 pairing (having to pay off loans vs. deferring loans). The baseline condition is shown in (iii) with a sentence marked by *also*, whose content does not contrast with any preceding information.

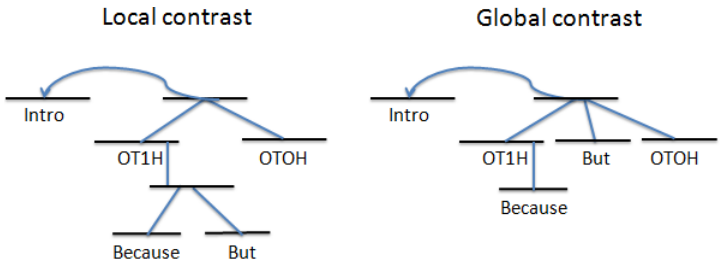

Figure 2: Attachment heights of the *But*-sentence in the local and global contrast conditions.

Figure 2 illustrates the attachment heights of the local and global contrast conditions. This distinction can be tested by omitting the because-clause: If the intervening sentence is still acceptable in the story, the contrast that this intervening sentence provides is a global one rather than a local contrast. To illustrate this, consider Example (6-ii) again, but without the because-clause: "On the one hand, he needs the money that this job will pay. But the loans could be deferred for a few more months." In this scenario, the loans mentioned in the local contrast are not introduced and the intervening sentence becomes unacceptable. A similar problem does not occur in the global condition, as illustrated by Example (6-i) without the because-clause: "On the one hand, he needs the money that this job will pay. But he could keep looking for a nicer, better-paying job."

If comprehenders are sensitive to these structural distinctions, their prefer-

ence or dispreference for a subsequent OTOH sentence is predicted to emerge in tasks that index biases about ensuing discourse material. Crucially, the global condition is predicted to disfavor a follow-on OTOH-sentence as an unnecessary or unexpected "third hand" compared to the baseline and local conditions.

Regarding the predictions from existing models of real-time discourse parsing, we note that there is in fact a lack of any such models for processing multi-sentence discourse. This means that our studies stand to test a fundamental question of how incremental structure building really is at the discourse level. Do comprehenders posit structure (embedded constituents, long-distance dependencies) in real-time or do they only retroactively impose structure once larger portions of the passage are available? We know that within-sentence structure building is incremental (see, for example, Altmann & Steedman, 1988; Altmann & Kamide, 1999; Boland, Tanenhaus, Garnsey & Carlson, 1995; Pickering, 1994), but within-sentence structures are necessarily constrained by syntactic rules. Within a sentence, the parser can expect certain components to be realized and can expect them in certain positions—e.g., there will eventually be a verb if the utterance is anything more than a fragment and a transitive verb in English must precede its direct object. But at the discourse level, the number of possible structures is rarely constrained—there is no analogous guarantee that a discourse will introduce any particular element in any particular position.

If discourse parsing differs from syntactic parsing due to the large number of possible structures and the limited number of governing constraints, we would expect to see little difference between (6i-iii), at least during comprehenders' incremental processing. If discourse parsing permits the generation of weak expectations of discourse (i.e., expectations of possible discourse relations but not of discourse structure), comprehenders are predicted to be surprised by OTOH in both (6i) and (6ii) because the expectation of some type of contrast has already been satisfied. If discourse parsing permits the generation of strong expectations over possible structures (i.e., expectations that are computed over parse trees whose generation may include embedded constituents), comprehenders are predicted to be most surprised by (6i), whose resulting 3-sister structure

violates the constraint imposed by OT1H for a binary OT1H–OTOH pairing.

Experiment 1: Norming study

Our studies rely on a within-items manipulation of the additional intervening sentence (see (6) where the global contrast condition is about better-paying jobs, the local contrast condition about loans, and the no-contrast condition about a car). Crucially, we want to avoid a scenario where any observed effects are attributable to basic properties of the information in those intervening sentences. The global and local conditions are discourse structurally the most interesting for our later experiments because they will allow us to distinguish comprehenders' awareness of embedding and attachment in the story continuations they write (Experiment 2) and in their reading times (Experiment 3). However, if participants find one story variant more entertaining or topically more interesting (jobs vs. loans?), their behavior in the subsequent experiments might reflect those judgments rather than the discourse parsing that we probe in Experiments 2 and 3.

We therefore collected naturalness ratings on a superset of stories, selecting for our later experiments a subset for which participants gave similar ratings to the two contrast conditions, under the assumption that they found the content of both similarly interesting and natural. The experiment had a 3×2 design: type of intervening sentence (global/local/no contrast) $\times$ presence/absence of the final OTOH-sentence. The presence of the OTOH-sentence was varied in order to determine the naturalness of the item at the precritical region, right before a participant would encounter (or write a continuation for) the final sentence. The OTOH-absent condition mirrors the materials for the story continuation task; the OTOH-present condition corresponds to the eyetracking task.

Participants

144 native English speakers (age range 18-75 years; mean age 35 years; 77 female), registered as 'workers' on the Mechanical Turk website hosted by Amazon, received monetary compensation for their participation ($1 per batch).

Participants had various educational levels ranging from high school to a Master's degree.

Materials and Procedure

Thirty-one items were created for this study, with the intention of selecting the 24 most suitable items. All had the same structure as (6). The presence or absence of the final OTOH-sentence was varied, creating six conditions. All experimental items can be found in Appendix C.

Each participant rated the naturalness of one version of 10 or 11 stories (because of the uneven number of items), and 8 fillers items on a scale of 1-7 (7 as most natural). Every version of each item was rated by eight people. Filler items consisted of short stories that were either unnatural or natural, to create a spectrum of naturalness for assessing participants' understanding of the rating scale. Natural filler stories were stories with temporal or causal coherence relations, without any discourse violations. These stories were less complex than the experimental items. Unnatural filler stories contained discourse violations such as an incorrect connective given the context ('since' where it should be 'even though') or an incorrect referring expression.

Analysis Methods

Results of all experiments reported in this paper were modeled using linear mixed-effect regression models (LMER; Baayen, Davidson & Bates, 2008). Models were evaluated using *lme*4 package within the statistical software R (Bates & Sarkar, 2007; R Development Core Team, 2008). For binary response variables, we used binomial mixed effects regression models. We always started out with maximal random effect structure and reduced random effects only in cases of non-convergence of the full model (Barr, Levy, Scheepers & Tily, 2013). In these cases, we first tried to simplify the model by removing correlation between random intercepts and random slopes (lmer model: Y $\sim$ X + (1+X || item) + (1+X || subject)), and then proceeded to removal of random intercepts or random slopes (lmer model: Y $\sim$ X + (1 | item) + (1 | subject)). All cases

Table 1: Mean rating (and standard deviation) of the naturalness of the items per condition.

Condition	OTOH-absent		OTOH-present	
	M	SD	M	SD
Global contrast	4.75	1.71	4.78	1.63
Local contrast	4.65	1.74	4.99	1.60
No contrast	4.23	1.89	5.65	1.47

where maximal models did not converge are reported in the article together with the results of the maximal converging model.

The factor for OTOH presence/absence was centered, and the 3-level factor for intervening sentence was deviation coded. Significance of fixed effects was evaluated by performing likelihood ratio tests, in which the fit of a model containing the fixed effect for each condition is compared to another model without it but that is otherwise identical in random effects structure. All pairwise comparisons were conducted using subsets of the data, only including the observations for the relevant pairs of conditions with re-centered predictors. For the fixed factors, we report the regression coefficient, the standard error, the t-value, and the corresponding p-value associated with the likelihood ratio test. For the fixed factors with more than two levels, we report only the p-value and the degrees of freedom, as the likelihood ratio test only evaluates the difference between models with/without that entire factor.

Results

To obtain a set of 24 similarly acceptable items for the remaining studies, the seven stories with the lowest rating in any of the conditions were excluded from this analysis and from the subsequent experiments. The mean ratings for the conditions, based on the 24 included items, are given in Table 1. Rating scores of the set of 24 items were modeled using linear mixed-effect regression models, as described above.

The ratings show several patterns. First, as per our goals, there is no main effect of intervening sentence type ($p = 0.42$). There is, however, a main effect of OTOH presence, whereby the OTOH-present condition was rated more highly than the OTOH-absent condition ($\beta = 0.584$, SE = 0.16, t = 3.64, $p < 0.001$), but this is driven by an interaction between OTOH presence and intervening sentence type (multi-level factor so only model comparison p-value reported: $p < 0.001$). To understand the nature of the interaction, we consider the OTOH-absent and OTOH-present conditions in turn.

Looking at only the OTOH-absent stories, we find that some contrast is better than having none, and the lowest ratings are for the no-contrast condition (a main effect of intervening sentence type ($p < 0.05$, 2 d.f.). Pairwise comparisons reveal a difference between ratings in the global condition and the non-contrastive condition ($\beta = 0.54$, SE = 0.18, t = 3.09, $p < 0.01$) and between the local condition and the non-contrastive condition ($\beta = 0.39$, SE = 0.18, t = 2.24, $p < 0.05$). Crucially, however, the two contrast types (global vs. local) do not differ significantly from each other (β = -0.12, SE = 0.17, t = -0.74, $p = 0.46$).

The OTOH-present stories provide the flip side of this: Stories with no intervening contrast received the highest ratings and again the two contrast types did not differ from each other. The likelihood ratio test showed a main effect of intervening sentence type in OTOH-present stories ($p < 0.001$, 2 d.f.). Pairwise comparisons revealed a difference between ratings in the global and non-contrastive conditions (β = -0.87, SE = 0.22, t = -3.90, $p < 0.001$) and between the local and non-contrastive conditions (β = -0.652, SE = 0.15, t = -4.26, $p < 0.001$). Again, no significant difference was found between the global and local conditions ($\beta = 0.223$, SE = 0.23, t = 0.98, $p = 0.32$).

Discussion

This story acceptability study served as a norming study for the following experiments. We wanted to ensure that any effect we might find in the remainder of this paper between the global and the local condition is not due to a

difference in naturalness of the stories. The results show no difference between the global and local condition, which leads us to assume that the acceptability of the conditions will not be a confounding factor in our other experiments. We also find that non-contrastive intervening sentences between OT1H/OTOH resulted in more natural-sounding stories than either locally or globally contrastive intervening sentences. This is not surprising, as OT1H and OTOH are typically used to express a relation with two contrastive situations, rather than three.

Considering that the stories were rated on a scale of 1 to 7, the scores of the experimental items (ranging between 4.23 and 5.65) can be considered quite low. This could be due to the presence of the filler items that were designed to create a spectrum of naturalness. Hence, some of these fillers were less complex stories, which might have influenced scores for the experimental items.

In the next sections, we look at expectations that the local and global conditions generate, and their time-course. Section 5 presents a story continuation study using the OTOH-absent materials and Section 6 presents an eyetracking study using the OTOH-present materials.

Experiment 2: Story continuation study

For the current experiment, we asked participants to write a story continuation to the version of the items without the OTOH-sentence. The goals of this study were twofold: First, it tests whether readers interpret both a globally and a locally contrastive sentence as satisfying the expectation for a contrast set up by OT1H; second, the current study gives us insight into which connectives participants prefer to use to signal contrast with the OT1H clause.

Participants

90 native English speakers (age range 18-61 years; mean age 33 years; 54 female), registered as 'participants' on the Prolific Academic website received

monetary compensation for their participation (1.50 GBP per batch). Participants had various educational levels ranging from high school to a Master's degree.

Materials and Procedure

The experimental stimuli consisted of the 24 stories that were selected based on acceptability judgments, see Section . The stories did not contain the OTOH sentence, and they appeared in three conditions: The sentence following the OT1H sentence contained a global contrast, a local contrast, or no contrast. Each version of each item was completed by ten people. Each participant saw one version of 8 stories, and 10 fillers items. Filler items consisted of short stories in the same format as the experimental items, without the markers OT1H and OTOH. Participants were asked to read the sentences and then write two sentences to continue the story.

Annotation

The continuations were manually coded for the following properties:

- Explicit marking: whether a connective was used, and if so, which one.
- Discourse relation: whether the content of the continuation contrasted with the OT1H-clause or not.

For determining the discourse relation, we considered the two sentences that participants were asked to write, and determined whether one of the provided sentences presented a contrast with the content in the OT1H-clause. To help determine this, a connective insertion test was used with the intervening sentence excluded: If a contrastive connective could be inserted directly between the OT1H-clause and the provided continuation, the continuation was marked as contrastive with OT1H. Passage 7 shows an example of the insertion test for a continuation in the local contrast condition (intervening sentence in brackets).

(7) *Prompt:* Frank is thinking about quitting his job at the supermarket after working there for five years. On the one hand, he thinks he could get

a more promising job at a multinational, because he studied accounting in college. [But he has no real work experience as an accountant.]

Continuation: (implicit 'On the other hand') Perhaps he could intern or do an apprenticeship. There are so many other options to consider.

Results

For the analysis of the continuations' discourse relations, we modeled the binary outcome of continuation type using mixed-effect logistic regression models, with likelihood ratio tests to compare models differing in the presence or absence of the fixed factor for condition. Models included random intercepts and random slopes.

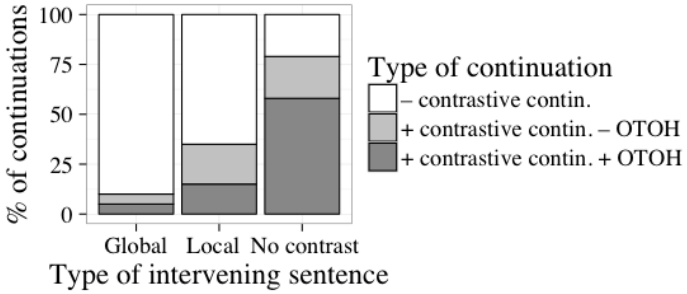

Figure 3: Types of continuations in the completion experiment, by type of intervening sentence.

Figure 3 shows the distribution of completions by condition. The global condition yielded the fewest continuations that contrasted with the OT1H sentence (10%, when +contrastive +OTOH continuations are collapsed with +contrastive -OTOH continuations); the local condition yielded more (35%), and the no-contrast condition yielded the most (79%). A likelihood ratio test confirmed a main effect of condition ($p < 0.001$, 2 d.f.). For convergence, we removed the correlations for the random slope of item, as well as the random slope for subject. Pairwise comparisons confirmed that the global condition differed from the local condition ($\beta = 2.69$, SE = 0.87, z = 3.1, $p < 0.001$), which in turn differed from the no-contrast condition (β = -3.66, SE = 0.81, z = -4.51, $p < 0.001$).

This means that participants were sensitive to the global / local contrast manipulation, and that especially the globally contrastive sentence strongly reduces the need for a subsequent contrast.

In 57.8% of all continuations, participants included an overt connective, allowing us to see how often OT1H is followed by the explicit OTOH expression and other connectives. The most common connective in continuations that contrasted with OT1H was OTOH (61.3% of +contrastive continuations). The most common non-OTOH markers used to signal +contrastive continuations were *however*, *but*, and *although*. Appendix B shows the full breakdown of connective use across continuation types and conditions.

If we analyze the continuations in terms of participants' use of explicit contrastive connectives rather than our own annotations of contrast in their continuations, the pattern of connective use follows the pattern of discourse relations reported above. The list of contrastive markers consisted of *on the other hand*, *however*, *but*, *although/though*, *on the downside*, and *conversely*. In a model of the binary outcome of presence/absence of a contrastive marker, we find a main effect of condition on the choice to use a contrastive marker ($p < 0.001$, 2 d.f.; the model did not contain a correlation for the random intercept and random slope of subject, nor a random slope for item). Pairwise comparisons showed a significant difference between the no-contrast condition and the global condition (β = -5.2, SE = 0.9, z = -5.81, $p < 0.001$; the model did not include a random slope of item), the non-contrastive condition and the local condition (β = -3.63, SE = 0.66, z = -5.52, $p < 0.001$), and the global condition and the local condition (β = 1.38, SE = 0.7, z = 1.98, $p < 0.05$). As expected, contrastive markers were used most often in the non-contrastive condition, followed by the local condition.

Discussion

The results from the story continuation study show that participants distinguish contrastive from non-contrastive intervening material (global/local vs. no-contrast) and, moreover, distinguish different types of intervening contrast

(global vs. local). These results suggest that people do anticipate a specific discourse structure after encountering OT1H and take into account the attachment height of the clauses in order to construct their formulation of a following utterance.

A similar picture emerges for participants' use of connectives: More contrastive connectives were used in the non-contrastive condition than in the other conditions. The local condition yielded more continuations with a contrastive connective than the global condition. Furthermore, although most of the contrastive markers were OTOH, other contrastive connectives were also used. This shows that although OTOH is the preferred marker after OT1H, other markers can also signal contrast2.

The current study has provided further insight into the readers' expectations after encountering a locally or globally contrastive sentence, and into the types of connectives that people use to complete stories with OT1H. However, it does not let us test whether real-time discourse parsing is predictive. In other words, do readers generate expectations of discourse structure immediately after having read OT1H, and do these expectations then affect early processing of OTOH? And are they able to immediately attach incoming (intervening) sentences to previously read content and adapt their expectations of the upcoming discourse structure accordingly? These questions are investigated in Experiment 3.

Experiment 3: Eye-tracking study

An eye-tracking study was conducted to test whether readers use OT1H as a cue to anticipate upcoming discourse structure, and how their online processing of OTOH is influenced by intervening material.

Participants

39 native speakers of English participated in this experiment, 7 of which had to be excluded (4 due to eye-tracking problems and 3 due to problems with the computer). Participants were recruited from the University of Edinburgh

community. Data from the remaining 32 participants (age range 18-30 years; mean age 22 years; 18 female) was analyzed. All participants had normal or corrected-to-normal vision. Participants were paid for their participation (15 GBP for 90 min) and were unaware of the purpose of the experiment.

Materials

The experimental stimuli consisted of the same 24 items used in the offline studies. The three conditions varied in the third sentence (the one intervening between the OT1H and OTOH sentences): global, local, and no contrast. We also included an additional filler condition in which OTOH occurred without the OT1H marker. For this condition we only used non-contrastive intervening sentences (marked by the connective *also*). This filler condition was included to ensure that OTOH was not always preceded by an OT1H cue in the experiment. Forty-eight stories for two unrelated experiments were included as fillers, along with 12 filler items containing aspects of all three experiments. In these additional 12 filler items, OT1H sometimes occurred without OTOH, or the other way around. This was done to approximate the rate at which OT1H and OTOH occur together in natural text. Note that even if participants come to expect OTOH after having seen OT1H, this can only serve to decrease our chances of seeing a difference across the global, local, and no-contrast conditions.

The stimuli were counterbalanced across four lists, with each story appearing in a different condition in each list. All participants saw each story in only one condition. The participants were randomly assigned to one of the lists, and for each participant the list was presented in a unique order.

Procedure

Participants were tested individually. They were seated at a distance of approximately 60 centimeters from the monitor. Participants' eye movements were recorded with SR Research Eyelink 1000 at the sampling rate of 500 Hz. Viewing was binocular, but only the participant's dominant eye, as determined by a parallax test prior to the experiment, was analyzed. Participants rested

their head on a chin-rest. Their movements were not restricted, but they were instructed to move as little as possible during the eye-tracking part of the experiment.

Each session started with an oral instruction, after which the eye-tracker was adjusted if necessary. A brief calibration procedure was then performed. This procedure was repeated after a short break halfway through the experiment and whenever measurement accuracy appeared insufficient. Upon successful calibration the experiment started with three practice trials. The participant was instructed to read the passage at a natural pace and press the space bar after reading the entire story. Before presentation, a fixation mark appeared, first in the middle of the screen and then at the position of the first word of the first sentence. The stories were presented randomly and in their entirety on the screen. It was ensured that the critical region "On the other hand" never occurred at the beginning or end of a line. To encourage participants to read carefully, each story was followed by a comprehension question. The questions were answered using the 'F' key for *no* and the 'J' key for *yes*. The eye-tracking component lasted approximately 45 minutes.

Analysis procedure

For analysis purposes the sentences were divided into four regions, as illustrated in 8:

(8) (But he could keep looking for a nicer,) / better-paying job. $_{pre-criticalregion}$ / On the other hand, $_{criticalregion}$ / he hates $_{spillover1}$ / the idea $_{spillover2}$ / (of cleaning out panda cages and lion's dens every day.)

The critical region was the OTOH region, as this is where the reader could have difficulties due to misanalysis depending on how the preceding sentence is aligned in the discourse structure. The two words preceding this region were the pre-critical region.[1] The first spillover region contained the two words following

[1]The two words preceding the pre-critical region were not analysed, but were included in Figure 4 for illustrative purposes.

the expression OTOH and the second spillover region contained the third and fourth word following OTOH.

Three reading time measures were computed: first pass duration, regression path duration and total reading time. First pass duration is the time spent in a region before moving on or looking back. This measure reflects the immediate processing difficulties a reader has when reading a region for the first time (Rayner, 1998). Regression path duration is the summed fixation duration from when the current region is first fixated until the eyes enter the next region on the right. This measure thus includes regressions to regions to the left of the current region. Regression path duration can be seen as reflecting the process of integrating the linguistic material with the previous context (Rayner, 1998). Total reading time is the total time spent in a region, including regressions to that region.

Prior to all analyses, skipped regions were treated as missing data. Additionally, fixations shorter than 80 milliseconds and longer than 2000 milliseconds were removed. It is assumed that the reader did not process any linguistic input during fixations shorter than 80 milliseconds, and that fixations longer than 2000 milliseconds reflect tracker losses or indicate that the participant was distracted. In all reading time measures, outliers were removed by excluding reading times more than three standard deviations from both the participant's mean and the condition's mean in a region (1.7% of all data).

Results

Reading times were modeled using linear mixed-effect regression models, with subjects and items as crossed random effects. As in the previous experiments, likelihood ratio tests were computed to compare mixed-effects models differing only in the presence or absence of the fixed factor for condition. Table 2 shows the mean reading time measures per condition and region; Figure 4 shows the first pass duration per type of intervening sentence.

No difference in reading time was found at the pre-critical region in the first pass duration ($p = 0.17$, 2 d.f.), regression path duration ($p = 0.26$, 2 d.f.), and

Table 2: Mean reading times and standard deviations per measure and region.

	Region							
	Pre-critical		*Critical*		*Spillover 1*		*Spillover 2*	
	M	SE	M	SE	M	SE	M	SE
First pass durations								
Global contrast	282	132	352	146	276	129	231	116
Local contrast	240	114	318	134	261	124	237	109
No contrast	259	139	322	131	242	106	232	106
Regression path durations								
Global contrast	409	228	391	190	306	173	288	216
Local contrast	382	255	369	227	297	215	298	235
No contrast	364	241	351	188	274	173	289	203
Total reading time durations								
Global contrast	335	148	453	221	326	170	272	148
Local contrast	311	152	400	178	309	176	285	160
No contrast	303	161	387	165	293	146	297	173

total fixation duration ($p = 0.38$, 2 d.f.).

At the critical region *On the other hand*, the results showed a main effect of intervening sentence type in the first pass duration ($p < 0.05$, 2 d.f.) and total fixation duration ($p < 0.05$, 2 d.f.). Pairwise comparisons were conducted to see whether the three conditions differed significantly from each other. These will be presented per reading time measure and condition. No significant difference in regression path duration was found at the critical region ($p = 0.19$, 2 d.f.).

For the first pass duration, the results revealed a significant difference between the global and no-contrast conditions ($\beta = 30.72$, SE = 14.18, t = 2.17, $p < 0.05$; the model did not include a correlation between the random intercept

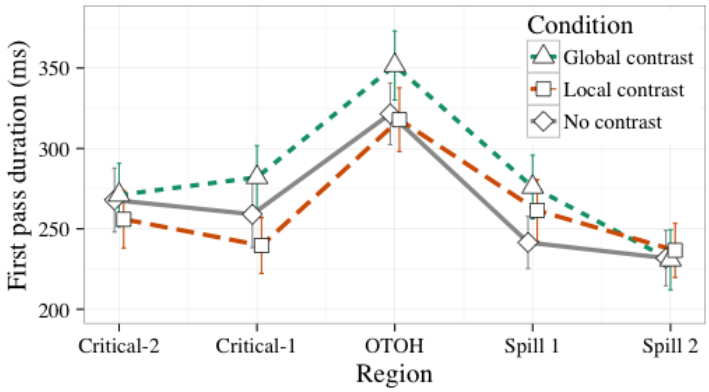

Figure 4: First pass duration by experimental region, per type of intervening sentence.

and random slope of item), and between the global and local conditions (β = -35.06, SE = 13.45, t = -2.61, $p < 0.05$): Reading times of OTOH were longer in the global condition than in the local and no-contrast conditions. No significant difference was found in first pass duration of OTOH between the local condition and the no-contrast condition (β = -3.29, SE = 14.83, t = -0.22, $p = 0.82$).

For the total fixation duration, a similar picture emerges: the global condition differs from the no-contrast condition (β = 67.68, SE = 24.15, t = 2.80, $p < 0.01$), as well as from the local condition (β = -50.86, SE = 20.5, t = -2.48, $p < 0.05$; the model did not include a correlation between the random intercept and random slope of item): The total fixation duration was longer in stories with the globally contrastive sentence than in stories with the locally contrastive or non-contrastive sentence. No significant difference was found between the no-contrast and local conditions (β = 16.93, SE = 20.23, t = 0.84, p = 0.4).

For spillover region 1, no effect of intervening sentence type was found in the first pass duration (p = .11, 2 d.f.), regression path duration (p = .43, 2 d.f.) or total reading time measure (p = .28, 2 d.f.). Similarly, no effect was found at spillover region 2 for first pass duration (p = .81, 2 d.f.), regression path duration (p = .92, 2 d.f.) or total reading time (p = .37, 2 d.f.).

Even though the reading times at the precritical region did not differ significantly between conditions, Figure 4 does show a slight difference in reading times at the precritical region. To ensure that any effect found at the critical OTOH region is not caused by a spillover effect from any (non-significant) difference at the precritical region, we ran new models that include the reading times at the precritical region as a predictor.

For the first pass duration, the results still showed a main effect of intervening sentence type ($p = 0.05$, 2 d.f.). Pairwise comparisons again showed a significant difference between the global and local conditions (β = -25.7, SE = 11.42, t = -2.25, $p < 0.05$; the model did not include a correlation between the random intercept and random slope of item). The difference in reading times in the global and no-contrast conditions were only marginally significant when precritical reading times were included as a predictor ($\beta = 21.18$, SE = 10.97, t = -1.84, $p = 0.06$). The difference between the local and no-contrast conditions remained non-significant (β = -5.55, SE = 11.19, t = -0.49, $p = 0.62$). For the total fixation duration, the previously significant effect became marginally significant when precritical reading times were included as a predictor ($p = 0.06$, 2 d.f.). This will be addressed in the discussion.

Discussion

The pattern in participants' eye movements shows that the local and global conditions differ reliably from each other in first pass reading times. This distinction demonstrates comprehenders' ability to build discourse structures that distinguish between superficially similar intervening constituents. Even though both the local and global intervening sentences start with *but*, the underlying structures are different; only the global condition satisfies the expectation for a relevant contrast, and participants show sensitivity to this difference. Below we discuss the precritical and spillover regions.

The results indicate some spillover effect from the precritical region on the reading times of the critical OTOH region: The difference between the global and no-contrast conditions, significant in the original model of first pass duration

at the critical region, is marginal in a model that includes precritical reading times; the main effect of intervening sentence type on total fixation duration is likewise marginal in the larger model. These findings could be attributed to differences in the final words of the intervening sentences (which differ across conditions), though this would require that the differences be systematic enough to differentiate the conditions and influence effects at the critical region. A review of the materials reveals no immediately evident bias in the construction of the intervening sentence. Alternatively, two other factors could be at play – parafoveal processing and the reduced power for models with an additional factor. Parafoveal processing (Rayner, 1998) is known to occur when readers process words surrounding the current fixation (the parafovea extends out to 5 degrees on either side of fixation). Considering that OTOH is a marked phrase (after being pre-activated by OT1H), it is likely easily identifiable from parafoveal vision. A follow-up experiment could address these explanations by postponing the presentation of OTOH: If another intervening sentence occurs before OTOH and is identical across conditions, any increase in reading times for the global condition at this new intervening sentence would provide evidence that the increase is caused by parafoveal processing. In addition to parafoveal processing, the inclusion of the precritical reading times as an extra factor in our post-hoc analyses may have also served to reduce the models' power.

Although effects are commonly observed in spillover regions, our results were limited to the critical region itself. This likely reflects the nature of the critical region: "On the other hand" is a marked and quite long phrase, possibly already pre-activated after encountering "On the one hand". The length of the region may have provided sufficient time for participants to resolve any difficulty in integrating OTOH after encountering an intervening contrast.

The results from the eye-tracking experiment can be summarized as follows. First, reading times of OTOH were longer when the expression was preceded by a globally contrastive sentence than when it was preceded by a non-contrastive sentence or a locally contrastive sentence. Crucially, the difference between the global and local conditions remained when taking into account the reading

times of the precritical region. No significant difference in reading times of OTOH were found between the no-contrast condition and the local contrast condition. Based on these results, we can conclude that the expectation for a contrast set up by OT1H is satisfied by a globally contrastive sentence, but not by a locally contrastive sentence. This means that readers construct structural expectations of the scope of contrast based on OT1H, and that they are able to build and update a specific prediction of discourse structure immediately while reading. When encountering a globally contrastive sentence, their prediction for a contrast is satisfied. This then leads to processing difficulty when encountering OTOH, which is reflected in additional reading times.

General discussion and conclusion

The current paper addresses the question of how comprehenders' discourse processing makes use of predictive cues to structure building in multi-sentence passages. More specifically, we investigated using offline and online tasks how structural readers' expectations of upcoming contrast are using contexts with the set of OT1H/OTOH markers.

The results of the offline story continuation study showed that passages with an OT1H-marked contrast1 and a locally contrastive sentence favor continuations that convey contrast2 significantly more than passages with a globally contrastive sentence. This indicates that participants are indeed sensitive to the attachment height of the locally and globally contrastive sentences. The results also showed that although most of the markers signalling contrast2 were OTOH or a compound thereof, other contrastive connectives such as *but* and *although* were also used. This suggests that the anticipation of contrast2 is not merely a surface expectation for the OTOH expression.

The eye-tracking-while-reading study tested whether readers construct structure-sensitive expectations of contrast during on-line processing. The results showed that OTOH is processed more easily following a non-contrastive or locally contrastive intervening sentence, compared to a globally contrastive intervening

sentence. This result provides further support that readers construct structure-sensitive expectations, whereby only a contrastive sentence that targets contrast1 specifically can satisfy the expectation for contrast2.

Taken together, the findings of our experiments provide evidence of comprehenders' anticipation of upcoming coherence relations in multi-sentence discourse. Moreover, they support the Right Frontier Constraint, since the evidence indicates that readers do indeed track the height of incoming input, and close off contrast1 after globally contrastive intervening sentence, but not after a locally contrastive intervening sentence. The current studies also support a model of discourse processing in which comprehenders construct and maintain structural discourse representations.

Based on these results, a discourse level model of language processing would minimally need to distinguish contrastive from non-contrastive coherence types, it would have to be able to incrementally build discourse structure with constituents, predicting a distribution of upcoming discourse relations as well as a distribution over specific discourse markers. In particular, any model that would only calculate discourse relational structure once a full clause of sentence has been completely processed, would not be compatible with the results of our experiments.

An open question regarding predictive discourse processing is how comprehenders manage long-range discourse expectations. Do they expect an (ideally OTOH-marked) contrast completion to appear as soon as a clause is finished, carrying this expectation forward through the discourse until it is satisfied? Or are those expectations suspended for as long as the discourse is discussing an intervening aspect, such that the outstanding expectation for OTOH is only reactivated once there has been some kind of completion signal for any intervening discourse structures, in analogy to syntactic structure? Given the very different levels of constraint in syntax vs. discourse, we think that these questions are worth exploring in future research. While our stimuli design allowed us to compare conditions with and without the marker OT1H in the eye-tracking experiment, the null effect of a lack of reliable difference between the two con-

ditions did not allow us to draw any conclusions; but pilot experiments show evidence for anticipation of a OTOH type cue following OT1H in Dutch.

The generalizability of discourse prediction effects observed here for the OT1H / OTOH construction should furthermore be tested for other text structuring markers such as list signals (e.g., *There are three things that are relevant. First... Second... Third...*). These might be particularly interesting because their distribution in a text (holding between larger text segments) might differ from the distribution of OT1H / OTOH.

What can we conclude, then, about the anticipation of discourse structure between sentences? We have shown that coherence markers do not only aid comprehension by functioning as processing signals that help establish local coherence, but also by enabling predictions about upcoming discourse structure. The findings here can therefore be interpreted as evidence for the relevance of anticipation of discourse structure in text processing. The fact that the expression *On the other hand* is dispreferred more when it follows a globally contrastive sentence than a locally contrastive sentence supports the hypothesis that readers are able to create specific expectations of discourse structure based on cues in the preceding context. Moreover, the fact that they are able to maintain such predictions across sentences can be taken as evidence that readers build predicted discourse structures immediately, rather than wait until the end of a sentence to integrate the full discourse structure. Readers are then able to update this expected structure as soon as they encounter evidence that their current representation is false, as evidenced by the longer reading times at OTOH.

Our study also addresses a persistent gap in psycholinguistic research: While there has been a wealth of research on syntactic processing, there has been a lack of structured psycholinguistic work on discourse processing. The reason for that may have been the field's perception that the possible dependencies at the discourse level are too unconstrained, too numerous, or just too ill-defined. The present work makes an important contribution, because it identifies and tests a type of cue that does constrain upcoming discourse in a way that can

be analysed and tested. Our study on OT1H/OTOH hence opens up a new domain for psycholinguistics by taking a first step towards an understanding of what comprehenders' strategies are for building up a representation of a whole discourse while incrementally perceiving new input.

Acknowledgements

This research was funded by the German Research Foundation (DFG) as part of SFB 1102 "Information Density and Linguistic Encoding" and the Cluster of Excellence "Multimodal Computing and Interaction" (EXC 284). We are grateful to Chris Cummins and Emily Nicolet for help on an earlier version of this study, to Patrick Sturt for advice on statistics, to Ines Rehbein for help with extracting data from ukWaC, and to Ted Sanders for reading an earlier draft.

A. Corpus example of OT1H and OTOH

The following example passage is taken from the ukWaC corpus.

(9) **On the one hand** there was no group of Yugoslavs to whom the phrasing of Robertson's order applied more completely than the Croat Ustachi, Domobranci and regular troops under German Army Group E, who at the time were approaching the Austrian border, armed and in very large numbers. **Certainly** Gen McCreery wanted specific authorization from AFHQ to deal with these Croats in precisely the terms which Robertson's order gave him. **But** we have no firm indication that, at the moment Robertson drafted the order, he was specifically aware of McCreery's request for such an authorization. **Nor** did Gen Robertson word his order to make it refer exclusively to Croats. **If** he had meant to, he surely would just have stated "Croats". That he did not so do indicates he had some other categories in mind as well. **On the other hand**, it is **also** clear that Robertson originally intended to

word his order in such a way that some dissident Yugoslavs should be excluded from the hand-overs, **and** these he categorized as "Chetniks".

In order to understand example (9), the reader must build a discourse structure that accommodates a contrastive relation based on OT1H. Before encountering the second argument of this relation, however, the reader has to process additional explicit discourse markers, namely *certainly,but*, *nor*, and *if*.

B. Appendix to Experiment 2: Occurrence of connectives

Table 3 presents the distribution of connectives in Experiment 2.

C. Material

Stimuli (sentence version (i) is a globally contrastive intervening sentence, version (ii) is a globally contrastive intervening sentence, and (iii) is a non-contrastive intervening sentence).

1. Michael heard that his favourite singer Beyoncé is coming to Edinburgh during her tour. On the one hand, he's thinking about taking two days off for the concert, because she'll only be in Edinburgh for one show. **(i)** But he might only be able to get one day off work. **(ii)** But she will probably be back next year for more concerts. **(iii)** Also, he still needs to finish some of his vacation days. On the other hand, he has a deadline coming up and really needs to get his work done.

2. On a rainy day, Gillian was thinking of asking her friend Mark to join her for a shopping trip. On the one hand, she was thinking that they could take the car, because he just passed his driver's license exam last month. **(i)** But she also knows that the bus ride is pretty quick. **(ii)** But he might not feel comfortable driving in the rain. **(iii)** Also, they will be able to play their own music in the car. On the other hand, she might get everything she needs faster if she just goes shopping by herself.

3. Jon is from Spain and is considering going to a Scottish ceilidh, to dance and listen to music. On the one hand, he thinks it might be a lot of fun, because he's heard great stories about these parties from his brother. **(i)** But he doesn't know

Table 3: Occurrence of connectives per condition: +contrastive continuations convey content that contrasts with the OT1H-clause. -contrastive continuations convey non-contrastive content or content that contrasts with material outside the OT1H-clause (Experiment 2).

Continuation type	Connective	global	local	no-contrast	Total
	on the other hand	11	31	128	170
+contrastive	but on the other hand	1	0	5	6
OTOH present	on the other	0	0	5	5
	and on the other hand	1	2	0	3
	however	2	3	25	30
	but	2	6	10	18
+contrastive	although/though	0	5	6	11
OTOH absent	alternatively	1	3	0	4
	also	0	1	1	2
	on the downside	0	0	1	1
	otherwise	0	0	1	1
	conversely	0	0	1	1
	(no connective)	6	33	9	48
	so	10	5	3	18
	also	4	9	2	15
	but	1	5	0	6
-contrastive	therefore	0	3	2	5
	however	2	1	0	3
	although/though	1	0	1	2
	then	2	0	0	2
	as a result	1	0	0	1
	(no connective)	195	133	40	368
Total		240	240	240	720

anybody else who will be there. **(ii)** But he does not have the same taste in music as his brother. **(iii)** Also, he would like to learn more about Scottish culture. On the

other hand, he's really worried about other people seeing his underwear when dancing with a kilt.

4. Mary is thinking about taking part in the whisky tasting at the Talisker distillery in Scotland. On the one hand, she's really curious about trying real Scottish whisky, because she's read a lot about the smoky smells. **(i)** But she is a lightweight and doesn't like getting drunk. **(ii)** But she has a blocked nose and doesn't smell much. **(iii)** Also, she has heard that it's different from American whisky. On the other hand, she feels like the whisky tasting could be too expensive for her travel budget.

5. Peter is looking for Scottish recipes with which he can impress his visitors from overseas. On the one hand, haggis would be a good dish to serve, because he used to love his mother's haggis when he was a child. **(i)** But not everyone likes sheep intestines and brains. **(ii)** But his mother is actually a much better cook than he is. **(iii)** Also, it is something that's very unique to Scotland. On the other hand, he won't be able to find the special utensils to prepare the haggis anyway.

6. John has been dating Sue for a few months and he's thinking about their future together. On the one hand, he'd like to buy a bigger house and move in with her right away, because she can cook amazingly well. **(i)** But he would like to enjoy the bachelor life a little longer. **(ii)** But she is terrible at remembering to wash the dishes. **(iii)** Also, she is great at making a house feel like a cozy home. On the other hand, he might just take things slow and give her the keys to his apartment.

7. Bob suggested a business merger with Jennifer's company, and now she's considering it. On the one hand, she'd like to join forces with Bob, because he already has many loyal, and even some famous customers. **(i)** But she would rather avoid the costly legal fees of a merger. **(ii)** But she is worried the celebrities will have high demands. **(iii)** Also, he is known for having a lot of business experience. On the other hand, she wants to make sure she can rise to power as CEO without competition.

8. Susan doesn't like her job at the warehouse and is mulling over what to do next with life. On the one hand, she might start a farm somewhere in New Zealand, because New Zealand is the sheep capital of the world. **(i)** But she could open a scuba diving school in Australia. **(ii)** But the wool industry has been going downhill these days. **(iii)** Also, she has heard the nature is beautiful in New Zealand. On the other hand, she could also move to India and join an Ashram in order to meditate.

9. Joseph got a job offer from the Edinburgh Zoo and he's pondering whether he should take it. On the one hand, he needs the money that this job will pay, because he should start paying off his student loans this year. **(i)** But he could keep looking for a nicer, better-paying job. **(ii)** But the loans could be deferred for a few more months. **(iii)** Also, his car needs to be serviced by the end of the month. On the other hand, he hates the idea of cleaning out panda cages and lions' dens every day.

10. Maryann is considering taking surfing lessons during her vacation in Hawaii next month. On the one hand, she loves the idea of surfing in Hawaii, because she's heard that there are many cute surfer boys there. **(i)** But she is afraid her painful back will not be up to it. **(ii)** But she already has a boyfriend whom she loves a lot. **(iii)** Also, she has heard that the waves are really high in Hawaii. On the other hand, relaxing at the beach with a nice cocktail sounds very good to her too.

11. Kate wants to go visit her brother in Aberdeen with her newborn baby during the weekend. On the one hand, she's thinking about driving there, because it's less of a hassle than traveling by train with a newborn baby. **(i)** But the traffic in Aberdeen is always busy and chaotic. **(ii)** But her baby has always been quiet on trains so far. **(iii)** Also, going by car will probably be faster than by train. On the other hand, she's thinking the baby might be too young to spend so much time traveling.

12. Frank is thinking about quitting his job at the supermarket after working there for five years. On the one hand, he thinks he could get a more promising job at a multinational, because he studied accounting in college. **(i)** But he would miss the personal contact with customers. **(ii)** But he has no real work experience as an accountant. **(iii)** Also, he knows someone who could get him an interview. On the other hand, he has a good chance at becoming a manager at the supermarket next year.

13. Lisa found out that she is unexpectedly pregnant and is unsure what to do with the baby. On the one hand, she's thinking that she'd like to keep the baby, because she always loves playing with her baby nephew. **(i)** But she is not sure whether her boyfriend wants a baby. **(ii)** But she usually sees her nephew for a few hours only. **(iii)** Also, she has always dreamed about being a good mother. On the other hand, she always wanted to be married and have a house before having a baby.

14. Daniel has been dating a girl from work and he's considering introducing her to his parents. On the one hand, he's sure his mother will be excited, because she

thinks that, at his age, he should be married already. **(i)** But he thinks his mother will dislike his girlfriend's tattoos. **(ii)** But he is unsure whether he'd want to marry this girl. **(iii)** Also, she is always curious about the girls he's seeing. On the other hand, his girlfriend might think it's too soon to meet his parents and get scared.

15. Nicole is turning 27 next week and she's mulling over what snacks to serve at her party. On the one hand, she'd like to prepare finger food, because she wants to show her mother that she has improved her cooking. **(i)** But she wants to spend very little time in the kitchen. **(ii)** But she is unsure if her mother will come to the party. **(iii)** Also, she thinks it's classy to serve finger food at a party. On the other hand, she can make it easy for herself and only serve some pretzels and nuts.

16. Henry's laptop is quite old and now he's debating whether or not he should get a new one. On the one hand, he can afford to buy one now, because he just heard from his boss that he'll receive a bonus this Christmas. **(i)** But the laptop that he has is actually still working fine. **(ii)** But he wanted to spend his bonus on a vacation to Bali. **(iii)** Also, he has some money saved up from his birthday. On the other hand, he would also like a tablet and could use that to check emails too.

17. Anthony woke up with a headache and now he's thinking about calling in sick for work today. On the one hand, he won't miss a lot, because he was only planning on attending a talk and had no other meetings planned. **(i)** But he would like to get ahead on his quarterly reports. **(ii)** But he was excited about learning more from the talk. **(iii)** Also, he can check his email and answer calls from home. On the other hand, he might feel a lot better already after taking an aspirin and some vitamins.

18. Johanna got an invitation from her aunt to visit her for two weeks in Tanzania this winter. On the one hand, she thinks it could be a great experience, because she would be able to go on a safari for the first time. **(i)** But she fears the African heat she's heard so much about. **(ii)** But she is very afraid of wild animals, especially lions. **(iii)** Also, she is curious about African culture and customs. On the other hand, she's not sure whether she can find someone to take care of her dog.

19. Melissa's friend lives at the seaside and now Melissa is planning her weekend trip there. On the one hand, she'd like to drive there directly after work on Friday, because her friends will have a party there that night. **(i)** But the roads to the seaside would be incredibly busy. **(ii)** But someone she doesn't like might also attend. **(iii)** Also, she wants to spend as much time there as possible. On the other hand, relaxing

at home on Friday evening would make her feel less stressed and rushed.

20. Nan is unsure of what she wants to do after she gets her Bachelor's degree in informatics. On the one hand, she might do a Master's at the same uni, because that'll make it easier to get a research position there. **(i)** But she would like to study abroad once in her life as well. **(ii)** But the research positions at her university are not well paid. **(iii)** Also, it will look good on her CV if she has a Master's degree. On the other hand, she could do a traineeship at Shell and learn more about the corporate world.

21. David is thinking about taking his girlfriend out somewhere to improve their relationship. On the one hand, he'd like to invite her to a rock music festival, because his old friends will be playing a short set there. **(i)** But he is not sure whether she really likes rock music. **(ii)** But they would spend a lot of time with his friends then. **(iii)** Also, he likes the other bands that will play at the festival. On the other hand, she might be happier if they watch a romantic movie together at the cinema.

22. Gary's favorite holiday is Christmas and now he's wondering how to celebrate it this year. On the one hand, he'd like to go to South Africa, because his parents recently moved there and he'd like to surprise them. **(i)** But he would rather not celebrate Christmas in warm weather. **(ii)** But he has been quarrelling a lot with his father lately. **(iii)** Also, he would like to experience the South African way of life. On the other hand, he heard that the public security of South Africa has a really bad reputation.

23. Lucy has a lot of savings and she's thinking about how to manage her personal finances. On the one hand, she'd like to invest in stocks, because her sister has experience in stocks investment and can give her advice. **(i)** But stocks investment is often accompanied by high risk. **(ii)** But her sister has actually made little profit from it so far. **(iii)** Also, she has been told that stocks can be very profitable. On the other hand, she could help out her best friend and invest her money in his business.

24. Helen found some signs that Ben, her colleague and friend, has violated the company rules. On the one hand, she's thinking she should report it to a superior, because the violation may lead to losses for the company. **(i)** But this will certainly have an effect on their friendship. **(ii)** But he may stop before he actually causes the losses. **(iii)** Also, she will be an accomplice if she doesn't say anything. On the other hand, she could talk to Ben about it first before reporting it to their superior.

References

Altmann, G., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*, 247–264.

Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, *30*, 191–238.

Arai, M., & Keller, F. (2013). The use of verb-specific information for prediction in sentence processing. *Language and Cognitive Processes*, *28*, 525–560.

Arnold, J. E. (1998). *Reference form and discourse patterns*. Ph.D. thesis Stanford University.

Asher, N. (1993). *Reference to Abstract Objects in Discourse*. Dordrecht: Kluwer.

Asher, N., & Lascarides, A. (2003). *Logics of conversation*. Cambridge University Press.

Asher, N., & Vieu, L. (2005). Subordinating and coordinating discourse relations. *Lingua*, *115*, 591–610.

Asr, F. T., & Demberg, V. (2015). Uniform information density at the level of discourse relations: Negation markers and discourse connective omission. In *IWCS 2015* (pp. 118–128).

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.

Baroni, M., Bernardini, S., Ferraresi, A., & Zanchetta, E. (2009). The WaCky wide web: a collection of very large linguistically processed web-crawled corpora. *Language resources and evaluation*, *43*, 209–226.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278.

Bates, D., & Sarkar, D. (2007). The lme4 package. *R package version*, *2*.

Boland, J. E., Tanenhaus, M. K., Garnsey, S. M., & Carlson, G. N. (1995). Verb argument structure in parsing and interpretation: Evidence from wh-questions. *Journal of Memory and Language*, *34*, 774–806.

Canestrelli, A. R., Mak, W. M., & Sanders, T. J. (2013). Causal connectives in discourse processing: How differences in subjectivity are reflected in eye movements. *Language and Cognitive Processes*, *28*, 1394–1413.

Carlson, L., Marcu, D., & Okurowski, M. E. (2003). Building a discourse-tagged corpus in the framework of rhetorical structure theory. In *Current and new directions in discourse and dialogue* (pp. 85–112). Springer.

Clifton, C., Frazier, L., & Connine, C. (1984). Lexical expectations in sentence comprehension. *Journal of Verbal Learning and Verbal Behavior*, *23*, 696–708.

Cristea, D., & Webber, B. (1997). Expectations in incremental discourse processing. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics* (pp. 88–95). Association for Computational Linguistics.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*, 1117–1121.

Dery, J. E., & Koenig, J. P. (2015). A narrative-expectation-based approach to temporal update in discourse comprehension. *Discourse Processes*, *52*, 559–584.

Drenhaus, H., Demberg, V., Köhne, J., & Delogu, F. (2014). Incremental and predictive discourse markers: ERP studies on German and English. In *Proceedings of the 36th Annual Conference of the Cognitive Science Society (CogSci)*.

Ehrlich, K. (1980). Comprehension of pronouns. *The Quarterly Journal of Experimental Psychology*, *32*, 247–255.

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, *41*, 469–495.

Garnham, A., Traxler, M., Oakhill, J., & Gernsbacher, M. A. (1996). The locus of implicit causality effects in comprehension. *Journal of Memory and Language*, *35*, 517–543.

Hobbs, J. R. (1979). Coherence and coreference. *Cognitive Science*, *3*, 67–90.

Hobbs, J. R. (1985). *On the coherence and structure of discourse*. Stanford: Center for the Study of Language and Information.

Holler, A., & Irmen, L. (2007). Empirically assessing effects of the right frontier constraint. In *Anaphora: Analysis, algorithms and applications* (pp. 15–27). Springer.

Hovy, E., & Maier, E. (1995). *Parsimonious or profligate: How many and which discourse relations?*. University of Southern California: Unpublished manuscript.

Jasinskaja, K., & Karagjosova, E. (submitted). Rhetorical relations. In L. Matthewson, C. Meier, H. Rullmann, & T. E. Zimmermann (Eds.), *The Companion to Semantics*. Oxford: Wiley.

Kehler, A. (2002). *Coherence, reference, and the theory of grammar*. CSLI publications Stanford.

Kehler, A., Kertz, L., Rohde, H., & Elman, J. L. (2008). Coherence and coreference revisited. *Journal of Semantics*, *25*, 1–44.

Kim, C. S. (2015). Presupposition satisfaction, locality and discourse constituency. In *Experimental Perspectives on Presuppositions* (pp. 109–134). Springer.

Köhne, J., & Demberg, V. (2013). The time-course of processing discourse connectives. In *Proceedings of the 34th Annual Meeting of the Cognitive Science Society (CogSci)*.

Koornneef, A. W., & Sanders, T. J. M. (2013). Establishing coherence relations in discourse: the influence of implicit causality and connectives on pronoun resolution. *Language and Cognitive Processes*, *28*, 1169–1206.

Koornneef, A. W., & Van Berkum, J. J. (2006). On the use of verb-based implicit causality in sentence comprehension: Evidence from self-paced reading and eye tracking. *Journal of Memory and Language*, *54*, 445–465.

Kuperberg, G. R. (2016). Separate streams or probabilistic inference? what the N400 can tell us about the comprehension of events. *Language, Cognition and Neuroscience*, *31*, 602–616.

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*, 32–59.

Mak, W. M., & Sanders, T. J. M. (2013). The role of causality in discourse processing: Effects of expectation and coherence relations. *Language and Cognitive Processes*, *28*, 1414–1437.

Mann, W. C., & Thompson, S. A. (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text-Interdisciplinary Journal for the Study of Discourse*, *8*, 243–281.

Pickering, M. J. (1994). Processing local and unbounded dependencies: A unified account. *Journal of Psycholinguistic Research*, *23*, 323–352.

Polanyi, L. (1988). A formal model of the structure of discourse. *Journal of Pragmatics*, *12*, 601–638.

Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A. K., & Webber, B. L. (2008). The Penn Discourse TreeBank 2.0. In *Proceedings*

of the 6th International Conference on Language Resources and Evaluation (LREC). Citeseer.

Pyykkönen, P., & Järvikivi, J. (2009). Activation and persistence of implicit causality information in spoken language comprehension. *Experimental Psychology*, (pp. 1–12).

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria. URL: `http://www.R-project.org` ISBN 3-900051-07-0.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372–422.

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

Rohde, H., Levy, R., & Kehler, A. (2011). Anticipating explanations in relative clause processing. *Cognition*, *118*, 339–358.

Sanders, T. J., Spooren, W. P., & Noordman, L. G. (1992). Toward a taxonomy of coherence relations. *Discourse processes*, *15*, 1–35.

Staub, A., & Clifton, C. (2006). Syntactic prediction in language comprehension: Evidence from *either...or*. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 425–436.

Stevenson, R., Knott, A., Oberlander, J., & McDonald, S. (2000). Interpreting pronouns and connectives: Interactions among focusing, thematic roles and coherence relations. *Language and Cognitive Processes*, *15*, 225–262.

Stewart, A. J., Pickering, M. J., & Sanford, A. J. (2000). The time course of the influence of implicit causality information: Focusing versus integration accounts. *Journal of Memory and Language*, *42*, 423–443.

Tyler, J. (2014). Prosody and the interpretation of hierarchically ambiguous discourse. *Discourse Processes*, *51*, 656–687.

Van Berkum, J. J., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 443–467.

Wolf, F., & Gibson, E. (2005). Representing discourse coherence: A corpus-based study. *Computational Linguistics*, *31*, 249–287.

Wolf, F., Gibson, E., & Desmet, T. (2004). Discourse coherence and pronoun interpretation. *Language and Cognitive Processes*, *19*, 665–675.

Xiang, M., & Kuperberg, G. (2015). Reversing expectations during discourse comprehension. *Language, Cognition and Neuroscience*, *30*, 648–672.

Xu, X., Jiang, X., & Zhou, X. (2015). When a causal assumption is not satisfied by reality: differential brain responses to concessive and causal relations during sentence comprehension. *Language, Cognition and Neuroscience*, *30*, 704–715.

Yoshida, M., Dickey, M. W., & Sturt, P. (2013). Predictive processing of syntactic structure: Sluicing and ellipsis in real-time sentence processing. *Language and Cognitive Processes*, *28*, 272–302.