Lexical and contextual cue effects in discourse expectations: Experimenting with German 'zwar...aber' and English 'true/sure...but'

Juliane Schwab   JSCHWAB@UNI-OSNABRUECK.DE
Institute of Cognitive Science
Osnabrück University

Mingya Liu   MINGYA.LIU@HU-BERLIN.DE
Department of English and American Studies
Humboldt University of Berlin

Abstract

Existing literature shows that readers and listeners rapidly adjust their expectations about likely discourse continuations through discourse markers, as well as through other narrow linguistic and discourse contextual cues. However, it is unclear whether (i) the facilitative effects of various linguistic cues differ in quality and (ii) whether the effects interact with one another in any principled manner. We conducted two self-paced reading experiments on concessive constructions in German and English wherein optional lexical and/or contextual cues appeared ahead of a concessive discourse connective. The results demonstrate that readers can use both types of cues to anticipate the upcoming connective. Thus, our study provides novel evidence for expectation-driven accounts of discourse processing and elucidates the functions of discourse signals. Furthermore, the results also show that the role a type of cue plays may be subject to cross-linguistic variation.

Keywords: Discourse processing, predictions, concessives, discourse signals, self-paced reading

1. Introduction

Although much of psycho- and neurolinguistic research is concerned with the processing of individual sentences, language processing in a natural setting involves more than the generation of meaning in isolated sentences. Instead, comprehenders swiftly relate propositions within a text or conversation to one another and the broader (extra-)linguistic context to build a coherent discourse representation.

A range of psycho- and neurolinguistic studies indicate that processing at the discourse level is aided by predictions for the subsequent discourse content and structure (Drenhaus et al., 2014; Köhne and Demberg, 2013; Rohde et al., 2011; Rohde and Horton, 2014; Scholman et al., 2017; van Bergen and Bosker, 2018; Xiang and Kuperberg, 2015). The current literature indicates that discourse expectations can be formed from a range of cue words or phrases, such as discourse markers and discourse connectives. Xiang and Kuperberg (2015), for instance, demonstrated that the discourse connective *even so* reverses readers' expectation for the sentence continuation, such that unlikely events, like partying after having failed an exam in (1), become expected. Similar effects

have also been found for the concessive connectives *however* and German *dennoch* 'nevertheless' (Drenhaus et al., 2014; Köhne and Demberg, 2013), as well as the causal connectives *therefore* and its German counterpart *daher* (*ibid.*).

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

Further, Scholman et al. (2017) have studied the processing of the two-part construction *'on the one hand...on the other hand'*. They found that the discourse marker *'on the one hand'* affected how readers integrated upcoming discourse relations. Among other things, the authors compared the reading times at *on the other hand* in conditions where the critical region was preceded by *on the one hand* to a control condition where it was not. In the initial study, they did not find a facilitating effect of the first connective on processing the second one. A follow-up eye-tracking study (Scholman et al., 2018) with an increased filler-to-item ratio and twice the amount of subjects, however, found a first-pass reading time effect pointing toward prediction, that is, *on the other hand* was read faster if *on the one hand* had appeared in a previous sentence. The volatility of the effect suggests that readers' adaptation to the structures they are exposed to during an experiment is a real concern when studying discourse expectations.

If, as shown by the two studies discussed above, comprehenders generate discourse expectations, an obvious question is whether the cues relevant for the formation of expectations extend beyond discourse connectives. An increasing amount of research focuses on the questions how often and why discourse relations are signaled by means other than discourse markers, such as lexical, syntactic, or graphical cues (such as enumerated lists in written text) (Das and Taboada, 2018; Taboada and Das, 2013). Critically, these discourse signals may appear with or without a discourse marker. In fact, Das and Taboada (2018) have shown that, at least within the RST (Rhetorical Structure Theory) Discourse Treebank (Carlson et al., 2002), 75% of discourse relations were marked exclusively with signals other than discourse markers, and 8% were marked with more than one signal. In light of these findings, it is clear that these discourse signals must serve an important function for establishing coherence relations. However, the precise mechanism via which signals come to indicate a specific discourse relation, as well as their interaction with discourse markers is so far not well understood.

Finally, a number of neurolinguistic studies have shown that the discourse context has an early impact on language comprehension. For instance, Nieuwland and Van Berkum (2006) showed that supportive discourse contexts can quench processing difficulties typically observed for local semantic violations. In addition, Hald et al. (2007) demonstrated that the discourse context has an immediate effect on sentence interpretation. In a visual world eye-tracking study, Kim et al. (2015) showed that the discourse context acts as a constraint on the alternatives computed on-line during sentences containing focus particles. Given this evidence for on-line effects of the discourse context during sentence processing, expectations for the discourse too may be affected by contextual signals, for instance through pragmatic inferences. One example for the immediate effect of pragmatic influences on discourse expectations comes from experimental work on lexically triggered pragmatic inferences. Implicit causality verbs, such as *scolded* in (2) were found to generate an expectation for an upcoming causal discourse relation, that is, a reason for the scolding, for instance. The generated expectation was evidenced by anticipatory eye movements (Rohde and Horton, 2014) and participants' on-line parsing decisions (Rohde et al., 2011).

(2)   Arthur scolded Patricia in the hallway.
        Cause continuation: She had put thumbtacks on the teacher's chair.

To summarize, past work has shown that discourse processing is expectation-based, such that both discourse connectives and pragmatic inferences can generate discourse expectations. Furthermore, these findings hint that additional discourse signals may have a similar effect on the on-line processing of discourse relations. So far, however, no study to our knowledge has either (i) directly contrasted the facilitative effects of different cues, or (ii) studied how comprehenders process sentences containing multiple cues.

In the current study, we investigate the role of two types of discourse signals that appear *ahead* of a discourse connective and their interaction in discourse relation processing. We address the role of what we call 'lexical' cues, such as *true* in (3a), and 'contextual' cues, such as the incompatibility between liking to run *outdoors* and owning a treadmill in (3b), for the generation of discourse expectations.

(3)   a.   Paul likes to run. (True,) he has a treadmill in the living room, but he often jogs in parks.
        b.   Paul likes to run (outdoors). He has a treadmill in the living room, but he often jogs in parks.

In particular, we used a self-paced reading task with subsequent naturalness ratings to answer two questions: First, do discourse cues generate an expectation that facilitates processing the discourse connective (e.g. *but*), as indexed by reduced reading times? Second, do the effects of multiple cues interact with one another in any principled manner? Does, for example, the occurrence of multiple cues result in additive facilitative effects? We take a cross-linguistic perspective (targeting German and English), as theories on the roots of expectation effects in language processing indicate that expectation effects arise from general cognitive processing principles (Levy, 2008; Lewis and Bastiaansen, 2015). Thus, cues that are similarly good indicators for the upcoming discourse relation in both languages under observation should also lead to a similar processing effect. In addition, the comparison between German and English builds upon existing work on the processing of concessive discourse relations in these two languages (Drenhaus et al., 2014) that indicates similarities with regard to the processing effects of concessive connectives. Based on the results of our study, we will argue that additional cues preceding the discourse connective may benefit the communication between speaker and listener by generating an anticipatory signal that can facilitate processing the upcoming discourse relation.

The article is organized as follows: in section 2, we introduce the discourse cues under investigation in both languages. In section 3, we present production and corpus-linguistic studies on the German and English lexical cues that inform our hypotheses for the experiments. In section 4, we present the details of the German and English experiments. As outlook for the results: in German, we found evidence for expectations generated from lexical and contextual cues. In English, however, we only found evidence for lexical cue-based expectations. In section 5, we discuss these findings, explanations for the divergence between German and English, and implications for future research on discourse phenomena.

2. The tested expressions

Our study concerns the processing of concessive discourse relations. Following König's (2006) definition, concessive clauses, for example in the form of 'p, but q' involve the assertion of two propositions p and q, despite a presupposed incompatibility of q with p. Concessives have also been described as *anti-causal* (Konig and Siemund, 2000), as p is said to give rise to a causal inference that is subsequently denied by q. For instance, in (3), p (*He has a treadmill...*) may give rise to the expectation that he extensively uses this treadmill to exercise. Q (*he often jogs in parks*) subsequently denies this expectation. We decided to target concessive discourse relations in order to follow up and extend upon the existing research summarized above on the role of connectives in the processing of relations of opposition such as concession and contrast (Asr and Demberg, 2020; Drenhaus et al., 2014; Köhne and Demberg, 2013; Scholman et al., 2017; Xiang and Kuperberg, 2015).

In this section, we will introduce the cues we refer to as lexical and contextual cues for concessive discourse relations in German and English. We start with the lexical cues[1]: the German discourse marker *zwar*[2] ('truly') and the English particles *true* and *sure*. The latter particles are not typically classified as discourse markers (they are listed in neither the Penn Discourse Treebank (Prasad et al., 2008) nor the English lexicon of discourse markers DiMLex-Eng (Das et al., 2018)). However, they have been described as markers of concessivity in the past (Antaki and Wetherell, 1999; König, 2006). We argue that *true/sure* are ambiguous at the turn-initial position, where besides marking a conceded argument, they can also express the speaker's sincere agreement with a previous speaker's assertion. However, in turn-medial positions and outside of dialogues, the latter reading is infelicitous, rendering *true/sure* a clear signal for a concessive relation. We lay out the details of our argument below.

Our discussion of the English lexical cues focuses on clause-modifying *true/sure*[3], where the particles simply assert that the speaker believes the clause the particles are modifying to be true. Signaling a concessive relation is not part of their core meaning. This becomes apparent when they appear at the turn-initial position: *true/sure* here are used to express agreement with the last speaker's assertion (4a). A concessive relation is possible, but optional (4b).

(4)   A: Mary has done a lot to improve her health recently.

   a.   B: True/Sure(, she even quit smoking).
   b.   B: True/Sure(, but she still smokes like a chimney).

We argue that *true/sure* in turn-medial positions and outside of dialogues become a clear indicator for a concessive relation: re-asserting that one believes in the truth of one's own utterance is infelicitous, as can be seen from the minimally modified example without speaker change (5a).

---

1. We use the term *lexical cue* as umbrella term, as our English and German lexical cues differ with regard to whether they are typically classified as discourse markers

2. The Etymological Dictionary of German (Pfeifer et al., 1993) indicates that *zwar* is a contraction of the proposition *zu* ('to') and the adjective *wahr* ('true'), from Old High German *zi wāre* 'in truth, indeed' (8th century), Middle High German *ze wāre/zewāre*, originally used to express agreement to an assertion. Its use as concessive marker seems to have emerged in Middle Low German

3. Obviously, *true/sure* have a discourse-unrelated function as adjectives. This meaning, however, is easily distinguished from the clause-modifying one and is not available in the structural configurations we are investigating. Therefore, we have nothing more to say on this matter.

Instead, *true/sure* can only function as markers of an argument that is conceded to the audience. The particles can appear both before the conceded argument (5b) or after it (5c).

(5)   Mary has done a lot to improve her health recently.

    a. #True/Sure(, she even quit smoking).
    b. True/Sure, she still smokes, but she started exercising regularly.
    c. She still smokes, true/sure, but she started exercising regularly.

Given these observations, we chose material wherein there is no ambiguity about the function of *true/sure* (see (3a)) for our study, that is, items where there is no apparent speaker (and thus no speaker change), and where *true/sure* only appears at the second sentence of our items, thus preventing participants from interpreting it as response to an omitted previous utterance.

The German lexical cue that we investigate in analogy to *true/sure* is the adverb *zwar* ('truly'). The German lexicon of discourse markers, DiMLex (Stede and Umbach, 1998), lists *zwar* as a discourse marker that appears at the first argument of a concessive relation[4]. Contrary to English *true/sure*, which appears before or after the proposition it modifies (5), German *zwar* usually appears immediately after the finite verb, that is, inside the proposition (6a). While the flexibility in the German word order allows fronting the adverbial phrase such that *zwar* appears in a sentence-initial position (6b), this is less frequent than sentence-medial *zwar*. Similar to the two-part marker *'on the one hand...on the other hand'*, marking the conceded argument with *zwar* typically requires the second argument to be marked with a concessive connective (e.g. *aber* ('but') in (6)). Just as for English *true/sure*, *zwar* is generally optional in establishing a concessive discourse relation. Nevertheless, the two-part *'zwar...aber'* construction is highly frequent both in formal and informal registers. To our intuition, the *'true/sure...but'* construction, on the other hand, is less frequent. Data from a parallel corpus that will be reported in section 3 confirms this intuition, but also demonstrates that *'true/sure...but'* is indeed analogous in meaning to German *'zwar...aber'*.

(6)   *Jens joggt gerne (draußen).*
      Jens jogs gladly (outdoors).
      'Jens likes to run (outdoors).

    a. *Er hat (zwar) ein Laufband im   Wohnzimmer, aber er joggt häufig im   Park.*
        He has (true) a   treadmill in-the living-room   but he jogs often in-the park
    b. *Zwar hat er ein Laufband im   Wohnzimmer, aber er joggt häufig im   Park.*
        True has he a   treadmill in-the living-room   but he jogs often in-the park
        'True, he has a treadmill in the living room, but he often goes running in the park.'

As what we call contextual cues, we employ adverbials added to the first sentence of our mini discourses (see (3b) and (6)). We refer to this as *contextual* cue as the adverbials themselves do not convey information regarding the discourse relation. Instead, the adverbial generates an incoherence between the first and the second proposition of our items that should give rise to an expectation for a concessive discourse relation.

Let us spell out the workings of this manipulation in detail: (7a) literally asserts that John likes running outdoors. At the same time, as the manner of running is explicitly stated through inclusion of the optional adverbial *outdoors*, the proposition evokes alternatives to *outdoors* (e.g.

---

4. German also has a related expression, *und zwar* (lit. 'and true', roughly 'namely'). This expression, however, is used to mark an expansion, rather than a concession (DiMLex, Stede and Umbach (1998)).

on a treadmill, at the gym) and implies that, circumstances permitting, John indeed prefers running outdoors over its alternatives. By contrast, the sentence without adverbial, (7b), is underspecified with regard to John's running exercise.

(7)   a.   John likes to run outdoors.
          b.   John likes to run.

Following (7), the second sentence, (8), asserts that John has a treadmill. Out of the blue, this sentence would be rather uninformative with regard to likely discourse continuations (one could for instance continue the discourse with: 'It was a gift from his sister, who didn't need it anymore'). However, given (7), (8) expands the comment on John's exercise regimen and implies that he uses his treadmill to run. On the one hand, as continuation of (7b), this is coherent with the immediately preceding sentence and thus does not generate an expectation for a concessive discourse relation (a continuation such as *'and he uses it every day'* would be perfectly coherent). As continuation of (7a), on the other hand, this generates an incoherence: (7a) implied that John typically runs outdoors, while (8) implies that he uses his treadmill.

(8)   He has a treadmill in the living room.

We argue that this discourse-based incoherence could generate an expectation for a concessive discourse continuation. This is because a concessive relation between (8) and a subsequent second argument reinstates the coherence of the discourse: in (9), the former becomes the conceded argument, for which, while true, the inference that John likes to use his treadmill to run, is denied.

(9)   He has a treadmill in the living room, but he often jogs in parks.

In our stimulus material, the adverbial in (7a) is instrumental to achieve an opposition between (7a) and (the inference of) (8), which in turn should generate an expectation for a concessive discourse relation. Note, however, that the same effect can be achieved with any set of two propositions in which the second one gives rise to an inference that is contradictory to what was asserted with the first one. An example can be seen in (10): having reversed the valence of the first sentence from *liking* to *hating to run*, the second proposition together with its inference that John uses the treadmill to run is suddenly rendered contradictory to the first sentence. As argued above, the saving grace to reinstate the coherence of the discourse is a concessive discourse relation, wherein the inference of the second proposition is denied.

(10)   John hates running. He has a treadmill in the living room ?(, but only his wife ever uses it.)

Given the existing evidence for comprehenders' ability to generate discourse expectations from pragmatic cues outlined in the introduction (Kim et al., 2015; Rohde et al., 2011; Rohde and Horton, 2014), we expect that comprehenders should be able to recognize the incoherence in the discourse and rapidly adjust their expectations about the likely discourse continuation. In our materials, this should surface as a reduced reading time at the connective *but*. It is to note that with the incoherence generated through the addition of the contextual cue, there is a specific expectation about the inference that will be denied, namely that John runs *indoors* on his treadmill. By contrast, one could argue that in contexts containing only a lexical cue (3a), it is unclear what the default inference from John owning a treadmill should be. Nevertheless, we take the position that, even then, the most obvious invited inference given the first sentence is that John likes to use his treadmill to run.

Consequently, the sentence becomes somewhat odd if the concession denies a different inference (11). Finally, while we agree that the contextual cue may set up the reader with a slightly more specific expectation regarding the *content* of the upcoming discourse, in both contextual and lexical cue conditions an expectation for the *structure* of the discourse, as measured at the connective *but*, should hold.

(11) a. True, John has a treadmill in the living room, but he's not a particularly fast runner.
      b. John likes to run. ?True, he has a treadmill in the living room, but he's not a particularly fast runner.

To summarize, in the present study, we target both lexical *and* contextual cues as the two indicate concessivity through crucially distinct mechanisms: German *zwar* and English *true/sure* mark a conceded argument. It is part of their lexical meaning to indicate concessivity. As such, a comprehender might automatically generate an expectation for a concessive discourse relation upon reading these lexical elements. By contrast, the contextual cue indicates concessivity through an incompatibility between two propositions and the inferences they give rise to. Comprehenders must identify this incompatibility on-the-fly and use it to re-evaluate the discourse structure if they are to anticipate the upcoming concessive relation. Our study contributes to the literature by investigating both types of cues within the same stimulus material and across two languages, thus providing insight into the effects of various cues in discourse relation processing.

In the next section, we focus on the lexical cues in German and English. Using corpus data and a production study, we first demonstrate that the lexical cues are indeed predictive of an upcoming concessive connective. The predictive power attested through this data forms the basis of our hypothesis that comprehenders should be able to generate discourse expectations on-line on the basis of these cues. Second, in a parallel corpus study on constructions using the German or English lexical cues, we show that *zwar* is much more frequent than *true/sure* and often does not receive an analogous lexical marker in English translations. However, *true/sure* represent a low-frequency, but reliable, option to instantiate a concessive discourse relation in English that is commonly translated to *'zwar...aber'*. The data reported in section 3 informs our hypotheses for the self-paced reading experiments reported in section 4.

3. Frequency and reliability of *zwar*, *true*, and *sure* as lexical cues

As our study concerns the question whether the lexical cues *zwar*, *true*, and *sure* have a facilitative effect on processing the following connectives *aber* and *but*[5] respectively, we report natural language data from German and English corpora that attest to the predictive strength of these expressions in sections 3.1 and 3.2. For the German expression *zwar*, we also include a production study testing the extent to which it triggers the use of *aber* or functionally related connectives like *doch/allerdings* ('however'). In section 3.3, we then estimate the frequency of *'zwar...aber'* and *'true/sure...but'* constructions using a parallel corpus.

---

5. We chose *aber* and *but* as connectives as they are the most frequent connectives following *zwar* and *true/sure*, see section 3.3. Of course, the concessive discourse relation can also be realized through other connectives, e.g. German *jedoch* / English *however*.

3.1 *zwar...aber* in German

The predictive strength of *zwar* as a lexical cue is attested in data collected and analyzed by Günthner (2016): in a data set of 91 informal conversations, Günthner (2016) found that *zwar* was part of a concessive relation in the vast majority of cases (96% of all extracted instances of *zwar*). Only 1.3% of her data were stand-alone instances of *zwar*, see an example in (12) (Günthner (2016), p.156; transcription simplified).

(12)   *H. recounts being asked to go out to a club*
        H: *Un ich so; ich hab* zwar *keine Zeit, ich war auch in Arbeitsklamotten,*
        And I was like; *true* I have no time, I was wearing my work clothes
        (Anna, Eri and Mari laugh)
        H: *dann äh nach Hause, schnell geduscht*
        then (uh) straight home, quickly showered
        E: *kann gar nich sein*
        can't be
        H: *sind in die Disse und da war das das*
        went to this club and that was it that

In (12), the expected connective *aber* is substituted with a recounting of events that instead illustrate a concession via H's subsequent actions. Günthner also cites an earlier study by Primatarova-Miltscheva (1986) that confirms this distribution of *zwar...aber*, although Primatarova-Miltscheva finds a slightly higher rate of stand-alone instances of *zwar*, with 18% of the analyzed spoken language instances of *zwar* not being part of a concessive relation.

To verify the predictability of *aber* by *zwar*, we also conducted a sentence continuation study. 26 students of Osnabrück University (16 female, mean age = 20.38, sd = 1.6) participated in the study. All participants were native speakers of German and had normal or corrected-to-normal vision. Participants gave written informed consent prior to participating and were compensated with partial course credit. Participants were shown sentence fragments such as (13) on a computer screen and used the keyboard to enter a grammatical continuation of the sentence into a text box below the fragment. There was no time or word limit for their responses. Each participant saw 4 items with *zwar* as in (13), pseudo-randomly interspersed with 28 fillers from another study. The results of the experiment are summarized in Table 1. We list both the connective used and its position, as German allows for the placement of a connective either before the subject position (Vor-Vorfeld, 'pre-prefield') or between the finite and nonfinite verb positions (Mittelfeld, 'middle field'). See examples for each in (14).

(13)   *Louisa hat zwar eine Plattensammlung...*
        Louisa has true   a   vinyl-collection...
        'True, Louisa has a vinyl collection...'

Three aspects of the results are notable: for one, participants *always* produced an adversative or concessive discourse marker. Second, *aber* is the most frequently used marker. In the cases where participants used other markers, these can be substituted by *aber* without change in meaning (e.g. (14)). As noted by Scheffler and Stede (2016) and Stede et al. (2019), the discourse connectives in Table 1 overlap in their discourse functions, such that all of them can designate a concessive relation. It is thus in line with these authors' work that we observed participants using a variety of

connectives. Third, the Vor-Vorfeld position, which is where the discourse marker appears in the material for our study, is the preferred position to mark the concessive relation.

(14)   a.   *Louisa hat zwar eine Plattensammlung, doch / aber sie hört   nicht gerne Musik.*
            Louisa has true a   vinyl-collection, yet / but she listens not gladly music
            'True, Louisa has a vinyl collection, but she does not like listening to music.' (Vor-Vorfeld connective)

    b.   *Louisa hat zwar eine Plattensammlung, ihr Plattenspieler funktioniert allerdings /*
            Louisa has true a   vinyl-collection, her record-player works   however /
            aber nicht mehr richtig.
            but not more correctly
            'True, Louisa has a vinyl collection. But her record player does not really work anymore.' (Mittelfeld connective)

Discourse marker	Frequency in Vor-Vorfeld position	Frequency in Mittelfeld position
aber ('but')	50 (48.08%)	24 (23.08%)
jedoch ('though')	10 (9.62%)	7 (6.73%)
allerdings ('however')	3 (2.88%)	4 (3.85%)
doch ('yet')	4 (3.85%)	0
dennoch ('still')	0	2 (1.92%)
Total	67 (64.42%)	37 (35.58%)

Table 1: Sentence-continuation task: Frequency of adversative/concessive discourse markers following a sentence fragment containing *zwar* with percentage of the total number of responses in parentheses.

Based on the literature and our production study, we assume that *zwar* functions as a good predictor for a concessive construction. Furthermore, as *aber* was the most frequent concessive marker in the production data, participants may come to expect the concessive relation to be lexically realized by *aber*. We thus chose to use *aber* in the material of our German experiment due to its high frequency both in general and in relation to *zwar*.

3.2 *true/sure...but* in English

The English particles *true* and *sure* signal an upcoming concession in a similar way. We searched in the 100,000,000-word The British National Corpus (2007), distributed by the University of Oxford on behalf of the BNC Consortium. The search was conducted via the Treebank.info parser (Uhrig and Proisl, 2011). First, we retrieved all instances of the particles *true* (398 in total) and *sure* (118 total). Then, we manually checked for concessive discourse moves both within the sentence and in the 5 sentences following the sentence marked by *true* or *sure*.

The results show that *true/sure* was commonly part of a concessive move, with *true* occurring as part of a concessive discourse segment in 86% (344 instances) of the data, and *sure* in 83% (98 instances) of the data. Of those, for both particles, intrasentential concession was the most common form of discourse continuation (69% of the 98 instances for *sure*, 59% of the 344 instances for *true*). In a smaller subset of the data, the second argument of the concessive move, i.e. the *but-*

clause, appeared in the immediately following sentence (22% of the data for *sure*, 26% for *true*), or other sentences intervene between the sentence marked with *true/sure* and the second argument (8% of the data for *sure*, 15% for *true*). Examples for both intra- and intersentential concessions are provided in (15)[6]. As the BNC contains some instances of reported speech, the 14% (*true*)/17% (*sure*) of the data that did not contain a concessive discourse relation included some instances of *true/sure* as responses to a previous speaker's assertion, a function also outlined in section 2. In other cases, the function of *true/sure* was not apparent to us.

(15)   a.   *True*, things have gone missing, **but** that does not mean that I am the thief. (BNC-K95-33)

           b.   *True*, she probably still had a long way to go. She was painfully thin and there was an insubstantiality about her. **But** there was no denying that today his wife was better than he had known her for many many months. (BNC-CDE-222-224)

In effect, the corpus data verifies that the particles *true* and *sure* are good indicators for a concessive discourse structure. The reported data from German and English demonstrate that the two English particles are both similar to each other and to the German *zwar* in how often they signal a concessive discourse relation. However, one potentially concerning facet we have thus far not addressed are differences in frequency between German *zwar* and English *true/sure*: while *zwar* is a highly frequent discourse marker, the use of the particles *true* and *sure* appears to be more limited. In the next section, we therefore report on a parallel corpus study that estimates their frequency in comparison to German *zwar* and discusses potential implications for our work.

3.3 Frequency of *zwar*, *true*, and *sure*

Our objectives for a parallel corpus study were to (i) gather an estimate of the frequency of German *zwar* and English *true/sure* from matching text sources, and to (ii) investigate how the German concessive construction using *zwar* is translated to English and vice versa.

We used the German-English archive of the parallel corpus Europarl (version 7) (Koehn, 2005) that contains 1,920,209 aligned sentences from the proceedings of the European Parliament published between 1996 and 2011. First, we extracted all instances of German *zwar* (excluding cases of *'und zwar'*, which, as mentioned in footnote 4 of section 2, marks an expansion rather than a concession), as well as all instances of English *true/sure*. As the corpus does not contain POS-tags or dependency annotations that would facilitate the exclusion of adjectival *true/sure*, we instead restricted our search in the English archive to only include capitalized instances (*'True'*), or instances delimited by commas on either side (*', true,'*). Then, we manually removed any remaining adjectival uses of *true/sure* (as in *'True friendship'*). While, admittedly, this procedure may have caused us to miss some instances of discourse-functional *true/sure*, it ensured a clean data set that reflects the usage of the lexical cue in our experimental material (see (3)).

The search results show that German *zwar* is highly frequent with 11,671 instances in the corpus. As noted in section 2, *zwar* can appear both sentence-initially and sentence-medially. Among the 11,671 hits for *zwar*, 1,601 (13.7%) are sentence-initial (i.e. capitalized). English *true/sure* are quite infrequent with 112 instances of *true* and only 3 of *sure*. Nevertheless, the data again support the discourse cue function of *true/sure* that we have reported above: in 105 of the 115 collective in-

---

6. Examples of usage taken from the BNC were obtained under the terms of the BNC End User Licence. Copyright in the individual texts cited resides with the original IPR holders.

Lexical cue (on conceded argument) in source text **true** (102) /**sure** (3)	Connective (on second argument) in source text					
		but	however	nevertheless/nonetheless	(al)though	even so/yet/still	no connective
Lexical cue in German	Connective on second argument in German						
zwar (31)	*aber* ('but')	17	1	1	–	–	–
	(je)doch ('however')	8	3	–	1	–	–
stimmt ('correct') (21)	*aber* ('but')	12	1	1	–	–	–
	allerdings ('however')	–	1	–	–	–	–
	dennoch ('nevertheless')	–	–	1	–	–	–
	(je)doch ('however')	2	1	–	–	–	1
	no connective	–	1	–	–	–	–
sicher(lich) ('sure(ly)') (9)	*aber* ('but')	5	–	–	–	–	–
	allerdings ('however')	–	1	–	–	–	–
	(je)doch ('however')	1	–	–	1	–	–
	dennoch ('nevertheless')	–	–	–	–	1	–
natürlich ('natural(y)') (8)	*aber* ('but')	3	–	–	–	–	–
	(je)doch ('however')	3	1	1	–	–	–
gewiss ('certain(ly)') (6)	*aber* ('but')	2	–	–	–	–	1
	dennoch ('nevertheless')	–	–	1	–	–	–
	no connective	–	–	–	1	1	–
other lexical cue (26)	*aber* ('but')	11	1	1	1	–	–
	allerdings ('however')	2	–	–	–	–	–
	dennoch ('nevertheless')	–	–	–	–	1	–
	(je)doch ('however')	4	3	–	–	–	1
	no connective	–	1	–	–	–	–
no lexical cue (4)	*aber* ('but')	2	–	–	–	–	–
	(je)doch ('however')	2	–	–	–	–	–

Table 2: Results of a parallel corpus analysis using the German-English aligned version of Europarl. The upper left corner indicates the lexical cue that was present in the original text source (*true/sure*). The two leftmost columns indicate the lexical cue on the conceded argument (with total frequency counts in parentheses) and the connective on the second argument that were identified in the German translation. Entries in the table provide a count of the frequency with which a combination of *true/sure*+connective was translated into a corresponding cue+connective in German. For example, the first entry of the table (17) shows that *'true/sure..., but'* was translated to *'zwar..., aber'* 17 times.

stances of *true/sure*, the cue was part of a concessive relation. In a second analysis step, we aligned the search results with the respective German or English translation. We used the aligned sentences to analyze how German *zwar*-constructions are translated to English and, conversely, how English *true/sure*-constructions are translated to German. Taking the 105 instances of *true/sure* in concessive relations, we manually categorized the aligned sentences by (i) the connective at the second argument in the original texts, (ii) the lexical cue at the conceded argument in the translations, and (iii) the connective at the second argument in the translations. The results are summarized in Table 2. As this procedure was not manageable for all 11,671 instances of *zwar*, we instead selected a random sample of 500 instances of *zwar* (including 87 sentence-initial ones) and followed the same procedure as outlined above. The results are reported in Table 3. For the sake of brevity, we grouped some cue phrases and connectives together (e.g. *nevertheless* and *nonetheless*) and conflated those that appeared in less than 5% of the samples into an *'other'* category. An expanded version of this table featuring the full list of cues can be found in the appendices. For all instances categorized under 'no connective', no connective was apparent in the text.

It is clear that the particles *true/sure* are much less frequent than German *zwar*. For a cross-lingusitic study on these cues, a reader may thus wonder whether the stark difference in the fre-

Lexical cue (on conceded argument) in source text **zwar (500)**		Connective (on second argument) in source text						
		aber	doch	jedoch	dennoch	allerdings	other connective	no connective
Cue/Connective in English	Connective on second argument in English							
while/whilst (81)	*however*	1	–	–	–	–	–	–
	nevertheless	1	1	–	1	–	–	–
	on the other hand	–	1	–	–	–	–	–
	no additional connective	37	16	17	3	1	2	–
(al)though (76)	*however*	–	–	1	–	–	–	–
	nevertheless	3	–	–	–	–	–	–
	on the other hand	–	–	–	–	–	1	–
	no additional connective	38	17	12	–	2	1	1
modal verb (*may*) (28)	*but*	16	4	3	–	–	–	–
	however	1	–	1	–	–	–	–
	on the other hand	–	1	–	–	–	1	–
	no connective	1	–	–	–	–	–	–
other lexical cue (67)	*but*	22	8	5	1	–	–	–
	however	2	1	2	–	1	–	–
	nevertheless/nonetheless	1	–	–	1	–	–	–
	on the other hand	1	–	–	–	–	–	–
	yet/still	1	–	1	–	–	–	–
	no connective	9	7	1	–	–	–	3
no lexical cue (248)	*but*	106	49	44	–	2	3	–
	however	1	2	2	–	1	–	–
	nevertheless	–	–	1	–	–	–	–
	yet	2	3	1	1	–	–	–
	no connective	13	5	2	2	–	1	7

Table 3: Results of a parallel corpus analysis using the German-English aligned version of Europarl. The upper left corner indicates the lexical cue that was present in the original text source (*zwar*). The two leftmost columns indicate the lexical cue on the conceded argument and the connective on the second argument that were identified in the English translation. Entries in the table provide a count of the frequency with which a combination of *zwar*+connective was translated into a corresponding cue+connective in English.

quency of the constructions we are investigating could confound our experimental results. Indeed, past research has often found that low-frequency words and structures can be harder to process (cf. Ferreira et al. (1996); Rayner and Duffy (1986); Schilling et al. (1998) on the effects of word and structural frequency on sentence processing). However, as we are not interested in the processing time at the lexical cue itself or at the regions immediately following the lexical cue, but instead on its ability to generate discourse expectations for an upcoming connective, transient processing difficulties associated with encountering a low-frequency construction should arguably subside before the critical region, the connective *but/aber*. Further, these connectives themselves are the most frequent ones for the instantiation of a concessive relation in both languages. Admittedly, one caveat remains: despite the fact that the corpus data indicates that a concessive connective is highly predictable after *true/sure* and after *zwar*, the low frequency of the former construction may mean that (at least some) English native speakers are not sufficiently familiar with it to rapidly recognize *true/sure* as part of a concessive relation and to generate discourse expectations from it. This could lead to reduced expectation-based effects in the English experiment compared to the German one. In any case, we will consider potential frequency-induced effects in the interpretation of our experimental findings.

The results of this corpus study also demonstrate that *true/sure..., but* is indeed commonly translated to analogous concessive constructions in German. The most frequent translation is *zwar...aber*, but we also observe variations, particularly on the lexical cue at the conceded argument. Other trans-

lations for *true* include typical adjectives and adverbs marking the truth of an assertion, like *stimmt* ('correct'), *sicher(lich)* ('sure(ly)'), or *natürlich* ('natural(y)') (see (16a)). This is not surprising given that they share semantic properties with *zwar* in terms of veridicality (Giannakidou (1998) and subsequent works), i.e. all the variants and *zwar* are veridical operators with regard to the modified proposition. *Zwar* differs from the others as it is exclusively used as concessive marker.

In our sample of *zwar*, on the other hand, the majority of English translations did not use an analogous two-part construction. Instead, English often solely relies on the realization of a single connective, either at the conceded argument (*while, although*) or at the second argument (*but, however,* and others). We also observe the common presence of the modal verb *may* at the conceded argument, typically followed by a connective like *but*. While we do not take a strong stance on how modal verbs signal concessive discourse relations, epistemic modality as signal for concessivity has been discussed for instance by Baranzini and Mari (2019), König (2006), and Souesme (2009).

(16)   a.   *Das ist natürlich ein erster Schritt, dieser bleibt **aber** weit hinter dem Bedarf*
           This is naturally a   first   step,   this   stays but far behind the need
           zurück. (Sentence #1202079, German)
           back
           'This is a first step, *true*, **but** it falls far short of the requirements' (Sentence #1202079, English)

      b.   *Das heißt, die Gesetze sind zwar da,   **aber** sie   müssen auch angewandt werden.*
           That means the laws   are true there but they must   also applied   become
           (Sentence #790455, German)
           'In other words, the laws are in place, **but** they still need to be enforced.'
           (Sentence #790455, English)

To summarize, the parallel corpus data indicate that most instances of *zwar* are not realized with an analogous lexical signal in English. Nevertheless, *true/sure* as lexical cues indeed represent a low-frequency option to overtly mark the conceded argument of English concessive constructions.

3.4 Summary

With respect to signaling a concessive relation, the data reported in sections 3.1 and 3.2 demonstrate that both the German and the English lexical cues are good indicators of a concessive discourse relation. Still, it is an open question whether the pattern observed in the corpus data will actually translate to comprehenders' use of these cues in on-line processing. Thus, the chosen lexical cues are well-suited for an investigation of discourse expectations generated by lexical cues. The parallel corpus data reported in section 3.3 further showed that concessive *true/sure* constructions are commonly translated to German *zwar* constructions. This supports our decision to compare the two in a cross-linguistic experiment. The latter data set also showed that *zwar* is much more frequent than *true/sure*. Potential frequency effects will thus have to be taken into consideration when interpreting the results of our study.

4. Testing expectations in discourse processing / Experiments

We conducted two experiments in German and in English, using a combination of self-paced reading and rating tasks with reading time and naturalness ratings as measures. Both experiments were

based on a fully factorial design wherein we varied the amount of cues towards the concessive relation. For the naturalness ratings, we predicted that contextual cues in particular would increase the naturalness as they create a coherent discourse by closely relating the two sentences of our experimental items. Furthermore, due to the lower frequency *true/sure...but*, items containing these lexical cues may be rated as less natural than the same constructions with no lexical cues, as well as less natural than the German *zwar....aber* construction.

Concerning the reading time, our hypotheses were as follows: First, if readers can anticipate the discourse relation from lexical or contextual cues, we should find a reduced reading time at the connective when the preceding context contains either of these cues, as compared to a baseline condition without cues. Second, lexical and contextual cues may not lead to equally large expectation-based effects. As outlined in section 2, the lexical cues are clear indicators of a conceded argument. Comprehenders in principle only need to recognize the cues as such in order to generate an expectation for a concessive discourse relation. The contextual cue condition, on the other hand, requires comprehenders to detect the incoherence between two propositions and to use it to adjust their discourse expectations. The two cues thus give rise to expectations through different means to which comprehenders in turn may not be equally sensitive. Third, if multiple cues predicting the same upcoming discourse relation occur together, we wondered whether they may function as cumulative evidence towards the upcoming discourse relation. If that is the case, we might find that processing is facilitated even beyond the effect of a singular cue.

4.1 German experiment

4.1.1 MATERIALS

28 items were generated based on a 2 x 2 design with the following manipulations: first, the initial sentence (S1) was either neutral with respect to the content of the following sentence (conditions 1 and 3, henceforth C1 and C3), or incongruent with it (C2 and C4). In the latter case, the incongruence acts as cue towards a concessive structure. Thus, conditions 2 and 4 contain a *contextual cue*, whereas conditions 1 and 3 do not. The second experimental manipulation concerned *lexical cues*, such that the second sentence (S2) was lexically cued for the upcoming contrast with *zwar* ('truly') (C1 and C2), or did not contain a lexical cue (C3 and C4). With these two manipulations our items looked as follows: the first sentence always introduced an action that the agent likes to perform. It consisted of a proper name, followed by an intransitive verb and adverbial modifiers. In the -*contextual cue* conditions, the verb was modified by only one adverb, namely *gerne* (lit. 'gladly', akin to English 'to like + [verb gerund]'). In the +*contextual cue* conditions, it was modified by two adverbs, namely *gerne* and a second adverb that specifics the condition under which the action is performed. Thus, as can be seen in (8), the context sentence *Jens läuft gerne* ('Jens likes to run.') in C1 and C3 is underspecified with regard to the running conditions that the agent prefers, whereas the explicit context sentence *Jens läuft gerne draußen* ('Jens likes to run outside.') in C2 and C4 specifies exactly the way that the agent prefers to do the running exercises. The following sentence then introduces a second fact about the agent, for instance that they own a treadmill. This fact is always congruent with the neutral context sentence, but incongruent with the contextual cue condition. In C1 and C2, the sentence was lexically cued with *zwar*. In C3 and C4, there was no lexical cue. Finally, a coordinating clause starting with the discourse connective *aber* ('but') introduces a concessive relation between the two conjuncts. The region containing the discourse connective is the critical region (CR) at which we expect differences in reading times depending on whether

subjects had anticipated the concessive relation. An example for a test item in all four conditions can be seen in (17).

(17)   1. *Jens / läuft / gerne. / Er / hat / **zwar** / ein Laufband / im   Wohnzimmer, / aber*
              Jens / runs / gladly. / He / has / **true** / a   treadmill / in-the living-room / but
              er$_{CR}$ / joggt / häufig / im   Park.
              he$_{CR}$ / jogs / often   / in-the park
              'Jens likes to run. **True**, he has a treadmill in the living room, but he often jogs in the park.'

        2. *Jens / läuft / gerne **draußen**. / Er / hat / **zwar** / ein Laufband / im   Wohnzimmer, /*
              Jens / runs / gladly **outdoors**. / He / has / **true** / a   treadmill / in-the living-room /
              aber er$_{CR}$ / joggt / häufig / im   Park.
              but he$_{CR}$ / jogs / often   / in-the park
              'Jens likes to run **outdoors**. **True**, he has a treadmill in the living room, but he often jogs in the park.'

        3. *Jens / läuft / gerne. / Er / hat / ein Laufband / im   Wohnzimmer, / aber er$_{CR}$ /*
              Jens / runs / gladly. / He / has / a   treadmill / in-the living-room / but he$_{CR}$ /
              joggt / häufig / im   Park.
              jogs / often   / in-the park
              'Jens likes to run. He has a treadmill in the living room, but he often jogs in the park.'

        4. *Jens / läuft / gerne **draußen**. / Er / hat / ein Laufband / im   Wohnzimmer, / aber*
              Jens / runs / gladly **outdoors**. / He / has / a   treadmill / in-the living-room / but
              er$_{CR}$ / joggt / häufig / im   Park.
              he$_{CR}$ / jogs / often   / in-the park
              'Jens likes to run **outdoors**. He has a treadmill in the living room, but he often jogs in the park.'

In addition, the experiment also used 68 filler items, each consisting of 2 sentences to ensure uniformity with the experimental items. We used two types of fillers that were tested for the purpose of another experiment not reported here. An example for each type of filler can be seen in (18). The order of experimental items and fillers was pseudo-randomized following a Latin-square design, such that every participant saw only one condition of each item and two experimental items were always interspersed with at least two fillers. Participants saw a total of 96 prompts in the experiment.

(18)   a. *Nils / sitzt / in einer Firmenkantine.   / Besonders häufig / trifft / einer der*
              Nils / sits / in a   company-canteen. / Very   often / meets / one the.GEN
              Mitarbeiter / in der Pause / einen der   Kollegen, / die / Kinder / haben.
              employees / in the break / one   the.GEN colleagues / who / children / have.
              'Nils sits in a company's canteen. One of the employees very often meets one of the colleagues who have children in the break.'

        b. *Hannes / geht / einkaufen. / Gerne / berät / derjenige Verkäufer, / dessen Laden*
              Hannes / goes / shopping. / Happily / advises / the-one   salesperson / whose store
              / echte Delikatessen / anbietet, / bei Unklarheiten / den Kunden.
              / real   delicacies   / offers   / at unclarities   / the customer
              'Hannes goes shopping. The salesperson, whose store offers real delicacies, happily advises the customer in case of any unclarities.'

4.1.2 PARTICIPANTS

57 students (40 female, mean age = 22.47, sd = 2.9) from Osnabrück University participated in the study. All participants were native speakers of German and had normal or corrected-to-normal vision. Participants gave written informed consent prior to participating and were compensated with partial course credits or monetary payment. The experiment was approved by the ethics committee of Osnabrück University.

4.1.3 PROCEDURE

The experiment was programmed and presented on Ibex Farm (Drummond, 2013). Items were presented on a screen in a moving-window self-paced reading paradigm. The items appeared in words or phrases such as indicated through the slashes in (17) and (18). For every item, the two sentences appeared in the middle of the screen, with the second sentence appearing below the first one. Before reading an item, subjects thus saw a white screen with two lines of grey dashes where the words would appear. While reading, they used the space bar to reveal the next section, causing the previous one to disappear. After reading an item, a question asking for a naturalness rating of the sentences would appear on a new screen. We used a 7-point Likert scale with the endpoints marked as 'unnatural' (1) and 'natural' (7). In 25 out of the 96 items (distributed randomly over the experiment) the naturalness rating was further followed by a comprehension question appearing on a new screen. This question could be answered with 'yes' or 'no' based on the content of the item the participant had just read. The experimental session began with 4 practice trials, during which participants were still allowed to ask questions about the experimental procedure. In total, the experiment took about 35 minutes.

4.1.4 RESULTS

Reading times and naturalness ratings were analyzed using linear mixed-effects regression models using the lme4 package (Bates et al., 2018) in R (R Development Core Team, 2019).[7] Prior to the analysis, all participants with an accuracy rate below 80% (<20 correct answers) on the comprehension questions were removed from the data set. We thus had to remove 7 out of 57 participants. All following analyses were performed on the data set with 50 remaining subjects. To remove outliers in the reading times, we took the whole data set and removed all reading times more than 2 standard deviations from the mean for each subject's reading time per condition and region. Furthermore, reading times below 150ms were removed. The outlier removal process affected 1.5% of our data. For both naturalness ratings and reading times, our predictors are the binary variables contextual cue (+contextual cue v. -contextual cue) and lexical cue (+lexical cue v. -lexical cue). Both factors were coded as sum contrasts and included as fixed effects (with interaction) in our models (Vasishth and Broe, 2011). Following the Box-Cox procedure (Box and Cox, 1964), we determined that a reciprocal transformation of reading times would be appropriate to ensure a normal distribution of residuals. We hence used the reciprocal RT as dependent measure, which was multiplied by 1000 to make the estimates more interpretable (this does not affect the results in any way). We used the maximal random effects structure that allowed our models to converge (Barr et al., 2013), which, unless otherwise noted, included random intercepts per subject and item. For the reading time analysis, we performed analyses on the critical region (*aber er/sie*, 'but he/she'), as well as the regions

---

7. All data and code associated with this experiment are available from: `https://osf.io/ux8de/`

Condition	Rating	RT CR-1	RT CR	RT CR+1
C1: -context cue,+lexical cue	6.21 (1.09)	737.36 (332.45)	550.05 (201.87)	477.76 (161.50)
C2: +context cue,+lexical cue	6.36 (1.05)	754.65 (367.82)	519.44 (145.40)	469.99 (126.80)
C3: -context cue,-lexical cue	6.13 (1.23)	738.15 (312.12)	583.51 (239.98)	486.48 (139.53)
C4: +context cue,-lexical cue	6.27 (1.12)	754.53 (350.05)	566.87 (213.26)	468.67 (118.75)

Table 4: Mean naturalness ratings (on a scale of 1-7) and reading times (in ms) for the pre-critical, critical, and post-critical region with standard deviations in parentheses

preceding (CR-1), and following it (CR+1). Effects in self-paced reading are often delayed or distributed across multiple (adjacent) regions (Smith and Levy, 2013; Witzel et al., 2012), therefore we include the post-critical region CR+1 to check for delayed effects and effects that may span across both the CR and CR+1 regions. Additionally, we include an analysis on the pre-critical region CR-1 to ensure that any effects found at the critical region are not caused by spillover from previous material. The mean ratings and raw reading times are provided in Table 4.

Naturalness ratings   The model revealed a main effect of contextual cues (β = -0.08, SE = 0.02, t = -3.35, p = 0.0008). Items with a contextual cue were rated more natural than those without. No other effect was significant.

Reading times   In Figure 1, we provide a visualization of the reading times over the course of both sentences, with a close-up on the regions of interest.

There were no significant effects at the pre- or postcritical regions. At the critical region we found a significant effect of contextual cues (β = -0.03, SE = 0.01, t = -2.42, p $<$ 0.05) and a highly significant effect of lexical cues (β = -0.05, SE = 0.01, t = -4.81, p $<$ 0.0001), but no interaction. The critical region was read faster if the sentence was explicitly cued with *zwar*, or if the context sentence contained a contextual cue, but the two effects were independent of each other. That is, the lexical and the contextual cues together did not cause a superadditive reduction in reading times.

4.2 English experiment

4.2.1 MATERIALS

24 items were generated based on the 2 x 2 design of the German experiment. Again, the first sentence was either neutral (C1 and C3) or incongruent with the second one (C2 and C4), that is, the former contained no contextual cue whereas the latter did. As in the German experiment, the first sentence always introduced an action that the agent likes to perform. In the +contextual cue conditions, this description was additionally modified by an adverb that specified how the agent prefers to do the action. The second sentence was either left unmarked for the upcoming contrast (C1 and C2) or contained a lexical cue (C3 and C4), namely a sentence-initial particle (*true* in 12 items, *sure* in the other 12 items). The crucial concessive relation was signaled by the discourse connective *but*. An example item in all four conditions can be seen in (19). We used 48 filler items, all containing 2 sentences for uniformity with the experimental items. An example for the fillers is given in (20). The order of experimental items and fillers was pseudo-randomized following a Latin-square design, such that every participant saw only one condition of each item and two experimental items were always interspersed with at least two fillers. Participants saw a total of 72 prompts in the experiment.
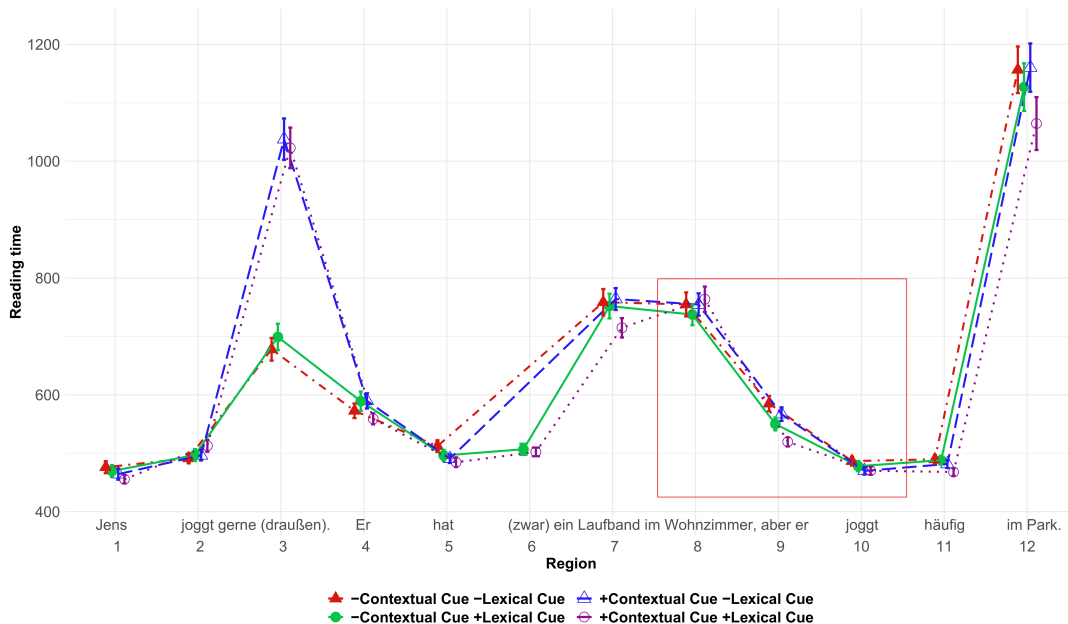
(a) Mean reading times for all regions
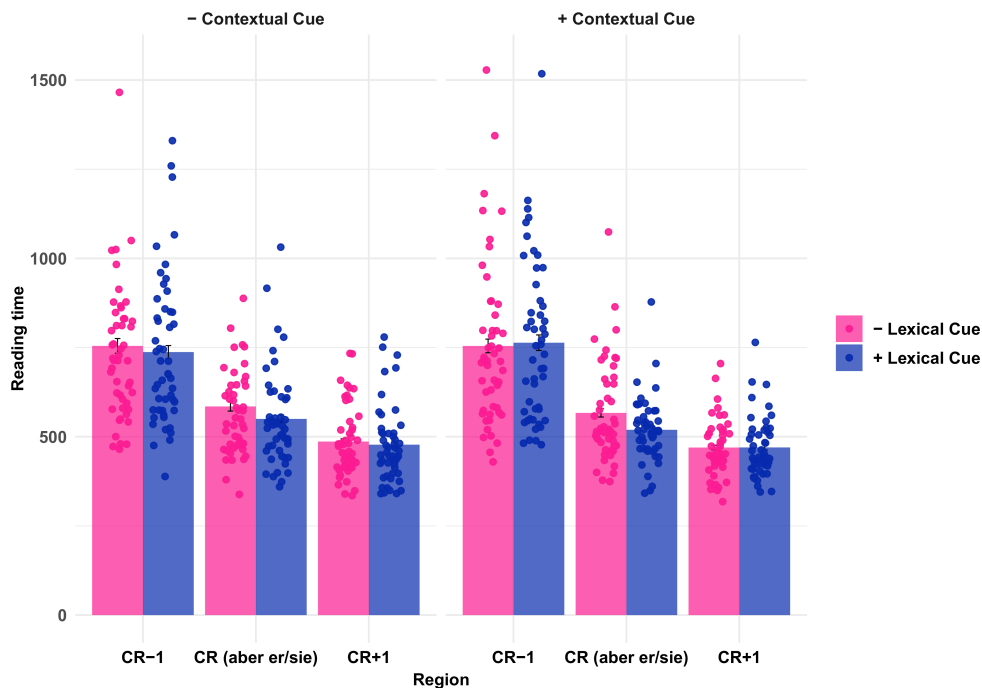
(b) Mean reading times for the critical region (*aber er/sie*, 'but he/she'), as well as the CR-1 and CR+1.

Figure 1: Mean reading times (ms) over (a) all regions in the 4 different conditions and (b) the pre-critical to post-critical region. The red rectangle in (a) marks the three regions displayed in (b). Error bars indicate standard errors. Each data point in (b) represents the average reading time of one subject.

(19) 1. James / likes /to run. / **True**, / he / has / a treadmill / in the living room, / but he_{CR} / often / jogs / in parks.

2. James /likes /to run **outdoors**. / **True**, / he / has / a treadmill / in the living room, / but he_{CR} / often / jogs / in parks.

3. James /likes /to run. / He / has / a treadmill / in the living room, / but he_{CR} / often / jogs / in parks.

4. James /likes /to run **outdoors**. / He / has / a treadmill / in the living room, / but he_{CR} / often / jogs / in parks.

(20) Vincent / lives / in a big city. / Because / he / knows / his way / around the city, / he / can give / directions / to pedestrians.

4.2.2 PARTICIPANTS

88 participants (32 female, 1 non-binary, mean age = 35.34, sd =10.42) were recruited via the crowd sourcing platform Amazon Mechanical Turk. The participants were located in the United States of America and declared that they were native speakers of English. Participants were informed about the general nature and duration of the experiment and gave consent to participating by ticking a box on the website. They were compensated with monetary payment. The experiment was approved by the ethics committee of Osnabrück University.

4.2.3 PROCEDURE

The experiment was programmed on Ibex Farm. The procedure was the same as in the German study. To ensure that all participants read attentively, the naturalness rating was always followed by a comprehension question appearing on a new screen. This question could be answered with 'yes' or 'no' based on the content of the item the participant had just read. The experimental session began with 4 practice trials. In total, the experiment took about 30 minutes.

4.2.4 RESULTS

Reading times and naturalness ratings were analyzed using linear mixed-effects regression models using the lme4 package in R.[8] Prior to the analysis, all participants with an accuracy below 80% (<58 correct answers) on the comprehension questions were removed from the data set. We thus had to remove 21 out of 88 participants. All following analyses were performed on the data set with 67 remaining subjects. To remove outliers in the reading times, we took the whole data set and removed all reading times more than 2 standard deviations from the mean for each subject's reading times per condition and region. Furthermore, reading times below 150ms were removed. The outlier removal process affected 3.7% of our data. The models for both the naturalness ratings and the reading times were constructed under the same specifications as in Experiment 1. The mean ratings and reading times are provided in Table 5.

As indicated above, 12 of our items used *true* as lexical cue, the other 12 used *sure*. Thus, our first objective in the analysis was to address whether *true* differed in any significant way from *sure* in either naturalness ratings or reading times. We constructed separate models for the ratings, the critical region and the post-critical region over the part of the data set that contained a lexical cue (conditions 1 and 2). In addition to the fixed and random effects specified above, we included

---

8. All data and code associated with this experiment are available from: `https://osf.io/ux8de/`

Condition	Rating	RT CR-1	RT CR	RT CR+1
C1: -context cue,+lexical cue	5.73 (1.26)	621.31 (424.19)	488.41 (251.35)	378.43 (143.84)
C2: +context cue,+lexical cue	5.65 (1.34)	619.06 (413.51)	470.49 (230.60)	370.71 (145.14)
C3: -context cue,-lexical cue	5.94 (1.23)	649.57 (463.89)	477.31 (211.80)	399.72 (222.36)
C4: +context cue,-lexical cue	6.07 (1.17)	679.00 (583.75)	478.26 (211.80)	387.92 (155.16)

Table 5: Mean naturalness ratings (on a scale of 1-7) and reading times (in ms) for the pre-critical, critical, and post-critical region with standard deviations in parentheses

a binary predictor for the type of lexical cue (that is, whether the item in question used *true* or *sure* as lexical cue) with an interaction term. None of the models showed any significant effect of the lexical choice (*true* vs. *sure*), nor any interaction with the other predictor. Nevertheless, we compared the models containing the type of lexical cue as predictor to models without said predictor using the likelihood ratio test, which assesses the goodness of fit between two models. The models containing the additional predictor were found not to have a better fit for our data in either the naturalness ratings (X^2 = 0.78, df = 2, p = 0.68), or the regions of interest for our reading time analysis (**CR**: X^2 = 2.03, df = 2, p = 0.36; **CR+1**: X^2 = 1.12, df = 2, p = 0.57). Thus, we decided to conflate the two lexical cues in all further analyses.

Naturalness ratings The model, whose fixed effects included the two binary predictors, and whose random effects structure included random by-item and by-subject intercepts, as well as random by-subject slopes for both fixed effects, revealed a significant interaction between the contextual and lexical cue (β = -0.05 SE = 0.02, t = -2.16, p < 0.05). The cause of this interaction effect was investigated in post-hoc paired comparisons via Tukey's HSD: in the presence of a contextual cue, the lexically cued condition was rated less natural than the uncued condition (β = 0.41, SE = 0.10, df = 112.9, t = 4.30, adjusted p < 0.001). However, if contextual cues were absent, lexically cued and uncued conditions received similar ratings (β = 0.21, SE = 0.10, df = 112.9, t = 2.18, adjusted p = 0.14). The condition containing only a contextual cue (C4) was rated more natural than the condition containing only a lexical cue (C1) (β = -0.34, SE = 0.10, df = 65.5, t = -3.55, adjusted p < 0.01)

Reading times We provide a visualization of the reading times over the course of both sentences, with a close-up on the regions of interest in Figure 2.

There were no significant effects at the pre-critical or the critical region. At the post-critical region (CR+1), we found a significant effect of lexical cues (β = -0.04, SE = 0.02, t = -2.58, p < 0.01), such that the region was read faster if the sentence had been cued with *true/sure*. There was no effect of contextual cues, and no interaction.

5. Discussion

The rapid identification of coherence relations is a key component of successful discourse processing. In the current study, we used a combination of self-paced reading and rating tasks to address the question whether the on-line process of establishing a discourse relation is aided by expectations generated from cues preceding the discourse connective. The results from the German experiment show that readers immediately integrate information from lexical and contextual cues to update their discourse expectations. The results from the English study confirm the importance of lexical
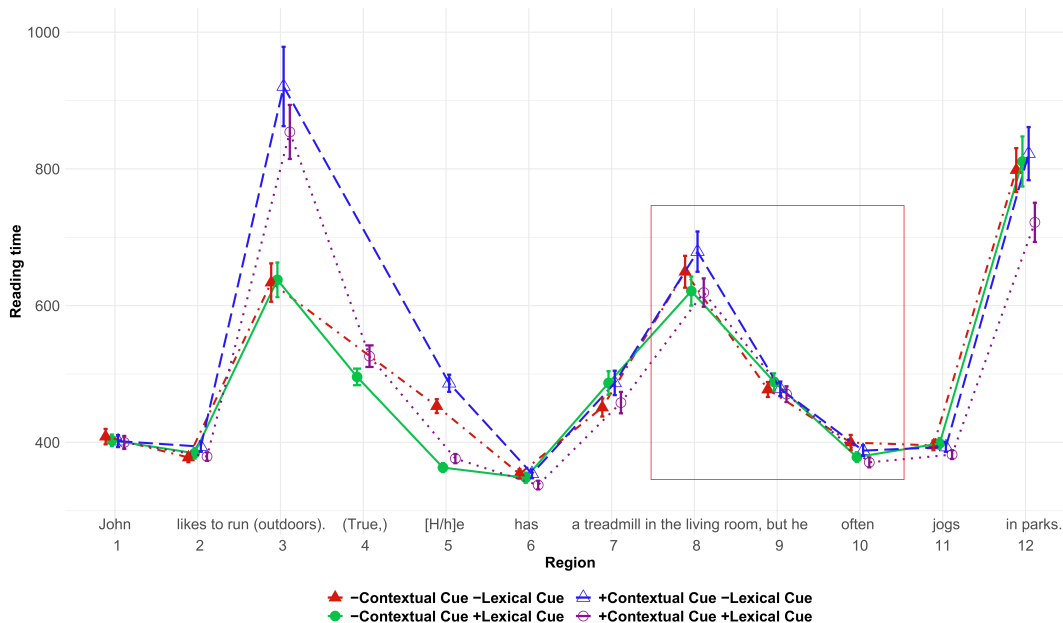
(a) Mean reading times for all regions
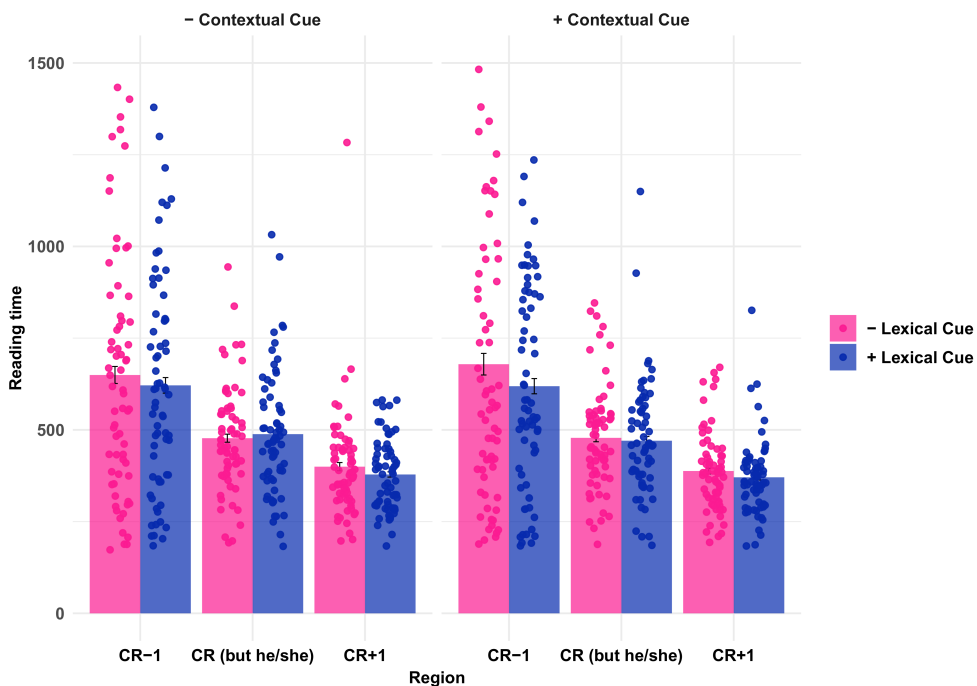
(b) Mean reading times for the critical region (*but he/she*), as well as the CR-1 and CR+1.

Figure 2: Mean reading times (ms) over (a) all regions in the 4 different conditions and (b) the pre-critical to post-critical region. The red rectangle in (a) marks the three regions displayed in (b). Error bars indicate standard errors. Each data point in (b) represents the average reading time of one subject.

cues, but do not show any facilitative effect of contextual cues, at least in on-line processing. In the following, we will discuss the findings for German and English in turn, and will conclude with an elaboration on the cross-linguistic variation we observed.

The German experiment can be summarized as such: the reading time results demonstrate that readers use both contextual and lexical cues to anticipate discourse relations, although lexical cues showed a numerically stronger effect. Given the reliability of the lexical cue attested by the corpus data presented in section 3, its power for generating discourse expectations conforms with our hypotheses. The more surprising aspect of our results is the effect of contextual cues: for one, in the absence of the lexical cue *zwar*, readers were able to immediately connect the content of the second sentence to its embedding context and to generate an expectation that mitigated the absence of any further discourse signals. Secondly, the contextual cue maintained an independent facilitative effect even when it appeared together with the lexical cue. Our study thus provides evidence that the joint presence of discourse cues from multiple linguistic sources can act as cumulative facilitators of discourse processing.

Lastly, the naturalness ratings for the German experiment showed that the items were rated more natural in the conditions containing a contextual cue, suggesting that the presence of contextual cues may have generated a more coherent discourse. This is also in line with the earlier observation that through the addition of the contextual cue the invited causal inference in the conceded argument is more specific than in conditions without contextual cues. Lexical cues had no effect on the ratings, which hints that the off-line judgments are not a direct reflection of processing ease, but rather an indicator for the overall coherence of the stimulus item. Given that the on-line measure operates on incomplete linguistic input to offer insight into incremental processing, while the off-line measure elicits a response after the full item has been presented, differences between the two measures are not entirely surprising.

In the English experiment, the reading times confirmed a facilitative effect of lexical cues on processing the concessive relation, albeit at delayed impact. However, contextual cues were found not to have any effect in the on-line processing results. Delayed effects are not uncommon in self-paced reading, but the obvious absence of an effect that was found in German requires further exploration. We will discuss possible explanations when we turn to the comparison of the German and English data below. Critically, despite the low frequency of *'true/sure...but'* constructions reported in section 3, the lexical cue had a facilitative effect on processing. Earlier, we have raised the possibility that (at least some) comprehenders may not be able to generate discourse expectation from *true/sure* due to limited familiarity with this type of construction. Overall, however, this is not the case. Interestingly, the effect of the lexical cue was smaller in English than in German, which might be a reflection of a reduced utility of *true/sure* as discourse cue. We leave this issue for future research.

While it is possible that the infrequency of the *true/sure* construction would cause general processing difficulties, such effects are not apparent in our regions of interest. Neither the pre-critical (CR-1), nor the critical region (CR) showed elevated reading times for the conditions with a lexical cue. Thus, even if subjects experienced a temporary processing difficulty at *true/sure*, these appear to have subsided by the time the critical region was reached. In any case, frequency effects would predict longer reading times if the sentence contained *true/sure*, not shorter ones. Thus, the reduced reading times at the post-critical region cannot be attributed to a confound of the structural frequency.

As for the naturalness ratings, several findings stand out: for one, the two conditions containing a lexical cue received the lowest numerical ratings (C1: 5.73 and C2: 5.65 on average). We believe that this may be a reflection of the relatively low frequency of '*true/sure,...*' constructions in natural language, as reported in section 3.3. Secondly, we find that the condition containing a contextual cue, but no lexical cue, was rated the most natural. In analogy to the German results, the context may thus have added to the discourse coherence. Further adding a lexical cue to contextually cued items significantly worsened the ratings, suggesting that subjects dispreferred lexically cueing a concessive discourse relation if the contextual information already signals a concession.

Together, the two experiments provide converging evidence for the effects of contextual and lexical cues in discourse processing. For both languages, the off-line ratings suggest that contextual cues improve the coherence of the discourse, while the on-line results indicate that lexical cues facilitate processing through the anticipation of upcoming discourse relations. The German study further demonstrated an on-line effect of contextual cues which was absent in English.

One may speculate whether general differences between the languages could have contributed to the divergence between German and English on the contextual cue effect. If anything, however, we would have expected English native speakers to rely on contextual cues more heavily, given that a two-part marking as with '*true/sure...but*' or '*zwar...aber*' is generally less frequent in English than it is in German. We therefore think this question deserves further investigation in the future.

For the current experiments, we suspect that a likely contributor to the divergence between German and English is the difference in data acquisition. We would like to draw attention to the fact the German experiment was conducted in the lab of the first author's home university, while the English experiment has been conducted online over Amazon Mechanical Turk. In general, the replicability of lab-based experiments in linguistics through crowdsourcing is well established (Enochson and Culbertson, 2015; Munro et al., 2010; Schnoebelen and Kuperman, 2010; Snow et al., 2008). However, there are a number of factors adding extra variability to crowdsourced data (broader demographics, e.g. on the age and educational level of participants, lack of control over experimental setting, and limited control over maintained attentiveness during the experiment). For subtle pragmatic effects like our context manipulation, these factors may indeed make a difference. As with most lab-based experiments, the participants for our German experiment were closely matched with regard to age and education level (undergraduate and graduate students with ages ranging from 18-30). In our English experiment, participants were on average older (age range 19-70). In addition, we used a comprehension question after every item as quality control measure, and removed all participants with a response accuracy below 80%. A full quarter of MTurk participants had to be excluded, hinting that they were on average less attentive readers than our lab participants. While those participants remaining in the sample passed our comprehension questions, we think it is likely that there still important differences remaining between the lab-based and crowdsourced data.

Past studies on the demographics of MTurk workers, for instance, suggest that they cover a broad range of education levels and socio-economic statuses (Ross et al., 2010). As predictive processing during language comprehension has been suggested to be affected by age (though the literature is unclear on the effect of age, with some suggesting a decreased use of prediction (Federmeier et al., 2010; Wlotko et al., 2012), while others have found either no clear effect (Dave et al., 2018), or an effect in the opposite direction (Cheimariou, 2016)), working memory (Huettig and Janse, 2016), as well as language and literacy skills (Kukona et al., 2016; Falkauskas and Kuperman, 2015), the increased variability in our English participants' backgrounds may mean that, across our sample, not all participants may engage in predictive language processing as reliably as the university students

participating in the German experiment. For that reason, we assume that a contextual cue effect could still be present in English, and detectable with a larger sample size. We defer a thorough investigation of this question to future studies.

It is a longstanding idea that discourse relations are signaled through complex means that include, but go beyond, the presence of discourse markers (Asr and Demberg, 2015; Das and Taboada, 2018; Hoek et al., 2018; Prasad et al., 2010). Our study provides experimental evidence that confirms that readers are sensitive to subtle discourse cues, both in judging the coherence of discourse and in processing discourse relations. We have shown that readers can rapidly integrate multiple cues to generate and update expectations on how the discourse will unfold. We would like to conclude this paper by discussing what we consider to be fruitful avenues for further research. As language is a means for interpersonal communication, we ought to integrate both speaker and listener data for a better understanding of the communicative function of discourse cues outside of discourse connectives. While we have identified and given evidence for one of their functions, namely that they provide a processing benefit to readers, we must ask ourselves (i) whether this processing benefit is among the deciding factors as to the speaker's choice to add or delete discourse signals or contextual information towards a discourse relation, and (ii) what other factors affect the speaker's choice. A speaker may be more inclined to include additional cues in their utterance when the cues result in a strong processing benefit for the reader or listener, that is, for instance in circumstances where establishing the discourse relation would be difficult otherwise. There are at least two possibilities. First, some connectives are ambiguous with regard to their signaled coherence relations, such that additional cues may allow the comprehender to disambiguate the intended relation. For example, *aber* can be used as a device to change topics, such as in (21).

(21) *Danke, mir geht es gut.   Aber wie geht es dir?*
      Thanks, me goes it good. But   how goes it you?
      'Thanks, I am doing fine. But how are you?'

The use of *aber* in (21) is incompatible with *zwar*. Thus, the presence of *zwar* can exclude alternative readings and narrow down the discourse relation to a concessive one. Second, in case of a long distance between the two related propositions, the prediction from discourse cues enables comprehenders to keep the first proposition active in memory until the expected discourse continuation is fulfilled. An example for the latter point are intersentential discourse relations with intervening sentences as in (15b), repeated as (22) below.

(22) *True*, she probably still had a long way to go. She was painfully thin and there was an insubstantiality about her. **But** there was no denying that today his wife was better than he had known her for many many months. (BNC_CDE_222-224)

Future research on a wider range of discourse relations and text/dialogue genres will have to determine if these factors can be shown to systematically predict whether a speaker will include additional discourse cues. Simultaneously, a thorough investigation of comprehenders' use of such cues in the on-line processing of various discourse relations can complement the speaker data. In the end, we believe that integrating insight from both perspectives is a promising approach towards a better understanding of discourse-level communication.

Acknowledgements

Disclosure statement

Funding

Notes on contributors

J.S. and M.L. conceived and planned all the studies in the paper. J.S. carried out the experiments and corpus studies, and conducted data analyses. Both authors contributed to the interpretation of the results. J.S. took the lead in preparing the manuscript, with contributions from M.L.

References

Charles Antaki and Margaret Wetherell. Show concessions. *Discourse Studies*, 1(1):7–27, 1999. doi: 10.1177/1461445699001001002.

Fatemeh Torabi Asr and Vera Demberg. Uniform information density at the level of discourse relations: Negation markers and discourse connective omission. In *Proceedings of the 11th International Conference on Computational Semantics*, pages 118–128, London, UK, 2015. URL `https://www.aclweb.org/anthology/W15-0117.pdf`.

Fatemeh Torabi Asr and Vera Demberg. Interpretation of discourse connectives is probabilistic: Evidence from the study of *but* and *although*. *Discourse Processes*, 57(4):376–399, 2020. doi: 10.1080/0163853X.2019.1700760.

Laura Baranzini and Alda Mari. From epistemic modality to concessivity: Alternatives and pragmatic reasoning per absurdum. *Journal of Pragmatics*, 142:116–138, 2019. doi: 10.1016/j.pragma.2019.01.002.

Dale J. Barr, Roger Levy, Christoph Scheepers, and Harry J. Tily. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3): 255–278, 2013. doi: 10.1016/j.jml.2012.11.001.

Douglas Bates, Martin Maechler, Ben Bolker, Steven Walker, Rune H. B. Christensen, Henrik Singmann, Bin Dai, Fabian Scheip, Gabor Grothendieck, Peter Green, and John Fox. Package 'lme4', 2018.

George E. P. Box and David R. Cox. An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(2):211–243, 1964. doi: 10.1111/j.2517-6161.1964. tb00553.x.

Lynn Carlson, Daniel Marcu, and Mary Ellen Okurowski. RST discourse treebank, LDC2002T07, 2002. URL `https://catalog.ldc.upenn.edu/LDC2002T07`.

Spyridoula Cheimariou. *Prediction in aging language processing*. PhD thesis, University of Iowa, 2016.

Debopam Das and Maite Taboada. Signalling of coherence relations in discourse, beyond discourse markers. *Discourse Processes*, 55(8):743–770, 2018. doi: 10.1080/0163853X.2017.1379327.

Debopam Das, Tatjana Scheffler, Peter Burgonje, and Manfred Stede. Constructing a lexicon of English discourse connectives. In *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue*, pages 360–365, Melbourne, AUS, 2018. doi: 10.18653/v1/W18-5042.

Shruti Dave, Trevor A. Brothers, Matthew J. Traxler, Fernanda Ferreira, John M. Henderson, and Tamara Y. Swaab. Electrophysiological evidence for preserved primacy of lexical prediction in aging. *Neuropsychologia*, 117:135–147, 2018. doi: 10.1016/j.neuropsychologia.2018.05.023.

Heiner Drenhaus, Vera Demberg, Judith Köhne, and Francesca Delogu. Incremental and predictive discourse processing based on causal and concessive discourse markers: ERP studies on German and English. In *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*, pages 403–408, Québec, CAN, 2014. URL `https://escholarship.org/uc/item/9q88v0zh`.

Alex Drummond. Ibex Farm, 2013. URL `http:spellout.net/ibexfarm`.

Kelly Enochson and Jennifer Culbertson. Collecting psycholinguistic response time data using Amazon Mechanical Turk. *PLoS ONE*, 10(3), 2015. doi: 10.1371/journal.pone.0116946.

Kaitlin Falkauskas and Victor Kuperman. When experience meets language statistics: Individual variability in processing English compound words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(6):1607–1627, 2015. doi: 10.1037/xlm0000132.

Kara D. Federmeier, Marta Kutas, and Rina Schul. Age-related and individual differences in the use of prediction during language comprehension. *Brain and Language*, 115(3):149–161, 2010. doi: 10.1016/j.bandl.2010.07.006.

Fernanda Ferreira, John M. Henderson, Michael D. Anes, Phillip A. Weeks, and David K. McFarlane. Effects of lexical frequency and syntactic complexity in spoken-language comprehension: Evidence from the auditory moving-window technique. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(2):324–335, 1996. doi: 10.1037/0278-7393.22.2.324.

Anastasia Giannakidou. *Polarity sensitivity as (non) veridical dependency*. John Benjamins Publishing Company, Amsterdam/Philadelphia, 1998. doi: 10.1075/la.23.

Susanne Günthner. Concessive patterns in interaction: uses of zwar...aber ('true...but')-constructions in everyday spoken German. *Language Sciences*, 58:144–162, 2016. doi: 10.1016/j.langsci.2016.02.009.

Lea A. Hald, Esther G. Steenbeek-Planting, and Peter Hagoort. The interaction of discourse context and world knowledge in online sentence comprehension. evidence from the N400. *Brain Research*, 1146(1):210–218, 2007. doi: 10.1016/j.brainres.2007.02.054.

Jet Hoek, Sandrine Zufferey, Jacqueline Evers-Vermeul, and Ted J.M. Sanders. The linguistic marking of coherence relations: Interactions between connectives and segment-interal elements. *Pragmatics & Cognition*, 25(2):276–309, 2018. doi: 10.1075/pc.18016.hoe.

Falk Huettig and Esther Janse. Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Language, Cognition and Neuroscience*, 31(1):80–93, 2016. doi: 10.1080/23273798.2015.1047459.

Christina S. Kim, Christine Gunlogson, Michael K. Tanenhaus, and Jeffrey T. Runner. Context-driven expectations about focus alternatives. *Cognition*, 139:28–49, 2015. doi: 10.1016/j.cognition.2015.02.009.

Philipp Koehn. Europarl: A parallel corpus for statistical machine translation. In *Machine Translation Summit X*, pages 79–86, Phuket, THA, 2005. URL http://www.mt-archive.info/MTS-2005-Koehn.pdf.

Judith Köhne and Vera Demberg. The time-course of processing discourse connectives. In *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*, pages 2760–2765, Berlin, GER, 2013. URL https://escholarship.org/uc/item/3ng7w640.

Ekkehard König. Concessive clauses. In Keith Brown, editor, *Encyclopedia of Language & Linguistics*, pages 820–824. Elsevier Ltd, 2006. doi: 10.1016/B0-08-044854-2/00277-7.

Ekkehard Konig and Peter Siemund. Causal and concessive clauses: Formal and semantic relations. In Elizabeth Couper-Kuhlen and Bernd Kortmann, editors, *Cause - Condition - Concession - Contrast: Cognitive and Discourse Perspectives*, pages 341–360. Mouton de Gruyter, Berlin/New York, 2000. doi: 10.1515/9783110219043.4.341.

Anuenue Kukona, David Braze, Clinton L. Johns, W. Einar Mencl, Julie A. Van Dyke, James S. Magnuson, Kenneth R. Pugh, Donald P. Shankweiler, and Whitney Tabor. The real-time prediction and inhibition of linguistic outcomes: Effects of language and literacy skill. *Acta Psychologica*, 171:72–84, 2016. doi: 10.1016/j.actpsy.2016.09.009.

Roger Levy. Expectation-based syntactic comprehension. *Cognition*, 106(3):1126–1177, 2008. doi: 10.1016/j.cognition.2007.05.006.

Ashley G. Lewis and Marcel Bastiaansen. A predictive coding framework for rapid neural dynamics during sentence-level language comprehension. *Cortex*, 68:155–168, 2015. doi: 10.1016/j.cortex.2015.02.014.

Robert Munro, Steven Bethard, Victor Kuperman, Vicky Tzuyin Lai, Robin Melnick, Christopher Potts, Tyler Schnoebelen, and Harry Tily. Crowdsourcing and language studies: The new generation of linguistic data. In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*, pages 122–130, 2010. URL `https://www.aclweb.org/anthology/W10-0719.pdf`.

Mante S. Nieuwland and Jos J.A. Van Berkum. When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience*, 18(7):1098–1111, 2006. doi: 10.1162/jocn.2006.18.7.1098.

Wolfgang Pfeifer, Wilhelm Braun, Gunhild Ginschel, Gustav Hagen, Anna Huber, Klaus Müller, Heinrich Petermann, Gerlinde Pfeifer, Dorothee Schröter, and Ulrich Schröter. Zwar. In *Etymologisches Wörterbuch des Deutschen (digitalisierte und von Wolfgang Pfeifer überarbeitete Version im Digitalen Wörterbuch der deutschen Sprache)*. 1993. URL `https://www.dwds.de/wb/zwar`.

R. Prasad, A. Joshi, and B. Webber. Realization of discourse relations by other means: Alternative lexicalizations. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, pages 1023–1031, Beijing, CHN, 2010. URL `https://www.aclweb.org/anthology/C10-2118`.

Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi, and Bonnie Webber. The Penn Discourse Treebank 2.0. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC2008)*, pages 2961–2968, Marrakech, MAR, 2008. URL `http://www.lrec-conf.org/proceedings/lrec2008/pdf/754_paper.pdf`.

Antoinette Primatarova-Miltscheva. Zwar, aber – ein zweiteiliges Konnektivum? *Deutsche Sprache*, 14(2):125–139, 1986.

R Development Core Team. R: A Language and Environment for Statistical Computing, 2019. URL `https://cran.r-project.org/`.

Keith Rayner and Susan A. Duffy. Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, 14(3):191–201, 1986. doi: 10.3758/BF03197692.

Hannah Rohde and William S. Horton. Anticipatory looks reveal expectations about discourse relations. *Cognition*, 133(3):667–691, 2014. doi: 10.1016/j.cognition.2014.08.012.

Hannah Rohde, R. Levy, and A. Kehler. Anticipating explanations in relative clause processing. *Cognition*, 118(3):339–358, 2011. doi: 10.1016/j.cognition.2010.10.016.

Joel Ross, Lilly Irani, M. Six Silberman, Andrew Zaldivar, and Bill Tomlinson. Who are the crowdworkers? Shifting demographics in Amazon Mechanical Turk. In *Conference on Human Factors in Computing Systems - Proceedings*, pages 2863–2872, 2010. doi: 10.1145/1753846.1753873.

Tatjana Scheffler and Manfred Stede. Mapping PDTB-style connective annotation to RST-style discourse annotation. In *Proceedings of the 13th Conference on Natural Language Processing*, pages 242–247, Bochum, GER, 2016. URL `https://www.linguistics.rub.de/konvens16/pub/31_konvensproc.pdf`.

Hildur E.H. Schilling, Keith Rayner, and James I. Chumbley. Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences. *Memory and Cognition*, 26(6):1270–1281, 1998. doi: 10.3758/BF03201199.

Tyler Schnoebelen and Victor Kuperman. Using Amazon Mechanical Turk for linguistic research. *Psihologija*, 43(4):441–464, 2010. doi: 10.2298/PSI1004441S.

Merel C. J. Scholman, Vera Demberg, and Hannah Rohde. Signalling with one hand: A cross-linguistic comparison of the facilitative effect of "on the one hand.". *Talk presented at XPRAG.de Workshop "Implicit and Explicit Marking of Discourse Relations", Osnabrück, GER*, 2018.

Merel C.J. Scholman, Hannah Rohde, and Vera Demberg. "on the one hand" as a cue to anticipate upcoming discourse structure. *Journal of Memory and Language*, 97:47–60, 2017. doi: 10.1016/j.jml.2017.07.010.

Nathaniel J. Smith and Roger Levy. The effect of word predictability on reading time is logarithmic. *Cognition*, 128(3):302–319, 2013. doi: 10.1016/j.cognition.2013.02.013.

Rion Snow, Brendan O. Connor, Daniel Jurafsky, Andrew Y. Ng, Dolores Labs, and Capp St. Cheap and fast — but is it good? Evaluating non-expert annotations for natural language tasks. In *EMNLP'08: Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 254–263, 2008. doi: 10.3115/1613715.1613751.

Jean-Claude Souesme. MAY in concessive contexts. In R. Salkie, P. Busuttil, and J. van der Auwera, editors, *Modality in English – Theory and Description*, pages 159–176. De Gruyter Mouton, Berlin/New York, 2009. doi: 10.1515/9783110213331.159.

Manfred Stede and Carla Umbach. DiMLex: A lexicon of discourse markers for text generation and understanding. In *Proceedings of COLING-ACL '98*, Montreal, CAN, 1998. URL `http://www.ling.uni-potsdam.de/~stede/Papers/coling98.pdf`.

Manfred Stede, Tatjana Scheffler, and Amália Mendes. Connective-Lex: A web-based multilingual lexical resource for connectives. *Discours. Revue de linguistique, psycholinguistique et informatique. A journal of linguistics, psycholinguistics and computational linguistics*, 24, 2019.

Maite Taboada and Debopam Das. Annotation upon annotation: Adding signalling information to a corpus of discourse relations. *Dialogue & Discourse*, 3(2):249–281, 2013. doi: 10.5087/dad.2013.211.

The British National Corpus. Version 3 (BNC XML Edition), distributed by Bodleian Libraries, University of Oxford, on behalf of the BNC Consortium, 2007. URL `http://www.natcorp.ox.ac.uk/`.

Peter Uhrig and Thomas Proisl. Treebank.info, 2011. URL `http://treebank.info/`.

Geertje van Bergen and Hans Rutger Bosker. Linguistic expectation management in online discourse processing: An investigation of Dutch inderdaad 'indeed' and eigenlijk 'actually'. *Journal of Memory and Language*, 103:191–209, 2018. doi: 10.1016/j.jml.2018.08.004.

Shravan Vasishth and Michael Broe. *The Foundations of Statistics: A Simulation-based Approach.* Springer Verlag, Berlin/Heidelberg, 2011. doi: 10.1007/978-3-642-16313-5.

Naoko Witzel, Jeffrey Witzel, and Kenneth Forster. Comparisons of online reading paradigms: Eye tracking, moving-window, and maze. *Journal of Psycholinguistic Research*, 41(2):105–128, 2012. doi: 10.1007/s10936-011-9179-x.

Edward W. Wlotko, Kara D. Federmeier, and Marta Kutas. To predict or not to predict: Age-related differences in the use of sentential context. *Psychology and Aging*, 27(4):975–988, 2012. doi: 10.1037/a0029206.

Ming Xiang and Gina Kuperberg. Reversing expectations during discourse comprehension. *Language, Cognition and Neuroscience*, 30(6):648–672, 2015. doi: 10.1080/23273798.2014.995679.

Appendix A
A1: Stimuli for Experiment 1

Below we list the test sentences used for the first experiment. Items were presented in 4 conditions that can be reconstructed by adding either one or both of the optional words in parentheses to the sentence.

(1) Gregor isst gerne (auswärts). Er besitzt (zwar) eine Küche mit Profiausstattung, aber er isst meistens im Restaurant.

(2) Jens läuft gerne (draußen). Er hat (zwar) ein Laufband im Wohnzimmer, aber er joggt häufig im Park.

(3) Frederike lernt gerne (alleine). Sie teilt (zwar) eine Lerngruppe mit Freunden, aber sie büffelt regelmäßig im Alleingang.

(4) Thomas schwimmt gerne (draußen). Er kennt (zwar) ein Hallenbad im Stadtzentrum, aber er fährt oft ans Meer.

(5) Marc wandert gerne (querfeldein). Er kennt (zwar) einen Wanderweg im Wald, aber er läuft oft durchs Walddickicht.

(6) Jan malt gerne (draußen). Er hat (zwar) ein Atelier im Dachgeschoss, aber er malt momentan im Garten.

(7) Daniela spielt gerne (online). Sie hat (zwar) eine Brettspielsammlung im Schrank, aber sie spielt momentan am Computer.

(8) Finn trainiert gerne (gemeinsam). Er besitzt (zwar) eine Hantelbank im Keller, aber er geht täglich ins Fitnessstudio.

(9) Ronja arbeitet gerne (zuhause). Sie teilt (zwar) einen Büroraum mit Kollegen, aber sie arbeitet häufig im Bett.

(10) Tim kickert gerne (draußen). Er nutzt (zwar) eine Sporthalle im Winter, aber er spielt jetzt auf Rasenplätzen.

(11) Nele verreist gerne (spontan). Sie plant (zwar) eine Rundreise im Sommerurlaub, aber sie entscheidet spontan übers Reiseziel.

(12) Petra fotografiert gerne (analog). Sie benutzt (zwar) eine Digitalkamera im Fotostudio, aber sie fotografiert häufig auf Filmrollen.

(13) Lucas arbeitet gerne (tagsüber). Er akzeptiert (zwar) einen Job im Nachtclub, aber er sucht aktuell nach Jobalternativen.

(14) Mayra schreibt gerne (handschriftlich). Sie hat (zwar) einen Computer im Wohnzimmer, aber sie schreibt meistens auf Papier.

(15) Mika arbeitet gerne (handwerklich). Er hat (zwar) einen Bürojob im Rathaus, aber er schreinert täglich im Hobbykeller.

(16) Dennis klettert gerne (sicherungsfrei). Er besucht (zwar) einen Seilgarten mit Freunden, aber er klettert meist im Gebirge.

(17) Sina trainiert gerne (frühmorgens). Sie leitet (zwar) einen Abendkurs im Sportverein, aber sie trainiert regelmäßig vorm Arbeiten.

(18) Albert shoppt gerne (online). Er besucht (zwar) einen Laden im Einkaufsviertel, aber er bestellt meistens übers Internet.

(19) Helena wandert gerne (alpin). Sie unternimmt (zwar) eine Wattwanderung im Urlaub, aber sie läuft oft im Gebirge.

(20) Jana lernt gerne (visuell). Sie besitzt (zwar) ein Textbuch zum Prüfungsthema, aber sie lernt momentan über Videos.

(21) Lena singt gerne (zuhause). Sie besucht (zwar) eine Karaokebar mit Freunden, aber sie verbleibt aktuell als Zuschauerin.

(22) Hannah liest gerne (abends). Sie bekommt (zwar) eine Zeitung am Morgen, aber sie liest täglich beim Abendessen.

(23) Susanne musiziert gerne (gemeinsam). Sie beherrscht (zwar) ein Solostück im Klavier, aber sie spielt jetzt im Orchester.

(24) Johannes spielt gerne (drinnen). Er kennt (zwar) einen Spielplatz im Stadtpark, aber er bleibt momentan im Zimmer.

(25) Hannes tanzt gerne (allein). Er besucht (zwar) einen Tangokurs mit Freunden, aber er tanzt regelmäßig im Nachtclub.

(26) Johanna reist gerne (international). Sie erwägt (zwar) einen Nordseeurlaub auf Langeoog, aber sie fliegt jetzt nach Japan.

(27) Nicole recherchiert gerne (online). Sie kennt (zwar) eine Bibliothek am Stadtrand, aber sie recherchiert aktuell am Computer.

(28) Ole performt gerne (live). Er produziert (zwar) ein Album im Studio, aber er spielt täglich auf Konzerten.

A2: Stimuli for Experiment 2

Below we list the test sentences used for the second experiment. Items were presented in 4 conditions that can be reconstructed by adding either one or both of the optional words in parentheses to the sentence.

(1) Gregory likes to eat (out). (True,) he has an induction stove for the kitchen, but he mostly eats in restaurants.

(2) James likes to run (outdoors). (Sure,) he has a treadmill in the living room, but he primarily runs in parks.

(3) Fred likes to study (alone). (True,) he joins a study group with the classmates, but he often studies at home.

(4) Anna likes to swim (outside). (Sure,) she knows an indoor pool in the city center, but she regularly drives to beaches.

(5) Emma likes to dance (outside). (True,) she attends a dance class in the campus dance room, but she primarily dances in parks.

(6) Hannah likes to paint (alone). (Sure,) she attends an art class with the classmates, but she mostly paints in solitude.

(7) John likes to hike (off-road). (True,) he knows a hiking trail in the mountains, but he often walks through thickets.

(8) William likes to travel (alone). (Sure,) he visits a friend for the summer vacation, but he mostly travels in solitude.

(9) Lucas likes to cook (alone). (True,) he takes a cooking class at the culinary school, but he mostly cooks in solitude.

(10) Chloe likes to read (poems). (Sure,) she reads a newspaper on the train, but she regularly reads in poetry collections.

(11) Leah likes to sing (live). (True,) she records an album in the studio, but she often performs at concerts.

(12) Maya likes to act (live). (Sure,) she has a television role on the regional station, but she regularly acts on stage.

(13) Daniel likes to hear music (live). (True,) he plays a radio in the bedroom, but he often goes to concerts.

(14) David likes to play games (online). (Sure,) he has a board game in the cupboard, but he mostly plays on computers.

(15) Catherine likes to shop (online). (Sure,) she visits a store in the mall, but she primarily shops in online stores.

(16) Leo likes to train (together). (Sure,) he has a weight set in the bedroom, but he often trains in fitness studios.

(17) George likes to bake (alone). (True,) he shares a kitchen in the flat, but he mostly bakes in solitude.

(18) Jessica likes to study (interactively). (Sure,) she owns a book about the exam topic, but she primarily studies with friends.

(19) Gabrielle likes to do research (online). (True,) she knows a library in the city center, but she primarily researches on computers.

(20) Jane likes to relax (inside). (Sure,) she has a balcony in the sun, but she primarily relaxes in bed.

(21) Jonah likes to sunbathe (outside). (True,) he has a tanning membership at the gym, but he regularly lays out at beaches.

(22) Mark likes to work (outside). (Sure,) he has an office job at the government, but he regularly gardens on weekends.

(23) Amelie likes to nap (inside). (True,) she has a hammock in the garden, but she often naps in bed.

(24) Sarah likes to play (outside). (True,) she has a toy collection in the living room, but she regularly goes to playgrounds.

Appendix B

Below, we report an extended version of Table 2. In this version, we lists all cues and connectives that had been grouped into the 'other' category.

Lexical cue (on conceded argument) in source text **true (102) /sure (3)**	Connective (on second argument) in source text									
Lexical cue in German	Connective on 2nd argument (German)	but	however	nevertheless	though	even so	yet	still	although	nonetheless	none
zwar	*aber* ('but')	17	1	–	–	–	–	–	–	1	–
(31)	*doch* ('however')	5	2	–	–	–	–	–	–	–	–
	jedoch ('however')	3	1	–	1	–	–	–	–	–	–
stimmt ('correct')	*aber* ('but')	12	1	1	–	–	–				
(21)	*allerdings* ('however')	–	1	–	–	–	–				
	dennoch ('still')	–	–	1	–	–	–				
	doch ('however')	1	–	–	–	–	–	–	–	–	1
	jedoch ('however')	1	1	–	–	–	–	–	–	–	–
	no connective	–	1	–	–	–	–	–	–	–	–
sicher ('sure')	*aber* ('but')	3	–	–	–	–	–				
(6)	*allerdings* ('however')	–	1	–	–	–	–				
	doch ('however')	–	–	–	1	–	–				
	dennoch ('still')	–	–	–	–	–	–	1 (*sure*)	–	–	–
sicherlich ('surely')	*aber* ('but')	2 (1 *sure*)	–	–	–	–	–	–	–	–	–
(3)	*jedoch* ('however')	1	–	–	–	–	–	–	–	–	–
natürlich ('natural(y)')	*aber* ('but')	3 (1 *sure*)	–	–	–	–	–	–	–	–	–
(8)	*doch* ('however')	3	–	–	–	–	–	–	–	–	–
	jedoch ('however')	–	1	1	–	–	–	–	–	–	–
gewiss ('certain(ly)')	*aber* ('but')	2	–	–	–	–	–	–	–	–	1
(6)	*dennoch* ('still')	–	–	1	–	–	–				
	no connective	–	–	–	1	1	–				
allerdings ('however)	*jedoch* ('however')	1	–	–	–	–	–	–	–	–	–
einverstanden ('agreed')	*aber* ('but')	–	–	–	–	–	–	–	–	1	–
in der Tat ('indeed')	*aber* ('but')	1	–	–	–	–	–	–	–	–	–
(2)	*allerdings* ('however')	1	–	–	–	–	–	–	–	–	–
ist wahr ('is true')	*aber* ('but')	1	–	–	–	–	–	–	–	–	–
(3)	*dennoch* ('still')	–	–	–	–	–	1	–	–	–	–
	jedoch ('however')	–	1	–	–	–	–	–	–	–	–
ja ('yes')	*aber* ('but')	1	1	–	–	–	–	–	–	–	–
(3)	no connective	–	1	–	–	–	–	–	–	–	–
nun gut ('well')	*aber* ('but')	1	–	–	–	–	–	–	–	–	–
richtig ('right')	*aber* ('but')3	–	–	–	–	–	–	–	–	–	–
(4)	*doch* ('however')	–	1	–	–	–	–	–	–	–	–
tatsächlich ('indeed')	*jedoch* ('however')	1	–	–	–	–	–	–	–	–	–
zutreffen ('to be the case')	*aber* ('but')	1	–	–	–	–	–	–	–	–	–
unbestritten ('uncontested')	*aber* ('but')	1	–	–	–	–	–	–	–	–	–
wenn auch ('even if')	*aber* ('but')	–	–	–	–	–	–	–	1	–	–
wirklich ('really')	*allerdings* ('however')	1	–	–	–	–	–	–	–	–	–
(2)	*doch* ('however')	–	1	–	–	–	–	–	–	–	–
zugegeben ('admittedly')	*aber* ('but')	1	–	–	–	–	–	–	–	–	–
zugegebenermaßen ('admittedly')	*doch* ('however')	–	–	–	–	–	–	–	–	–	1
zweifellos ('undoubtedly')	*aber* ('but')	1	–	–	–	–	–	–	–	–	–
(3)	*doch* ('however')	1	–	–	–	–	–	–	–	–	–
	jedoch ('however')	1	–	–	–	–	–	–	–	–	–
no lexical cue	*aber* ('but')	2	–	–	–	–	–	–	–	–	–
(4)	*doch* ('however')	1	–	–	–	–	–	–	–	–	–
	doch ('however')	1	–	–	–	–	–	–	–	–	–

Table 6: Results of a parallel corpus analysis using the German-English aligned version of Europarl. The upper left corner indicates the lexical cue that was present in the original text source (*true/sure*). The two leftmost columns indicate the lexical cue on the conceded argument and the connective on the second argument that were identified in the German translation. Entries in the table provide a count of the frequency with which a combination of *true/sure*+connective was translated into a corresponding cue+connective in German. For example, the first entry of the table (17) shows that *'true/sure..., but'* was translated to *'zwar..., aber'* 17 times.

Cue/Connective in English	Connective on second argument in English	aber	doch	jedoch	dennoch	allerdings	andererseits	auch wenn	dagegen	hingegen	nichtsdestrotz	sondern	trotzdem	wiederum	wohingegen	none
while (61)	*nevertheless*	1	1	–	1	–	–	–	–	–	–	–	–	–	–	–
	on the other hand	1	1	–	–	–	–	–	–	–	–	–	–	1	1	–
	no further connective	28	11	13	2	1	–	–	–	–	–	–	–	–	–	–
whilst (20)	*however*	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	no further connective	9	5	4	1	1	–	–	–	–	–	–	–	1	1	–
although (68)	*however*	–	–	1	–	–	–	–	–	–	–	–	–	–	–	–
	nevertheless	2	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	on the other hand	–	–	–	–	–	–	–	1	–	–	–	–	–	–	–
	no further connective	34	14	12	–	2	–	–	–	–	1	–	–	–	–	1
though (8)	*however*	–	–	1	–	–	–	–	–	–	–	–	–	–	–	–
	nevertheless	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	on the other hand	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	no further connective	4	3	–	–	–	1	–	–	–	–	–	–	–	–	1
modal (may/might) (28)	*but*	16	4	3	–	–	–	–	–	–	–	–	–	–	–	–
	however	1	1	1	–	–	–	–	–	–	–	–	–	–	–	–
	on the other hand	1	1	–	–	–	–	–	–	–	–	–	–	–	–	–
	no connective	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–
admittedly (8)	*but*	4	1	2	–	–	–	–	–	–	–	–	–	–	–	–
	still	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	no connective	–	–	–	–	–	–	–	–	–	–	–	–	–	–	1
to be aware that	*but*	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–
certainly (10)	*but*	4	1	3	–	–	–	–	–	–	–	–	–	–	–	–
	no connective	2	–	–	–	–	–	–	–	–	–	–	–	–	–	–
clearly	*but*	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–
conditional (if...then)	no connective	3	4	–	–	–	–	–	–	–	–	–	–	–	–	–
despite	no connective	–	1	–	–	–	–	–	–	–	–	–	–	–	–	–
even if	no connective	1	2	–	–	–	–	–	–	–	–	–	–	–	–	1
even though	no connective	1	–	1	–	–	–	–	–	–	–	–	–	–	–	–
granted	*but*	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–
to know that	*but*	2	1	–	–	–	–	–	–	–	–	–	–	–	–	–
in fact (3)	*but*	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	however	–	1	–	–	1	–	–	–	–	–	–	–	–	–	–
indeed (4)	*but*	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	nonetheless	–	–	–	1	–	–	–	–	–	–	–	–	–	–	–
	yet	–	–	1	–	–	–	–	–	–	–	–	–	–	–	–
	no connective	–	1	1	–	–	–	–	–	–	–	–	–	–	–	–
it is true (13)	*but*	3	5	–	1	2	–	1	–	–	–	–	–	–	–	–
	however	1	–	1	–	1	–	–	–	–	–	–	–	–	–	–
	nevertheless	1	–	–	–	–	–	–	–	–	–	1	–	–	–	–
	no connective	–	–	–	–	–	1	–	–	–	–	–	–	–	–	1
obviously	*however*	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–
of course (4)	*but*	3	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	however	–	1	1	–	–	–	–	–	–	–	–	–	–	–	–
on the one hand	*on the other hand*	1	–	–	–	–	1	–	–	–	–	–	–	–	–	–
to recognize that	*but*	–	1	–	1	–	–	–	–	–	–	–	–	–	–	–
to be sure	*but*	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–
no lexical cue (248)	*but*	106	49	44	–	2	–	–	–	–	–	–	–	–	–	–
	however	1	2	2	–	1	–	–	–	–	–	–	1	–	–	–
	nevertheless	–	–	1	–	–	–	–	–	–	–	1	–	–	–	–
	yet	2	3	1	1	–	–	–	–	–	–	–	–	–	–	–
	no connective	13	5	2	2	1	1	–	–	–	–	–	–	–	–	7

Table 7: The upper left corner indicates the lexical cue that was present in the original text source (*zwar*). The two leftmost columns indicate the lexical cue on the conceded argument (with total frequency counts in parentheses) and the connective on the second argument that were identified in the English translation. Entries in the table provide a count of the frequency with which a combination of *zwar*+connective in German was translated into a corresponding cue+connective in English. For example, the first entry of the table (1) shows that '*zwar..., aber*' was translated to '*while..., however*' 1 time in the total sample of 500 instances of *zwar*.