GLOSSA psycholinguistics

Scholman, M. C. J., Rohde, H., & Demberg, V. (2024). Facilitation of a lexical form or a discourse relation: Evidence from pairs of contrastive discourse markers. *Glossa Psycholinguistics*, 3(1): 25, pp. 1–29. DOI: https://doi.org/10.5070/G60111353

eScholarship University of California

Facilitation of a lexical form or a discourse relation: Evidence from pairs of contrastive discourse markers

Merel C. J. Scholman, Saarland University, DE; Utrecht University, NL, m.c.j.scholman@lst.uni-saarland.de

Hannah Rohde, University of Edinburgh, UK, hannah.rohde@ed.ac.uk

Vera Demberg, Saarland University, DE, vera@lst.uni-saarland.de

Research has shown that people anticipate upcoming linguistic content, but evidence regarding expectations of specific lexical markers is mixed. We use the Dutch pair of discourse markers *Aan de ene kant…Aan de andere kant* ('On the one hand…On the other hand') and *Enerzijds… Anderzijds* (also equivalent to 'On the one hand…On the other hand') to test whether readers generate predictions of an upcoming contrast dependency based on the lexical marker for the first contrastive segment, and whether processing of the lexical marker for the second segment is facilitated (i) when the first segment contains a lexical marker to signal the upcoming contrast, and (ii) when that marker directly matches that of the second segment. In a self-paced reading study, we show that readers do generate expectations for upcoming discourse markers, in that the presence of a marker on the first segment facilitates processing of the marker on the second segment, but that a directly matching lexical form does not yield further facilitation. In an eye-tracking study, we replicate the facilitative effect of the first marker of a lexical pair on the processing of the second marker, and show that this effect occurs in immediate processing. These results establish expectation-driven effects at the discourse level in early reading time measures, showing that comprehenders are aware of the discourse dependency established by a discourse marker and are flexible in identifying and integrating discourse relations with different markers.

1. Introduction

A large body of research has provided evidence that readers and listeners generate expectations about upcoming content, and, in some cases, even predict specific words, during language comprehension (for a review, see Kuperberg & Jaeger, 2016). The predicted content tends to be processed faster once encountered, presumably due to the pre-activation of concepts or specific words. However, research on predictive processing has focused mainly on processing at the level of syntax and semantics. In the current contribution, we focus on discourse predictions, which require comprehenders to establish dependencies within and between sentences, as well as across passages. Discourse expectations can encompass a number of kinds of expectations, but in the current work, we use the term to refer to expectations about upcoming *discourse relations* (e.g., the meaning relations that connect two text segments, such as cause–consequence and contrast) as opposed to predictions about entities or situational content. We thus test discourse dependencies that hold between entire sentences, rather than sentence-internal dependencies that hold between syntactic elements or meaning-driven expectations about the upcoming mention of a relevant entity or situation.

Previous work suggests that comprehenders can indeed make predictions about upcoming discourse relations (e.g., Asr & Demberg, 2020; Barthel et al., 2022; Crible, 2021; Dery & Koenig, 2015; Hoek et al., 2020; Köhne-Fuetterer et al., 2021; Rohde & Horton, 2014; Scholman et al., 2020; Schwab & Liu, 2020; Xiang & Kuperberg, 2015; Yi & Koenig, 2021). Such relations can be explicitly signalled by discourse connectives and cue phrases, such as *although* or *as a result*, but they need not be. Most studies have focused on whether readers can use a discourse connective to predict upcoming dependencies and upcoming content. For example, Xiang and Kuperberg (2015) found that the presence of the connective *even so* in passages like (1) can reverse comprehenders' content expectations quickly, thereby making comprehenders more receptive to the mention of less typical responses to failing a test.

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

In the current work, we focus on the processing of the discourse markers themselves, testing whether prior material can induce facilitation of the second discourse marker and, furthermore, whether this facilitation pertains to an upcoming lexical item (i.e., a particular form of discourse marker) or the discourse relation itself (i.e., a specific type of relation is predicted, but any marker that expresses that relation can suffice). This study builds on findings reported in Scholman et al. (2017) and Scholman (2019), who investigated discourse predictions using the discourse markers *On the one hand* (OT1H) and *On the other hand* (OTOH) in English. These markers can be used to overtly express two contrasting perspectives, which we refer to as *contrast1* and *contrast2* (see (2)). The marker OT1H signals to the comprehender that a contrast2 will follow, an expectation

which typically must be satisfied for the passage to constitute a felicitous discourse. However, there is variability in the position and marking of contrast2: contrast2 does not need to follow contrast1 immediately (i.e., intervening content can occur), and neither contrast1 nor contrast2 need to be marked overtly with OT1H or OTOH, respectively.[1]

(2) Helen found some signs that Ben, her colleague and friend, has violated the company rules.
On the one hand,$_{[OT1H]}$ she's thinking she should report it to a superior.$_{[Contrast1]}$
On the other hand,$_{[OTOH]}$ she could talk to Ben about it first.$_{[Contrast2]}$

Scholman et al. (2017) tested passages like (2) and found that participants can generate fine-grained predictions of upcoming discourse dependencies based on OT1H: in a story continuation and eye-tracking study, they found evidence that readers quickly integrate discourse connectives and relations during processing. If a sentence intervened between the OT1H and OTOH segments and conveyed a plausible contrast2 (e.g., an example for (2): *But this will certainly have an effect on their friendship.*), the reading times of OTOH increased, indicating processing difficulty for encountering a subsequent contrastive segment at that point in the discourse. Hence, readers not only built, but also maintained, fine-grained predictions of upcoming contrast relations based on OT1H.

Results from Scholman et al. (2017), thus, indicate that readers exploit OT1H to facilitate processing of upcoming content. A logical extension of this finding would be that readers can also use OT1H to facilitate processing of an upcoming OTOH marker specifically: processing the first part of the two-part connective (OT1H) should automatically pre-activate the second part of the two-part connective (OTOH). This initial activation should result in facilitated processing of OTOH compared to a condition where OT1H is not present to pre-activate OTOH. Surprisingly, however, results from the eye-tracking study did not reveal such a facilitative effect: there was no difference in processing times of OTOH between a condition where OT1H was present versus absent (results for these additional conditions were reported in Scholman, 2019, Chapter 9).

The absence of a facilitative effect of OT1H on OTOH, combined with the finding that OTOH was dispreferred after a plausible intervening contrast (Scholman et al., 2017), would suggest that readers do pre-activate a discourse dependency (thus expecting an upcoming contrast2), but they might not pre-activate the specific marker used to typically signal that dependency (thus, not necessarily pre-activating OTOH specifically). In the current study, we follow up on this (lack of) facilitative effect of OT1H on the processing of OTOH, using similar pairs of Dutch markers.

---

[1] Scholman et al. (2017) found in a corpus study that in 15% of the OT1H-marked data, OT1H and OTOH are presented in adjacent sentences; and in 7% of the OT1H-marked data, other sentences intervene between the OT1H-sentence and the OTOH-sentence. In 17% of the data, OT1H is not followed by OTOH. In the majority of the data, however, OT1H and OTOH occur in the same sentence (61%).

1.1 Prediction and integration

Predictions play a crucial role in language processing, enabling efficient comprehension by leveraging contextual cues and prior knowledge (for a review, see Kuperberg & Jaeger, 2016). Predictions are specific anticipations about upcoming linguistic elements, such as words, formed in real-time based on the immediate context. For instance, upon hearing "The boy went outside to fly a…", one might predict that a noun is likely to follow, that this noun will refer to an object that can be flown, and perhaps even that the noun might take the form of the word *kite*, due to its high probability given the context. These predictions facilitate rapid processing by pre-activating neural representations of the anticipated elements, thereby allowing comprehenders to perform some of the processing ahead of time and, thus, reducing the cognitive load when these elements are encountered.

The focus on prediction in language processing that is currently popular in the psycholinguistic literature differs from more traditional language processing views, which focussed instead on how comprehenders integrate incoming input with previous input (Ferreira & Chantavarin, 2018; Pickering & Gambi, 2018). Specifically, *integration* refers to the process by which new linguistic information is incorporated into the existing mental representation of the sentence or discourse (i.e., the context). For example, upon hearing "The boy went outside to fly a…", comprehenders would activate properties of the context that are relevant to *kite* (that it describes a flying event, and that a boy is the agent of this event). Crucially, however, comprehenders would not pre-activate aspects of the linguistic representation of the word *kite*. These are only activated once comprehenders encounter *kite*, and it is only at that point that they are facilitated, because it is easier to integrate the meaning of the newly encountered *kite* with the already processed context (compared to the meaning of a different word, such as *airplane*) (Pickering & Gambi, 2018).

It can be difficult to distinguish prediction from integration, and, in particular, to find evidence that is compatible with prediction but not integration (see Pickering & Gambi, 2018). In fact, many of the reported findings supporting a prediction account could also be due to integration. In relation to the current article, we can view a hypothesized facilitative effect of OT1H on OTOH as arising from any number of mechanisms: a specific prediction of OTOH, a general expectation of a contrastive marker, or easier integration of a (specific or general) contrastive marker due to the comprehender's mental representation already containing OT1H. In our case, the results from the current studies cannot distinguish between prediction and integration, and so we will largely frame the discussion of the effects found in this study in terms of facilitated processing, though in what follows, we describe a number of existing findings in terms of predictions and expectations, to align with the language used in those studies. Ferreira and Chantavarin (2018) address this indeterminacy regarding attributing results to prediction or integration by viewing them as two sides of the same coin: the rich representation that is created as a result of integration of prior material enables the comprehender to prepare to receive new information, which will lead to precise predictions in (rare) cases of strong constraint. We will return to this in Section 5.

1.2 Lexical versus content-based predictions

Many studies have demonstrated that people can predict some information about upcoming words during comprehension. The question that arises is whether these expectations take the form of specific lexical predictions, which involve the activation of the semantic or grammatical features of a word, or even the exact form, before it appears in the discourse (for a review, see Nieuwland et al., 2018). Results from a range of studies are taken to indicate that comprehenders can indeed predict specific lexical items in certain contexts (e.g., DeLong et al., 2005; Fleur et al., 2020; Foucart et al., 2014; Fruchter et al., 2015; Husband, 2022; Ito et al., 2016, 2020; Kim & Lai, 2012; Laszlo & Federmeier, 2009; Martin et al., 2018; Nicenboim et al., 2020; Otten & Van Berkum, 2009; Otten et al., 2007; Thornhill & Van Petten, 2012; Van Berkum et al., 2005; Wicha et al., 2003, 2004). These studies focus mainly on predictions of gender noun class or phonological form. For example, De-Long et al. (2005) presented a seminal study on specific form prediction, using a paradigm in which semantic plausibility was linked with a non-semantic lexical feature: they exploited the phonological regularity of the *a/an* alternation in English (*an* precedes nouns beginning with vowel sounds, and *a* precedes nouns beginning with consonant sounds). Their results showed N400 effects for unexpected articles (e.g., on *an* for (3)), which was taken to indicate that readers can pre-activate individual, specific lexical forms in a graded fashion in correspondence to that word's probability.

(3) The day was breezy so the boy went outside to fly [a kite / an airplane].

Van Berkum et al. (2005) found similar results in an ERP paradigm and a self-paced reading (SPR) paradigm, using Dutch materials. Thus, there is evidence from electrophysiological and behavioral paradigms suggesting that comprehenders are able to make specific lexical predictions.

However, other studies have challenged this view, and suggested that comprehenders may rely more on general expectations and statistical regularities in a language, rather than precise lexical predictions based solely on individual word forms (e.g., Ito et al., 2017; Kochari & Flecken, 2019; Koornneef, 2021; Luke & Christianson, 2016; Nieuwland et al., 2018). For example, Nieuwland et al. (2018) failed to replicate the effect found in DeLong et al. (2005) in a large-scale replication study. Their results showed that listeners' brain activity was modulated by general syntactic and semantic constraints, such as word category and thematic roles, rather than specific lexical predictions. Moreover, Koornneef (2021) was unable to replicate the lexical prediction results from van Berkum et al.'s (2005) SPR study in two eye-tracking experiments. Koornneef (2021) took these results to indicate that only highly-constrained contexts in which readers process the incoming information at a relatively slow pace (as is typical for SPR designs) will induce strong, all-or-none lexical predictions (see also Luke & Christianson, 2016).

Overall, the available evidence regarding lexical predictions indicates that comprehenders can make specific, rather than merely conceptual, predictions, but other factors, such as general

expectations and statistical regularities, also come into play. The current study extends these findings by studying predictions at the discourse coherence level, exploiting statistical regularities of the co-occurrence of markers. This will be studied in both a self-paced reading paradigm and an eye-tracking paradigm.

1.3 Discourse-level predictions

Research on discourse has shown that comprehenders make predictions about upcoming relations between the ideas expressed in clauses or sentences, referred to as *discourse relations* (e.g., Asr & Demberg, 2020; Barthel et al., 2022; Crible, 2021; Dery & Koenig, 2015; Hoek et al., 2021; Köhne-Fuetterer et al., 2021; Rohde & Horton, 2014; Scholman et al., 2020; Schwab & Liu, 2020; Xiang & Kuperberg, 2015; Yi & Koenig, 2021). Comprehenders are able to predict specific discourse content or discourse relations based on discourse connectives, such as *because*, as well as other discourse relational signals, such as negation. For example, Köhne-Fuetterer et al. (2021) found in a visual world paradigm that the presence of the connective *however* in passages such as (4) (originally in German) can quickly reverse comprehenders' expectations of an expected result (getting something sweet to eat), thereby making an unexpected concession (getting something savoury to eat) expected. Further, ERP results using English data indicated that this mental representation update occurred immediately after encountering the concessive connective *however* (Köhne-Fuetterer et al., 2021).

(4) Marc fancies a small snack. He feels like having something sweet. [However], he gets…

Although discourse is more flexible in terms of dependencies (i.e., dependencies can occur within and between sentences, and even across passages) than sentence-internal syntax, most of the established prediction effects in discourse were found using contexts where the placement of the signal and the content was less flexible. For example, *however* leads the reader to expect an upcoming concessive relation, which can only be expressed in the clause immediately following the marker *however* (Köhne-Fuetterer et al., 2021). Similarly, for the pair *either…or*, the lexical cue *either* signals to the reader to expect an upcoming alternative, and, specifically, that this alternative will be marked by *or* and will appear in the same sentence (Staub & Clifton, 2006). Much of the prior research showing prediction effects, therefore, utilized lexical signals that are subject to structural constraints.

There are a few exceptions (Scholman, 2019; Scholman et al., 2017; Schwab & Liu, 2020). As discussed in Section 1, eye-tracking results revealed no difference in processing times of *On the other hand* between a condition where *On the one hand* was present versus where this cue was not present (Scholman, 2019). However, unlike Scholman (2019), Schwab and Liu (2020) did find evidence of connective facilitation, using a pair of similarly flexible German and English discourse markers, namely, *zwar…aber* and *true/sure…but* (see (5)). In a self-paced reading study,

they found that reading times on the connective *aber* ('but') were faster when the connective was preceded by the lexical cue *zwar* ('sure/true'), compared to when it was not. They also found a similar effect in English on the spillover region.

(5) James likes to run outdoors. [**True**,Ø] he has a treadmill in the living room, **but** he often jogs in parks.

Note, however, that even though *zwar…aber* and *true…but* do allow for flexibility in when the second argument will appear in the discourse, in Schwab and Liu (2020), these markers consistently occurred together in the same sentence. It therefore remains an open question whether these facilitation effects also arise across sentences. Across sentences, the possible relationships to be established are much more flexible than those afforded by sentence-internal syntactic and lexical-semantic constraints. Given this flexibility, readers might be less motivated to manage expectations regarding cross-sentence relationships.

Furthermore, in Schwab and Liu (2020), participants were asked to provide a meta-level naturalness rating after having read every item, which could have had an impact on reading goals and, consequently, on discourse processing and inferencing (Noordman et al., 1992). For example, providing naturalness ratings might have drawn the reader's attention to how the prompts are linguistically encoded, and readers might consequently have become more sensitive to the pair of markers and their co-occurrence in the prompts. Another open question is, thus, whether facilitative processing is engaged during reading, even when additional meta-level tasks are removed. Finally, since the study reported in Schwab and Liu (2020) was a self-paced reading paradigm, the methodology leaves open the question whether such expectation-driven effects can be seen in more natural reading paradigms and in early reading time measures. Koornneef (2021), for example, argues that SPR leads to stronger prediction effects because readers adapt to the word-by-word presentation mode by resorting to a more incremental processing strategy, in which they more rapidly use the information afforded by each word to generate a lexical prediction than they would do in unconstrained reading. The current study addresses these issues while extending prior work on discourse-level facilitative processing.

1.4 Current study

The current study investigates whether comprehenders use a discourse marker as a cue to facilitate the processing of specific upcoming words or to facilitate the processing of a type of discourse relation more broadly. We test this in Dutch, because this language has multiple pairs of markers that can express contrast: *aan de ene kant* ('on the one side', referred to from now on as *OT1H*, in keeping with Scholman et al. (2017)) and *aan de andere kant* ('on the other side', *OTOH*), and *enerzijds* (also equivalent to 'on the one hand') and *anderzijds* (also meaning 'on the other hand'). *Enerzijds* and *anderzijds* will be referred to as *EZ* and *AZ*, respectively. The two pairs

are highly similar in meaning and usage. As we will see in Section 2, these markers typically co-occur with their parallel partner to signal a contrastive dependency, but mixing the pairs (e.g., EZ~OTOH) can happen in natural language, as well.

For the scope of the research presented here, we assume a notion of discourse *contrast* in terms of an evaluation of a series of alternatives (thus, not limited to binary opposition). Logically speaking, contrast is the juxtaposing of multiple alternatives, which can be positions, arguments, options, or other propositional content (for a detailed discussion of the definition of contrast, see Molnár, 2002; Umbach, 2004).

There are many different ways in which contrast can be expressed, such as antonyms (*good-bad*), grammatical structure (i.e., parallelism), or prototypical connectives (e.g., *but, whereas*). The current study focuses on two pairs of semi-fixed discourse markers (OT1H~OTOH and EZ~AZ), which are lexical signals that facilitate the establishment of a contrast between two given alternatives (see also Hinnell, 2019). The markers *aan de ene kant* and *enerzijds* trigger a set of alternatives, and thereby allow the reader to establish an expectation for two contrastive perspectives (referred to as *contrast1* and *contrast2*); in addition, readers might also establish an expectation for a marker expressing contrast2. In turn, the markers *Aan de andere kant* and *Anderzijds* allow the reader to integrate the upcoming content as contrasting with the previously provided contrast1. In the absence of OT1H/EZ, however, it is also left up to the reader to identify an earlier provided contrast1 as such, which may require the reader to re-analyse the earlier encountered content upon encountering OTOH, as well as update possible expectations for the upcoming content of contrast2.

To illustrate this, consider (6), which contains three variants of marking for contrast1. In (6c), contrast1 is unmarked. Before encountering OTOH, readers might expect the discourse to continue with an explanation for why Helen feels she needs to speak to her supervisor, or a conclusion of the outcome of the situation. When contrast1 is marked ((6a) and (6b)), readers can anticipate a contrast dependency, which should facilitate processing of the discourse marker signalling contrast2 (no cue > EZ/OT1H). However, if comprehenders use the contrast1 marker to facilitate processing of specific discourse markers, the facilitative effect of OT1H on the processing of OTOH should be greater than that of EZ on OTOH (no cue > EZ > OT1H).

(6) Helen kwam erachter dat Ben, haar collega en vriend, de bedrijfsregels had overtreden.
   a. **Aan de ene kant** denkt ze dat ze het bij een leidinggevende moet melden. [OT1H]
   b. **Enerzijds** denkt ze dat ze het bij een leidinggevende moet melden. [EZ]
   c. Ze denkt dat ze het bij een leidinggevende moet melden. [no cue]
   Aan de andere kant kon ze ook eerst met Ben erover praten.
   'Helen found some signs that Ben, her colleague and friend, has violated the company rules. [*Aan de ene kant / Enerzijds / ∅*] she's thinking she should report it to a superior. On the other hand, she could talk to Ben about it first.'

In what follows, we present corpus data on the marker pairings in natural language, showing that *aan de ene kant* is followed by it partner *aan de andere kant* and *enerzijds* is followed by its partner *anderzijds* in most cases. A lexical account predicts that comprehenders are sensitive to the relative frequencies with which the pairs of markers occur in the language (Asr & Demberg, 2020), and, therefore, that processing of *aan de andere kant* is facilitated when it is preceded by *aan de ene kant* rather than *anderzijds*. This is tested in two experiments. Experiment 1 is a self-paced reading study, in which we show that readers process OTOH faster after having read a contrastive marker signalling contrast1, but no difference was found between the facilitative effect of *aan de ene kant* and *enerzijds*. Experiment 2 presents an eye-tracking study focusing on the time-course of the facilitative effect of *aan de ene kant* on *aan de andere kant*. We find an effect of *aan de ene kant* on early processing of *aan de andere kant*. Taken together, these results show that the presence of a discourse marker leads to facilitated processing of the discourse relation, rather than merely surface-based lexical facilitation. We argue that these results establish expectation-driven effects even in the absence of any task-specific goals and in the earliest possible measures, showing comprehenders' awareness of the relational dependencies established by specific discourse markers.

2. Corpus distribution of OT1H and OTOH

We analyzed how the markers OT1H and EZ are distributed in naturally occurring texts, to better understand the degree to which OT1H and EZ are predictive of OTOH. We analysed fragments containing at least one of the markers[2] along with a context of 150 words, in the Dutch reference corpus SoNaR, a 500 million-word corpus of contemporary written Dutch (Oostdijk et al., 2013). This amounted to 3,553 instances of OT1H and 12,419 instances of EZ.

The data shows that the appearance of OT1H or EZ does not wholly determine the presence of OTOH or AZ. **Table 1** displays the frequencies of OT1H and EZ in the corpus. Examples of each configuration are available on OSF.[3] The majority of the OT1H-marked data is followed by OTOH (73% of all OT1H-marked instances), and an even larger proportion of EZ-marked data is

Table 1: Percentage distributions (and raw counts) of OT1H and EZ co-occurring with their (non-)parallel partners in natural language.

	OTOH		**AZ**		**Neither**	
OT1H	73%	(2,594)	3%	(106)	24%	(853)
EZ	1%	(124)	88%	(10,929)	11%	(1,366)

[2] Instances where OT1H was followed by *van* ('of') were removed from the dataset; these correspond to the usage of the marker as a location indicator in Dutch (e.g., *aan de andere kant van de straat*, 'on the other side of the street').

[3] https://osf.io/y5afc.

followed by AZ (88%). This supports our hypothesis that OT1H is predictive of OTOH, and EZ is predictive of AZ, in naturalistic data.

However, a substantial proportion of OT1H-marked data is not followed by OTOH (27%), and the same goes for the EZ-marked data (12%). This suggests that natural text often requires readers to identify contrast2, even when it appears in a sentence that is not marked by the corresponding marker of the lexical pair. **Table 1** also shows that the pairs of markers can be mixed, although this does not occur frequently: 3% of OT1H-marked instances are followed by AZ, and 1% of EZ instances are followed by OTOH. Hence, it is unlikely that upon encountering a marker for contrast1, readers generate strong specific lexical predictions of the non-parallel marker, given that they do not frequently appear together.

Instances that do not contain OTOH or AZ are often marked by alternative connectives. We annotated a random sample of 200 instances marked by OT1H but not OTOH, and 200 instances marked by EZ but not AZ. For each instance, we annotated whether a contrast2-segment was present and which connective was used to mark contrast2.[4] For OT1H, we found that *maar* ('but') is the most common contrast2 marker when OTOH was not present (43.5%), followed by *anderzijds* and *tegelijkertijd* ('at the same time'). For EZ, *maar* ('but') is also the most common contrast2 marker (23%), followed by *daarnaast* ('besides'). In the EZ-sample, 19% of contrast2 is not marked explicitly, as in the following example (where 'the evolution of money supply' is not explicitly marked with *anderzijds*):

(7) Het cijfermateriaal dat aangeeft dat er tussen economische activiteit enerzijds en de evolutie van de geldhoeveelheid niet het minste verband bestaat, is overweldigend.
English gloss: *The numerical evidence that indicates that there between* <u>*economic activity*</u>$_{[contrast1]}$ ENERZIJDS *and* <u>*the evolution of the money supply*</u>$_{[contrast2]}$ *not the slightest connection exists, is overwhelming.*
Translation: 'The numerical evidence indicating that there is not the slightest connection between economic activity ENERZIJDS and the evolution of the money supply is overwhelming.'

In 15% of the annotated OT1H-subset (which would equal 3.6% of the full OT1H-marked dataset) and 6.5% of the EZ-subset (equalling 0.7% of the full EZ-marked dataset), contrast2 could not be identified at all. This can be attributed to the nature of the data: the corpus contains some informal, unfinished comments, such as interviews or forum conversations. In such less-structured discourse, there are instances where the topic shifted after contrast1 was provided.

[4] For this, we used the notion of contrast as defined in 1.4: contrast is the juxtaposing of two alternatives, which can be positions, arguments, options, or other propositional content. What constitutes contrast2 is, therefore, that it must present an alternative to the content expressed in contrast1.

In sum, the data indicates that OT1H and EZ are almost perfect predictors of contrast. More than 96% of OT1H markers (i.e., excluding the 3.6% of the dataset that does not contain contrast2) and 99% of EZ markers are followed by a contrast2, and 73% of OT1H-marked contrast1 are followed specifically by OTOH. These results attest to the relative activation strength of the pair, which should be reflected in facilitated processing of OTOH after processing OT1H, if comprehenders engage in generating predictions of upcoming discourse markers. By contrast, only 1% of EZ-marked contrast1 are followed by OTOH specifically, which would suggest that EZ should not facilitate processing of OTOH specifically as well as OT1H should, if readers generate strong lexical predictions rather than discourse-level predictions. Hence, an experimental design that compares the facilitative effect of OT1H as well as EZ on the processing of OTOH can provide insight into whether comprehension is facilitated by combined grammatical use (i.e., OTOH typically follows OT1H) or whether comprehension is facilitated by discourse-level integration rather than only the frequent pairing of the parallel marker pairs (i.e., OT1H and EZ both mark contrast, so any contrastive discourse marker signalling contrast2 should facilitate integration).

3. Experiment 1: Self-paced reading study

This web-based self-paced reading study tests whether comprehenders use discourse markers to facilitate processing of later content, and whether the facilitation is contingent on a particular form of contrast (i.e., a surface-based lexical facilitation), or whether this is actually a discourse-level facilitation. Under a lexical account, readers should find it easier to integrate specific discourse markers based on the form of the marker used to express contrast1, in which case we should see faster reading times on OTOH following OT1H compared to another form of OT1H (that can be used to mark contrast but does not form a lexical pair with OTOH). However, under what we will call the *relational account*, people should readily integrate a contrast relation rather than a particular surface form, in which case any contrastive marker should suffice.

3.1 Participants

A total of 138 native speakers of Dutch (age range 18–52 years; mean age 27 years; 59 female) participated in this experiment. Data from two additional participants were excluded because they didn't pass the comprehension question accuracy check (they scored $<70\%$). Participants were recruited via Prolific and received 3.00 GBP for their participation. They were unaware of the purpose of the experiment.

3.2 Materials

The experimental stimuli consisted of twelve passages. Each item consisted of an introductory sentence, followed by contrast1 and contrast2. The introduction or contrast1 included a linguistic

signal to indicate uncertainty about the situation (e.g., 'was considering', 'might', 'was thinking'). Contrast2 was always marked with OTOH; contrast1 was marked with OT1H, marked with EZ, or unmarked. The three conditions are illustrated in (8).

(8) De backpackster was haar trip naar Hawaii aan het plannen.
   a. **Aan de ene kant** leek het haar leuk om daar te surfen. [OT1H]
   b. **Enerzijds** leek het haar leuk om daar te surfen. [EZ]
   c. Het leek haar leuk om daar te surfen. [no cue]
   Aan de andere kant leek het haar ook leuk om te snorkelen in de oceaan.
   'The backpacker was planning her trip to Hawaii. [*Aan de ene kant / Enerzijds / ∅*] / she thought it would be fun to surf there. On the other hand, she thought it also seemed fun to snorkel in the ocean.

We adapted two aspects in our experiment compared to Scholman (2019). First, we decreased the proportion of items containing OTOH in the experiment by including fewer experimental items, and presenting these to a larger number of participants. This was done to preserve statistical power and minimize a possible effect of expectation adaptation throughout the study (Fine et al., 2013). Second, we removed intervening (non-contrastive) material that occurred between the pair, which was necessary for the design used in Scholman (2019), but was not necessary here. By removing this intervening material, we optimize the conditions for a facilitative effect of OT1H.

The items were interspersed with 60 filler items for an unrelated study focusing on causal and list relations. None of these fillers contained contrastive discourse connectives. The stimuli were counterbalanced across three lists, with each story appearing in a different condition in each list. The participants were randomly assigned to one of the lists.

3.3 Procedure

Participants were recruited via Prolific, after which they were directed to a website hosted by PCIbex (Schwarz & Zehr, 2021), where they completed the moving window self-paced reading experiment on their own laptop or desktop. Items were initially displayed as a fixation cross. After pressing the space bar, participants were presented with a series of horizontal lines; the length of the lines corresponded to the length of the regions. By pressing the space bar on their keyboard, participants revealed the next region of the item. Items were presented non-cumulatively; when a new region was revealed, the previous region was replaced by lines.

OTOH was always presented as an individual chunk and never occurred as the first or final region of a line. (9) illustrates the spatial configuration of target stimuli on the screen, with slashes demarcating the regions.

(9) De backpackster was haar trip naar Hawaii /
aan het plannen. / Aan de ene kant / leek het haar leuk /
om daar te surfen. / Aan de andere kant / leek het haar /
ook leuk / om te snorkelen / in de oceaan.
'The backpacker was her trip to Hawaii /
planning. / On the one hand, / she thought it would be fun /
to surf there. / On the other hand, / she thought /
it also seemed fun / to snorkel / in the ocean.'

Participants were presented with a verification statement after 25% of the items. An example of a verification statement for (9) is *De backpackster wilde surfen in Florida* ('The backpacker wanted to surf in Florida'). Participants responded to the statement by pressing either 'j' for TRUE or 'f' for FALSE on their keyboard. On two occasions during the experiment, participants were instructed to take a short break. When they were ready to continue, participants clicked a "proceed" button at the bottom of the screen. The entire study lasted, on average, 15 minutes.

3.4 Analysis procedure

Reading times were measured on two regions: *Aan de andere kant* (the critical OTOH region) and the three words following the critical region (the spillover region). The critical region is where a possible facilitative effect of OT1H is expected to be found: if readers anticipate OTOH based on OT1H, reading times of OTOH should be faster in the OT1H present condition than in the OTOH absent condition. The spillover region is of interest because (i) in SPR studies, effects can be delayed, and (ii) longer reading times on this region might also reflect difficulty updating the expectations in the OT1H-absent condition to integrate a previously unpredicted contrastive relation.

Prior to all analyses, items with passage reading times longer than 60 seconds (average reading time per item was 7 seconds) were removed (2 cases) because participants had likely taken a break during the presentation of these items. Reading times shorter than 80 ms (1 case) or longer than 5000 ms (1 case) were removed. Outliers were removed by excluding reading times on words more than 2.5 times the standard deviation from a participant's mean in a region (2.0% of the data points).

We analyzed the reading times on the OTOH region and the spillover region separately in linear mixed-effects models with log-transformed reading times.[5] The variable for condition was Helmert-coded to allow for a comparison between the no cue condition to the EZ and OT1H condition

[5] The effects reported based on transformed reading times were replicated using raw reading times, thus supporting the stability of the effects.

together, and the EZ and OT1H conditions. A centered covariate of trial order was included in the models to account for any variance due to participants' reading times speeding up over the course of experiment.[6] We present the analyses per region and apply Bonferroni corrections to account for multiple comparisons. The adjusted threshold for significance is $\alpha = .025$.

3.5 Results

Figure 1 shows the reading times on the critical and spillover region per condition. The results of the linear mixed-effects regression models are summarized in **Table 2**. On both the critical and the spillover region, the effect of condition was significant: OTOH and the spillover region were read faster when they followed EZ or OT1H, compared to when no cue was used to express contrast1. Crucially, while the numerical difference between the EZ and OT1H condition trends in the expected direction, there was no significant difference between reading times of OTOH and the spillover region when they were preceded by EZ or OT1H. This suggests that comprehenders readily interpret any type of contrastive marker to express contrast2, rather than having a preference for the parallel partner matching the contrast1 marker.

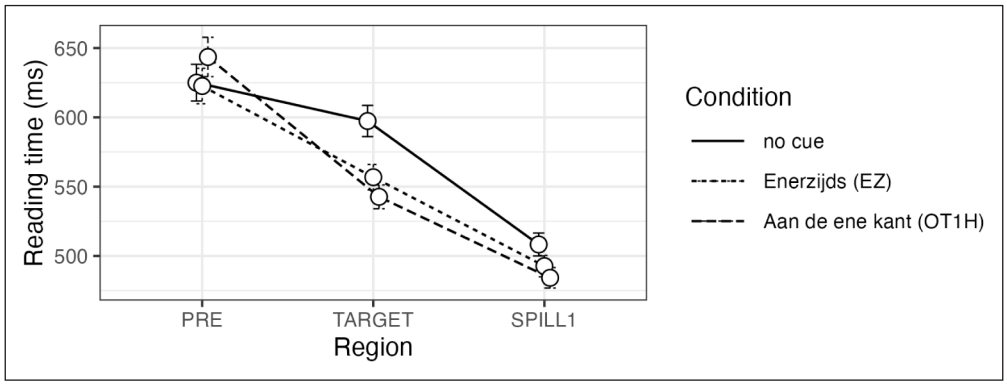

Figure 1: Mean reading times and error bars (SE) per region and condition, Exp. 1.

3.6 Follow-up: Bayesian analyses

The frequentist analysis provided above yielded non-significant p-values for the comparison between the EZ and OT1H conditions, indicating failure to reject the null hypothesis that the type of lexical contrast1 marker has no effect on reading times of OTOH. However, **Figure 1** does show a numerical facilitation trend of the OT1H condition, compared to the EZ condition. Moreover, non-significant p-values do not quantify evidence that there is no difference in the effect of EZ and OT1H on the processing of OTOH. Therefore, we follow up on these results by conducting a Bayesian reanalysis of the data.

[6] Model specification: `log(rt) ~ trial+cond + (1+cond|subject) + (1+cond|item)`.

Table 2: Regression coefficients and test statistics from linear mixed-effects models for the effect of condition for each region, Exp. 1. α significant at .025.

	Region							
	Critical				***Spillover***			
	β	**SE**	**t**	***p***	***β***	**SE**	**t**	***p***
(Intercept)	6.27	0.02	282.16		6.15	0.02	271.744	
Trial order	–0.00	0.00	–13.28	**<.001**	–0.00	0.00	–14.78	**<.001**
No cue vs. EZ & OT1H	–0.07	0.01	–5.47	**<.001**	–0.04	0.01	–2.84	**.02**
EZ vs. OT1H	–0.03	0.01	–1.90	.07	–0.02	0.01	–1.75	.09

The data were analysed using Bayesian mixed-effects regression models with the *brms* package in R (Bürkner, 2017), with a lognormal familiy, a Helmert-coded predictor for condition (no cue vs. EZ + OT1H, and EZ vs. OT1H), a covariate for trial order, and principled weakly-informative priors. Maximal random effects were estimated with an LKJ(2) prior on the random effects correlation (Lewandowski et al., 2009). Four chains were run, using 10,000 iterations (including 2,000 warm-up). We report posterior estimates and 95% credible intervals (CRI). All parameter estimates for the reported models converged with $\hat{R} = 1$.

Posterior distributions for the estimated parameters are depicted in **Figure 2**; the model results are shown in **Table 3**. We observed an effect of cue presence at the critical region as well as the spillover region, such that these regions were read faster when they were preceded by either EZ or OT1H. For the comparison between EZ and OT1H, the posterior shows a small effect on both regions, such that these regions were read faster after the OT1H condition than the EZ condition. Note, however, that the 95% CRI includes 0 (see also **Figure 2**), meaning there is a chance that a difference between OT1H and EZ is non-existent.

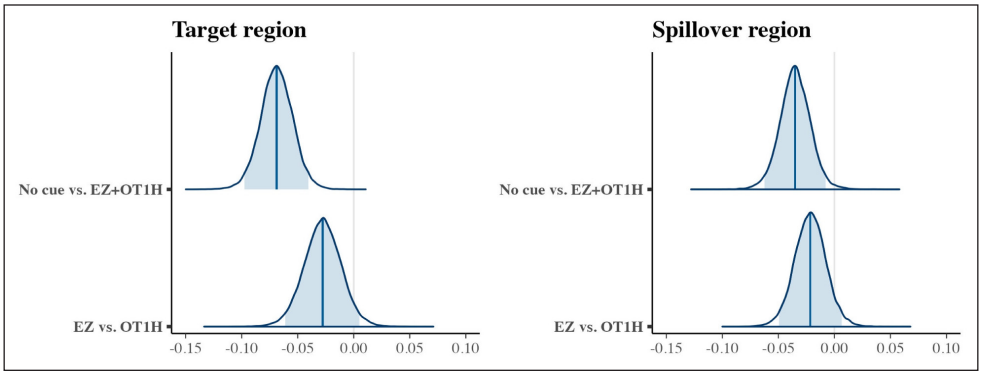

Figure 2: Posterior distributions for the comparisons between conditions. The dark blue line shows the median posterior effect estimate; the light blue area indicates the 95% credible interval.

Table 3: Coefficients from Bayesian mixed-effects model for the effect of condition for each region, Exp. 1.

	Region					
	Critical			***Spillover***		
	Posterior	**CRI low.**	**CRI up.**	**Posterior**	**CRI low.**	**CRI up.**
(Intercept)	6.27	6.23	6.32	6.15	6.10	6.20
Trial order	0.00	0.00	0.00	–0.00	–0.00	–0.00
No cue vs. EZ + OT1H	–0.07	–0.10	–0.04	–0.02	–0.04	–0.01
EZ vs. OT1H	–0.03	–0.06	0.00	–0.01	–0.03	0.00

To quantify the evidence in favour of these observed effects, we conducted Bayes factor analyses. Following state-of-the-art recommendations for Bayes factor computation (Schad et al., 2022), we employed bridge-sampling tools in *brms*, and used empirical priors for the fully-specified model. To derive these empirical priors, we obtained model fits for data from a study that is currently under submission, which compares reading times on *On the other hand* in English. Two conditions were present in this study: the *On the one hand*-cue was either present or absent. We used this distinction to derive the estimate for the no cue vs. EZ + OT1H condition in the current study, and additionally estimated the prior for the EZ/OT1H condition by dividing the prior for the cue present/absent condition – this corresponds to what can be considered a graded lexical prediction account, whereby a facilitative effect of OT1H over EZ, as well as a facilitative effect of EZ over no cue is expected (i.e., no cue > EZ > OT1H).[7]

For each region, we compared a fully-specified model (with both Helmert-coded conditions included) to reduced models in which the effects were assumed to be 0 – that is, the effect term for either the no cue vs. EZ + OT1H or EZ vs. OT1H was removed. The first comparison (no cue vs. EZ + OT1H) provided very strong evidence for a facilitative effect of having a contrast1 marker present compared to having no marker present on the critical OTOH region, and strong evidence for this effect on the spillover region (critical region: $BF_{10} = 44.54$ in favour of a model with both conditions; spillover region: $BF_{10} = 13.05$). Crucially, the second comparison (EZ vs. OT1H) provided anecdotal evidence for a facilitative effect of OT1H over EZ (critical region: $BF_{10} = 1.81$ in favour of a model with both conditions, compared to a model without the EZ vs. OT1H condition; spillover region: $BF_{10} = 2.17$). In other words, these results indicate that although it appears there might be a greater facilitative effect of OT1H compared to EZ (i.e., the alternative hypothesis is slightly favored), we do not have sufficiently strong evidence from the data to reject or accept this hypothesis.

[7] Specification of the informative prior: Intercept = normal(6.2, 0.6); no cue vs. EZ/ + OT1H = normal(–0.016, 0.012); EZ vs. OT1H = normal(–0.008, 0.012).

3.7 Discussion: Self-paced reading study

This study was designed to test whether comprehenders use OT1H to facilitate processing of specific lexical elements, or whether any (appropriate) lexical element marking contrast2 would suffice. The results provide tentative evidence for the latter: OTOH and the spillover region were read faster when contrast1 was marked, but there was only a weak numerical effect of OT1H in comparison to EZ. This indicates that comprehenders integrate new material at the discourse level after having encountered a marker for contrast1, with surface-based lexical facilitation playing a less strong role.

An open question is why this facilitative effect of a contrast1 cue did not appear in the eye-tracking study reported in Scholman (2019), which did not show a facilitative effect of OT1H on the processing of OTOH in English. One major difference between the two studies is the paradigm: the current study was a self-paced reading paradigm, whereas the study reported in Scholman (2019) was an eye-tracking study. In the eye-tracking paradigm, the presentation of the signal OT1H was not as visually salient as it was in this SPR experiment, where it appeared as an individual chunk. Moreover, participants are less likely to skip over the OTOH region in an SPR design, as they have to press a button to reveal each chunk. Further, it has been argued that chunk-by-chunk presentation, as is typical for SPR studies, may enable and encourage a greater reliance on predictions than natural reading (Koornneef, 2021; Luke & Christianson, 2016): mean reading times in eye-tracking research tend to be lower than single-word presentation times in SPR research. This could be attributed to the fact that self-paced reading times are not purely reading times; they also represent response times – they include the time a participant needs to plan and execute button presses. Comprehenders might adjust their reading process when they need to anticipate the next button press, as compared to when they move their eyes in a more natural setting. Therefore, it could be argued that participants might use the extra time afforded by the SPR design to try to predict the next word, which they might not do in more natural reading circumstances.

It is unclear whether the effect found in the current self-paced reading paradigm would replicate in a more natural reading paradigm. To further explore the effect of OT1H on OTOH, we repeat the reading task as an eye-tracking study in Experiment 2. We focus on the comparison between the OT1H present and OT1H absent conditions and, thus, leave out the EZ condition. This was done to keep the design more similar to that of Scholman (2019), and to gain more observations per condition per item, which allowed us to have greater statistical power.

4. Experiment 2: Eye-tracking study

In the previous study, we established that EZ and OT1H both lead to facilitated processing of OTOH. This section presents an eye-tracking study to investigate whether the facilitative effect of OT1H on the processing of OTOH can be replicated in a more naturalistic reading paradigm, and to gain further insight into whether such effects can already be seen in early reading time measures.

4.1 Participants

Eighty-five native speakers of Dutch (age range 18–55 years; mean age 24 years; 71 female) took part in this experiment. Participants were recruited from the Utrecht University community. All participants had normal or corrected-to-normal vision. Participants were paid for their participation and were unaware of the purpose of the experiment.

4.2 Materials and procedure

The materials were identical to those used in the self-paced reading study, without the EZ condition. Hence, this study used 12 experimental items, and the condition consisted of two levels: OT1H absent versus OT1H present. These items were interspersed with the same filler items used in Experiment 1.

Participants were tested individually. They were seated at a distance of approximately 60 cm from the monitor and rested their head on a chin-rest. Eye movements were recorded with SR Research Eyelink 1000 at a sampling rate of 500 Hz. The experiment lasted approximately 30 minutes. Each session started with an oral instruction, after which the eye-tracker was adjusted, if necessary. A 9-point calibration procedure was then performed. Upon successful calibration, the experiment started with two 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 at the position of the first word of the first sentence. Participants were instructed to fixate this mark, after which the story appeared. The stories were presented randomly and in their entirety on the screen. The critical region (*Aan de andere kant*) never appeared at the beginning or end of a line. A verification statement about the story followed 25% of the items, to ensure that the participants read the passages carefully. Participants indicated whether the statement was correct or incorrect by pressing a button on a button box.

4.3 Analysis procedure

For analysis purposes the sentences were divided into three regions, as illustrated in 10:

(10) (…) / andere ondernemer. $_{pre\text{-}critical}$ / Aan de andere kant $_{critical}$ / wilde hij graag $_{spillover}$ / (…)
(…) / *other entrepreneur.* $_{pre\text{-}critical}$ / *On the other hand* $_{critical}$ / *he would like* $_{spillover}$ / *(…)*

We also consider the pre-critical region in addition to the critical and spillover regions, because the critical region is consistent and quite salient; it is possible that readers parafoveally process the critical region while they fixate on the pre-critical region.

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). Note that we focus on first pass duration rather than first fixation duration as an early measure, because of the size of the critical region. *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 removed from the dataset. Fixations shorter than 80 ms (31 cases) and longer than 5,000 ms (3 cases) were recoded as missing data. In all reading time measures, outliers were identified as reading times more than 2.5 standard deviations from the participant's mean in a region (0.76% of the data points for the first pass duration, 2.16% for the regression path duration, and 0.68% for the total reading time duration). These outliers were recoded as missing data.

Log-transformed reading times were modelled per measure on the pre-critical region, the OTOH region and the spillover region separately in linear mixed-effects models.[8] The variable for condition was deviation-coded and included as fixed effect, together with a centered covariate of trial order, to account for presentation order of the item in the experiment.[9] We present the analyses per region and measure, and we apply Bonferroni corrections for multiple comparisons. The adjusted threshold for significance is $\alpha = .005$.

4.4 Results

Figure 3 shows the mean reading time measures per condition and region. **Table 4** displays the model results for all measures and regions.

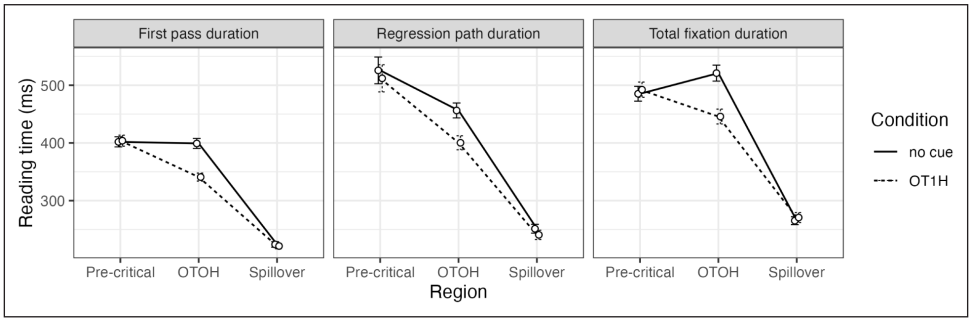

Figure 3: Mean reading times per measure, condition and region, Exp. 2. The error bars represent the 95% confidence interval.

[8] The effects that were found based on transformed reading times were replicated using raw reading times, thus supporting the stability of the effects.

[9] Model specification: `log(rt) ~ trial+cond + (1+cond|subject) + (1+cond|item).`

Table 4: Regression coefficients and test statistics from linear mixed-effects models for the effect of condition for each measure, Exp. 2. α significant at $<.005$.

	Region											
	Pre-critical				***Critical***				***Spillover***			
	β	**SE**	**t**	***p***	***β***	**SE**	**t**	***p***	***β***	**SE**	**t**	***p***
First pass duration												
Intercept	5.89	0.07	78.74		5.82	0.03	171.36		5.32	0.03	178.68	
Trial order	–0.00	0.00	–4.16	**<.001**	–0.00	0.00	–2.29	.02	–0.00	0.00	–2.02	.04
Condition	–0.02	0.03	–0.80	.44	–0.15	0.03	–5.19	**<.001**	–0.00	0.03	–0.12	.91
Regression path duration												
Intercept	6.05	0.07	91.51		5.92	0.04	163.69		5.38	0.03	154.27	
Trial order	–0.00	0.00	–2.95	**<.005**	–0.00	0.00	–2.49	.01	–0.00	0.00	–1.76	.08
Condition	–0.05	0.03	–1.63	.10	–0.14	0.04	–3.89	**<.001**	–0.03	0.03	–0.91	.37
Total fixation duration												
Intercept	6.06	0.07	87.30		6.04	0.04	158.38		5.46	0.04	143.24	
Trial order	–0.00	0.00	–3.95	**<.001**	–0.00	0.00	–4.25	**<.001**	–0.00	0.00	–1.66	.10
Condition	–0.01	0.03	–0.23	.82	–0.18	0.04	–4.40	**<.005**	–0.01	0.03	–0.35	.73

At the critical region, the results showed a significant difference between the conditions with and without OT1H in the first pass duration, regression path duration and total reading time duration: the reading times on OTOH were shorter when OT1H was present, compared to when there was no cue. The results showed no significant effect of condition in the pre-critical region or spillover region.

4.5 Discussion: Eye-tracking study

The results showed a facilitative effect across the eye-tracking measures: first pass duration, regression path duration, and total fixation duration on OTOH were shorter when OT1H was present, compared to when it was not. These results indicate that OT1H allows readers to anticipate an upcoming contrast, which then facilitates the processing of OTOH.

5. General discussion and conclusion

The work reported here investigated whether discourse markers facilitate discourse processing at a lexical level or at the content level (irrespective of the type of contrastive marker that is used). This was done by manipulating the co-occurrence of pairs of discourse markers in

Dutch: *Aan de ene kant* (OT1H) and *Aan de andere kant* (OTOH), and *Enerzijds* (also equivalent to *on the one hand*, referred to as *EZ*) and *Anderzijds* (also meaning 'on the other hand', referred to as *AZ*).

A self-paced reading study revealed a facilitative effect of the presence of a discourse marker on the processing times of the second marker during online processing: OTOH was read faster when it was preceded by a contrastive marker signalling contrast1, compared to when it was not. Crucially, a Bayesian analysis revealed only anecdotal evidence supporting a difference between a condition with OT1H and a condition with an alternative contrast1 marker, EZ. These results indicate that markers facilitate processing at the discourse level (i.e., of upcoming discourse relations), rather than solely at a surface-based level. Statistical frequencies of co-occurrences of the marker pairs appear to have only a weak effect on reading times.

An eye-tracking study provided more insight into the time-course of the facilitative effect, showing a facilitative effect of OT1H already in an early processing measure. This is consistent with a prediction account of language processing. Note that the studies reported here included a relatively large sample of participants who were presented with a smaller number of items, to address the risk of repeated exposure effects. Given that the results appear to replicate across methods, the chance of a spurious effect due to sparse data is reduced. We take this to support the robustness and reliability of the effects found in the current study. Taken together, these results show that comprehenders are able to pre-activate upcoming discourse markers and relations.

5.1 Connection to previous literature

The conclusion that comprehenders can pre-activate upcoming discourse markers and relations is in line with findings reported by Schwab and Liu (2020), and extends their results by showing that these expectations occur across sentences, in the absence of additional meta-level tasks, and in early processing measures during more natural reading.

The next question that arises is why this finding did not appear in earlier work studying the facilitative effect of OT1H on OTOH (Scholman, 2019; Scholman et al., 2017). Across a series of studies, Scholman and colleagues found that comprehenders were able to build fine-grained discourse structures based on OT1H, but no evidence was found that the presence of OT1H facilitated processing of OTOH (Scholman, 2019). Several explanations seem possible: First, the proportion of items containing OTOH differed significantly between the current study and the previous work. The proportion of OTOH-items was 43% in Scholman (2019), compared to 16% in the current study. It is possible that the frequency of OTOH in the earlier study led to rapid expectation adaptation and thereby concealed a possible facilitative effect of OT1H. Specifically, if readers came to expect OTOH to occur throughout the study, any possible surprisal upon

encountering an unmarked OTOH might disappear, because the marker was already primed to a sufficiently great extent (i.e., there would be no more room for a facilitative effect of OT1H, because OTOH would already generally be expected).

A second hypothesis is that the intervening material included in Scholman (2019) affected the strength of the pre-activation of OTOH. Specifically, in their study, contrast1 consisted of one additional clause and was followed by an additional non-contrastive sentence. Given that comprehenders were able to maintain predictions of the contrastive dependency in Scholman et al. (2017), Scholman (2019) hypothesised that this additional non-contrastive material should not diminish the pre-activation of the marker OTOH. The question of how long comprehenders are able to maintain discourse expectations remains an interesting one to explore, which we will address in future work.

Third, we need to consider a possible cross-linguistic effect. Indeed, Schwab and Liu (2020) found a stronger facilitative effect of discourse markers in German than in English, which could be attributed to differences between the languages in terms of the co-occurrences of the discourse markers. If OT1H occurs more often without OTOH in English than in Dutch, we would expect a weaker pre-activation of OTOH based on OT1H in English, which could explain why no effect was found in the English study, whereas we did find an effect in the current Dutch study. However, corpus data from Scholman et al. (2017) and Section 2 of the current study indicate that the opposite is true: the markers co-occur more frequently in English than in Dutch. In other words, we should find that readers create stronger expectations of an upcoming OTOH after encountering OT1H in English than in Dutch, and, hence, we should see a stronger facilitative effect of OT1H in English. Therefore, we do not consider this a viable explanation for the difference in effects between the studies.

Nevertheless, cross-linguistic differences do present an interesting avenue for future research. In Chinese, different pairs of markers, such as ‘because’ / ‘then’ and ‘although’ / ‘but’, co-occur frequently (Steele & Specia, 2014; Xue, 2005). It would, therefore, be interesting to explore whether the facilitative effect of the first marker in a pair of discourse markers can be replicated in Chinese, or perhaps might even be stronger. Research into phenomena such as these can provide more insight into the effect of language-specific factors on the processing of discourse structure, as well as a better understanding of discourse processing in general.

5.2 Nature of the effects

Tying these results back to the question of what form predictions can take, the current study does not provide evidence for the all-or-nothing activation of a specific lexical item before it is encountered in the linguistic input. Rather, our results support a graded lexical prediction account, which assumes the formation of partial predictions that do not necessarily include word identity (Luke & Christianson, 2016). It appears comprehenders use contrast1 markers, such as

OT1H and EZ, to create content-based expectations of upcoming content that are not necessarily narrowed down to a specific lexical item. In addition to these content-based expectations, comprehenders appear to make much weaker predictions of specific lexical items, such as OTOH and AZ. The current study adds to the existing body of literature on prediction effects by showing that comprehenders can generate predictions of discourse relations across sentences.

Nieuwland et al. (2018) failed to replicate evidence supporting specific lexical predictions. Their results suggested that comprehenders' predictions are not based solely on individual word forms, but rather on more general patterns and regularities in language. Interestingly, our results suggest that discourse processing goes beyond relying strongly on statistical patterns between word correspondences in natural language, at least when it comes to discourse markers and their corresponding relations. The corpus study reported in Section 2 showed that OT1H is strongly predictive of OTOH. If the facilitative effect of OT1H on the processing of OTOH was solely an effect of combined grammatical use (i.e., based on the general pattern of OT1H being followed by OTOH in natural language), then this facilitative effect should have been larger for OT1H than for EZ. Our results provided only anecdotal evidence for such an effect, however. This indicates that comprehenders are sensitive to meaning at the discourse level when generating expectations of upcoming content – that is, they generate expectations for an upcoming contrast relation irrespective of the prototypical marker that should follow the specific contrast1 marker according to patterns found in natural language.

However, other explanations of the data need to be considered, as well. As mentioned in Section 1, our results cannot distinguish between prediction and integration: we measured reading times of the target marker OTOH, but the clearest demonstration of prediction occurs when a study reveals the activation of a linguistic representation of a word before the comprehender encounters that word (Pickering & Gambi, 2018). Thus, our results are also in line with an integration account that assumes that OTOH is easier to integrate with the context if OT1H was present. Or, to adopt terminology from Ferreira and Chantavarin (2018), OT1H allows readers to establish a mental representation of a discourse containing a contrast, which, in turn, allows readers to prepare for an upcoming contrast relation, and possibly also predict how this relation will be marked. This preparation then makes OTOH easier to integrate with prior material, once the marker is encountered. Even if one would interpret the findings in this article as reflecting integration or content-based predictions, rather than specific lexical predictions, the results still provide an important contribution to the field of discourse: they add to the evidence that comprehenders do generate expectations at the discourse level, contrary to an account in which discourse expectation maintenance could be considered ineffective, due to the extensive flexibility and optionality at play at the discourse level.

Finally, it is possible that the effects found are not necessarily indicative of facilitation by the contrast1 marker, but could instead be considered as inhibitory slow-down effects caused

by OTOH: upon encountering an OTOH that was not pre-signalled, readers might be surprised that a contrast relation occurred in the discourse (rather than a different type of relation, such as a causal relation), and, thus, have to update their mental representation. This framing focuses on what other expectations readers might have had. A story continuation study reported in Scholman (2019, Chapter 11) showed that 40% of the continuations that participants produced in the condition without a contrast1 marker were, in fact, contrastive (marked by other connectives than OTOH). This indicates that the prompts could already prime participants to expect a contrast. Thus, it is not the case that a contrast2 was necessarily unexpected and, therefore, elicited a slow-down in the absence of OT1H. A more probable theory is that OT1H facilitates processing of OTOH, such that the OT1H/EZ marker leads readers to build an expectation for OTOH that causes the reader not to be surprised upon encountering OTOH.

In sum, the results from the current study further elucidate the nature of discourse processing, by showing that comprehenders do create expectations of upcoming connectives at the discourse level. This work thereby extends previous findings that comprehenders generate expectations of discourse content, to show that they can also pre-activate markers of discourse relations. This pre-activation results in an immediate processing advantage, as evidenced by an effect of OT1H in the first pass duration in the eye-tracking experiment, and by effects on the marker OTOH, but not on the words following OTOH, in both the SPR and eye-tracking experiments.

Data availability

The items and data are made available online: https://osf.io/y5afc.

Ethics and consent

All human subjects experimentation reported here was deemed exempt by the ethics board of the Deutsche Gesellschaft für Sprachwissenschaft Ethics Committee. All participants provided informed consent. The identity of research subjects was anonymised.

Funding information

This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Grant 232722074 – SFB 1102.

Author contributions

MS, HR, and VD conceptualized the study and decided on the methodology. MS conducted the main investigation, and was responsible for data curation, analysis and visualization. HR and VD supervised throughout. MS wrote the original draft, and collaborated with HR and VD to revise and address reviews.

References

Asr, F. T., & Demberg, V. (2020). Interpretation of discourse connectives is probabilistic: Evidence from the study of *but* and *although*. *Discourse Processes*, *57*(4), 376–399. DOI: https://doi.org/10.1080/0163853X.2019.1700760

Barthel, M., Tomasello, R., & Liu, M. (2022). Online comprehension of conditionals in context: A self-paced reading study on *wenn* ('if') versus *nur wenn* ('only if') in German. *Linguistics Vanguard*, *8*(4). DOI: https://doi.org/10.1515/lingvan-2021-0083

Bürkner, P.-C. (2017). Brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*, 1–28. DOI: https://doi.org/10.18637/jss.v080.i01

Crible, L. (2021). Negation cancels discourse-level processing differences: Evidence from reading times in concession and result relations. *Journal of Psycholinguistic Research*, *50*(6), 1–26. DOI: https://doi.org/10.1007/s10936-021-09802-2

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*(8), 1117–1121. DOI: https://doi.org/10.1038/nn1504

Dery, J. E., & Koenig, J.-P. (2015). A narrative-expectation-based approach to temporal update in discourse comprehension. *Discourse Processes*, *52*(7), 559–584. DOI: https://doi.org/10.1080/0163853X.2014.966293

Ferreira, F., & Chantavarin, S. (2018). Integration and prediction in language processing: A synthesis of old and new. *Current Directions in Psychological Science*, *27*(6), 443–448. DOI: https://doi.org/10.1177/0963721418794491

Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLoS One*, *8*(10), 1–18. DOI: https://doi.org/10.1371/journal.pone.0077661

Fleur, D. S., Flecken, M., Rommers, J., & Nieuwland, M. S. (2020). Definitely saw it coming? The dual nature of the pre-nominal prediction effect. *Cognition*, *204*, 104335. DOI: https://doi.org/10.1016/j.cognition.2020.104335

Foucart, A., Martin, C. D., Moreno, E. M., & Costa, A. (2014). Can bilinguals see it coming? Word anticipation in L2 sentence reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(5), 1461. DOI: https://doi.org/10.1037/a0036756

Fruchter, J., Linzen, T., Westerlund, M., & Marantz, A. (2015). Lexical preactivation in basic linguistic phrases. *Journal of Cognitive Neuroscience*, *27*(10), 1912–1935. DOI: https://doi.org/10.1162/jocn_a_00822

Hinnell, J. (2019). The verbal-kinesic enactment of contrast in North American English. *The American Journal of Semiotics*, *35*(1–2), 55–92. DOI: https://doi.org/10.5840/ajs20198754

Hoek, J., Rohde, H., Evers-Vermeul, J., & Sanders, T. J. (2020). Scolding the child who threw the scissors: Shaping discourse expectations by restricting referents. *Language, Cognition and Neuroscience*, 1–18. DOI: https://doi.org/10.1080/23273798.2020.1852292

Hoek, J., Rohde, H., Evers-Vermeul, J., & Sanders, T. J. (2021). Expectations from relative clauses: Real-time coherence updates in discourse processing. *Cognition*, *210*, 104581. DOI: https://doi.org/10.1016/j.cognition.2020.104581

Husband, E. M. (2022). Prediction in the maze: Evidence for probabilistic pre-activation from the English *a/an* contrast. *Glossa Psycholinguistics*, *1*(1). DOI: https://doi.org/10.5070/G601153

Ito, A., Corley, M., Pickering, M. J., Martin, A. E., & Nieuwland, M. S. (2016). Predicting form and meaning: Evidence from brain potentials. *Journal of Memory and Language*, *86*, 157–171. DOI: https://doi.org/10.1016/j.jml.2015.10.007

Ito, A., Gambi, C., Pickering, M. J., Fuellenbach, K., & Husband, E. M. (2020). Prediction of phonological and gender information: An event-related potential study in Italian. *Neuropsychologia*, *136*, 107291. DOI: https://doi.org/10.1016/j.neuropsychologia.2019.107291

Ito, A., Martin, A. E., & Nieuwland, M. S. (2017). How robust are prediction effects in language comprehension? Failure to replicate article-elicited N400 effects. *Language, Cognition and Neuroscience*, *32*(8), 954–965. DOI: https://doi.org/10.1080/23273798.2016.1242761

Kim, A., & Lai, V. (2012). Rapid interactions between lexical semantic and word form analysis during word recognition in context: Evidence from ERPs. *Journal of Cognitive Neuroscience*, *24*(5), 1104–1112. DOI: https://doi.org/10.1162/jocn_a_00148

Kochari, A. R., & Flecken, M. (2019). Lexical prediction in language comprehension: A replication study of grammatical gender effects in Dutch. *Language, Cognition and Neuroscience*, *34*(2), 239–253. DOI: https://doi.org/10.1080/23273798.2018.1524500

Köhne-Fuetterer, J., Drenhaus, H., Delogu, F., & Demberg, V. (2021). The online processing of causal and concessive discourse connectives. *Linguistics*, *59*(2), 417–448. DOI: https://doi.org/10.1515/ling-2021-0011

Koornneef, A. (2021). The processing signature of anticipatory reading: An eye-tracking study on lexical predictions. *Linguistics*, *59*(2), 449–479. DOI: https://doi.org/10.1515/ling-2021-0014

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59. DOI: https://doi.org/10.1080/23273798.2015.1102299

Laszlo, S., & Federmeier, K. D. (2009). A beautiful day in the neighborhood: An event-related potential study of lexical relationships and prediction in context. *Journal of Memory and Language*, *61*(3), 326–338. DOI: https://doi.org/10.1016/j.jml.2009.06.004

Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, *100*(9), 1989–2001. DOI: https://doi.org/10.1016/j.jmva.2009.04.008

Luke, S. G., & Christianson, K. (2016). Limits on lexical prediction during reading. *Cognitive Psychology*, *88*, 22–60. DOI: https://doi.org/10.1016/j.cogpsych.2016.06.002

Martin, C. D., Branzi, F. M., & Bar, M. (2018). Prediction is production: The missing link between language production and comprehension. *Scientific Reports*, *8*(1), 1079. DOI: https://doi.org/10.1038/s41598-018-19499-4

Molnár, V. (2002). Contrast – from a contrastive perspective. In *Information structure in a cross-linguistic perspective* (pp. 147–161). Brill. DOI: https://doi.org/10.1163/9789004334250_010

Nicenboim, B., Vasishth, S., & Rösler, F. (2020). Are words pre-activated probabilistically during sentence comprehension? Evidence from new data and a Bayesian random-effects meta-analysis using publicly available data. *Neuropsychologia*, *142*, 107427. DOI: https://doi.org/10.1016/j.neuropsychologia.2020.107427

Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., Von Grebmer Zu Wolfsthurn, S., Bartolozzi, F., Kogan, V., Ito, A., et al. (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *ELife*, *7*, e33468. DOI: https://doi.org/10.7554/eLife.33468

Noordman, L. G. M., Vonk, W., & Kempff, H. J. (1992). Causal inferences during the reading of expository texts. *Journal of Memory and Language*, *31*(5), 573–590. DOI: https://doi.org/10.1016/0749-596X(92)90029-W

Oostdijk, N., Reynaert, M., Hoste, V., & Schuurman, I. (2013). The construction of a 500-million-word reference corpus of contemporary written Dutch. In *Essential speech and language technology for Dutch: Results by the STEVIN-programme*. Springer Verlag. DOI: https://doi.org/10.1007/978-3-642-30910-6_13

Otten, M., Nieuwland, M. S., & Van Berkum, J. J. (2007). Great expectations: Specific lexical anticipation influences the processing of spoken language. *BMC Neuroscience*, *8*(1), 1–9. DOI: https://doi.org/10.1186/1471-2202-8-89

Otten, M., & Van Berkum, J. J. (2009). Does working memory capacity affect the ability to predict upcoming words in discourse? *Brain Research, 1291*, 92–101. DOI: https://doi.org/10.1016/j.brainres.2009.07.042

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin, 144*(10), 1002. DOI: https://doi.org/10.1037/bul0000158

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372–422. DOI: https://doi.org/10.1037/0033-2909.124.3.372

Rohde, H., & Horton, W. S. (2014). Anticipatory looks reveal expectations about discourse relations. *Cognition, 133*(3), 667–691. DOI: https://doi.org/10.1016/j.cognition.2014.08.012

Schad, D. J., Nicenboim, B., Bürkner, P.-C., Betancourt, M., & Vasishth, S. (2022). Workflow techniques for the robust use of Bayes factors. *Psychological Methods*. DOI: https://doi.org/10.1037/met0000472

Scholman, M. C. J. (2019). *Coherence relations in discourse and cognition: Comparing approaches, annotations and interpretations* [Doctoral dissertation, Saarland University].

Scholman, M. C. J., Demberg, V., & Sanders, T. J. (2020). Individual differences in expecting coherence relations: Exploring the variability in sensitivity to contextual signals in discourse. *Discourse Processes, 57*(10), 844–861. DOI: https://doi.org/10.1080/0163853X.2020.1813492

Scholman, M. C. J., Rohde, H., & Demberg, V. (2017). "On the one hand" as a cue to anticipate upcoming discourse structure. *Journal of Memory and Language, 97*, 47–60. DOI: https://doi.org/10.1016/j.jml.2017.07.010

Schwab, J., & Liu, M. (2020). Lexical and contextual cue effects in discourse expectations: Experimenting with German 'zwar… aber' and English 'true/sure… but'. *Dialogue & Discourse, 11*(2), 74–109. DOI: https://doi.org/10.5087/dad.2020.203

Schwarz, F., & Zehr, J. (2021). Tutorial: Introduction to PCIbex – An Open-Science Platform for Online Experiments: Design, Data-Collection and Code-Sharing. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 15–16.

Staub, A., & Clifton, C. (2006). Syntactic prediction in language comprehension: Evidence from either…or. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(2), 425–436. DOI: https://doi.org/10.1037/0278-7393.32.2.425

Steele, D., & Specia, L. (2014). Divergences in the usage of discourse markers in English and Mandarin Chinese. *International Conference on Text, Speech, and Dialogue*, 189–200. DOI: https://doi.org/10.1007/978-3-319-10816-2_24

Thornhill, D. E., & Van Petten, C. (2012). Lexical versus conceptual anticipation during sentence processing: Frontal positivity and N400 ERP components. *International Journal of Psychophysiology, 83*(3), 382–392. DOI: https://doi.org/10.1016/j.ijpsycho.2011.12.007

Umbach, C. (2004). On the notion of contrast in information structure and discourse structure. *Journal of Semantics, 21*(2), 155–175. DOI: https://doi.org/10.1093/jos/21.2.155

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*(3), 443. DOI: https://doi.org/10.1037/0278-7393.31.3.443

Wicha, N. Y., Bates, E. A., Moreno, E. M., & Kutas, M. (2003). Potato not Pope: human brain potentials to gender expectation and agreement in Spanish spoken sentences. *Neuroscience Letters*, *346*(3), 165–168. DOI: https://doi.org/10.1016/S0304-3940(03)00599-8

Wicha, N. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience*, *16*(7), 1272–1288. DOI: https://doi.org/10.1162/0898929041920487

Xiang, M., & Kuperberg, G. (2015). Reversing expectations during discourse comprehension. *Language, Cognition and Neuroscience*, *30*(6), 648–672. DOI: https://doi.org/10.1080/23273798.2014.995679

Xue, N. (2005). Annotating discourse connectives in the Chinese Treebank. *Proceedings of the Workshop on Frontiers in Corpus Annotations II: Pie in the Sky*, 84–91. DOI: https://doi.org/10.3115/1608829.1608841

Yi, E., & Koenig, J.-P. (2021). Grammar modulates discourse expectations: Evidence from causal relations in English and Korean. *Language and Cognition*, *13*(1), 99–127. DOI: https://doi.org/10.1017/langcog.2020.29